JIPET

Journal of Intelligent Procedures in Electrical Technology
Vol. 17/ No. 65/ Spring 2026
P-ISSN: 2322-3871, E-ISSN: 2345-5594, http://jipet.iaun.ac.ir/

https://dorl.net/dor/20.....
Research Article

An Algorithm with Low Execution Time for Anonymizing Graph Degree Sequence Based on Weighting Graph Edges

Maryam Kiabod[1,2], *Ph.D. Student*, Mohammad Naderi Dehkordi[1,2], *Assistant Professor*, Behrang Barekatain[1,2], *Assistant Professor*

[1]Faculty of Computer Engineering- Najafabad Branch, Islamic Azad University, Najafabad, Iran
[2]Big Data Research Center- Najafabad Branch, Islamic Azad University, Najaf Abad, Iran
m_kiabod@sco.iaun.ac.ir, naderi@iaun.ac.ir, behrang_barekatain@iaun.ac.ir

Abstract
Social networks have been introduced as an attractive, low-cost and accessible environment for communication between users. Analyzing the information collected in these networks, as one of the main goals, can violate the privacy of users. For this purpose, several algorithms have been presented to anonymize the graph of social networks, which try to protect the privacy of social network users. However, no method has been proposed that considers the simultaneous improvement of graph utility and execution time. In this regard, the current research tries to overcome this problem by combining two algorithms, time Saving k-degree anonymization method (TSRAM) and NaFa4KDA. The first algorithm reduces the time of anonymization of the degree sequence by drawing a compact tree from the degree sequence of the graph, and the second algorithm uses an effective method to reduce the number of scans of selection edges and also increase the accuracy of the algorithm in selecting the most suitable edge. for editing, it simultaneously improves the execution time of the algorithm and the usefulness of the graph. The results of comparing the proposed algorithm with other similar algorithms show that the proposed combined algorithm has significantly increased and decreased the utility of the graph and the execution time in the process of anonymization. On average, the evaluation results show a 34% improvement in execution time and a 10% improvement in graph utility.

Keywords: optimization, utility, social networks, privacy preserving

Received: 22 January 2023
Revised: 18 February 2023
Accepted: 25 April 2023

Citation: M. Kiabod, M. Naderi Dehkordi, B. Barekatain, "An algorithm with low execution time for anonymizing graph degree sequence based on weighting graph edges", Journal of Intelligent Procedures in Electrical Technology, vol. 17, no. 65, pp. 1-14, June 2026 (in Persian).

https://dorl.net/dor/20.....
مقاله پژوهشی

ارائه الگوریتمی با زمان اجرای کم برای گمنام‌سازی دنباله درجه گراف مبتنی بر وزن‌دهی به یال‌های گراف

مریم کیابد[1]، دانشجوی دکتری، محمد نادری دهکردی[1،2]، استادیار، بهرنگ برکتین[1،2]، استادیار

۱- دانشکده مهندسی کامپیوتر- واحد نجف آباد، دانشگاه آزاد اسلامی، نجف آباد، ایران
۲- مرکز تحقیقات کلان داده- واحد نجف آباد، دانشگاه آزاد اسلامی، نجف آباد، ایران
m_kiabod@sco.iaun.ac.ir, naderi@iaun.ac.ir, behrang_barekatain@iaun.ac.ir

چکیده: شبکه‌های اجتماعی بعنوان یک محیط جذاب، کم هزینه و قابل دسترس برای ارتباط بین کاربران معرفی شده‌اند. تجزیه و تحلیل اطلاعات جمع‌آوری شده در این شبکه‌ها، بعنوان یکی از اصلی‌ترین اهداف، می‌تواند حریم خصوصی کاربران را نقض کند. به این‌منظور، الگوریتم‌های متعددی برای گمنام‌سازی گراف شبکه‌های اجتماعی ارائه شده است که سعی در حفظ حریم خصوصی کاربران شبکه اجتماعی دارند. با این وجود، تاکنون روشی که بهبود همزمان معیارهای سودمندی گراف و زمان اجرا را مد نظر داشته باشند، مطرح نشده است. در این راستا، تحقیق جاری با تلفیق دو الگوریتم روش ناشناس‌سازی تصادفی صرفه جویی در زمان (TSRAM) و نافا برای گمنام‌سازی دنباله درجه گراف (NaFa4KDA)، تلاش دارد تا بر این مشکل مذکور فائق آید. الگوریتم اول از طریق رسم یک درخت فشرده از دنباله درجه گراف، زمان گمنام‌سازی دنباله درجه را کاهش و الگوریتم دوم با بکارگیری یک روش موثر برای کاهش تعداد اسکن‌های یال‌های انتخاب و همچنین افزایش دقت الگوریتم در انتخاب مناسب‌ترین یال‌ها برای ویرایش، بطور همزمان زمان اجرای الگوریتم و سودمندی گراف را بهبود می‌دهد. نتایج مقایسه الگوریتم پیشنهادی با الگوریتم‌های مشابه دیگر نشان می‌دهد که الگوریتم ترکیبی پیشنهادی در فرآیند گمنام‌سازی، سودمندی گراف و زمان اجرا را به ترتیب افزایش و کاهش قابل توجهی داده است. بطور میانگین، نتایج ارزیابی،۳۴ درصد بهبود در زمان اجرا و ۱۰ درصد بهبود سومندی گراف را نشان می‌دهد.

کلمات کلیدی: بهینه‌سازی، سودمندی،شبکه اجتماعی، حفظ حریم خصوصی

تاریخ ارسال مقاله: ۱۴۰۱/۱۱/۲
تاریخ بازنگری مقاله: ۱۴۰۱/۱۱/۲۹
تاریخ پذیرش مقاله: ۱۴۰۲/۲/۵

۱- مقدمه

امروزه اکثر ارتباطات بین انسان‌ها از طریق شبکه‌های اجتماعی انجام می‌شود. کاربران از طریق این شبکه‌ها می‌توانند با افراد دیگر در اقصی نقاط جهان در ارتباط باشند و اطلاعات خود را با آن‌ها به اشتراک بگذارند. بنابراین، این شبکه‌ها منبع مهمی از اطلاعات در مورد فعالیت کاربران در شبکه اجتماعی هستند. از این اطلاعات، می‌توان در جهت بهبود خدمت‌رسانی به کاربر بهره‌برداری نمود. با این حال، این اطلاعات ممکن است منجر به استنتاج اطلاعات حساس کاربران شوند. برای مثال، جنسیت فرد می‌تواند از طریق فعالیت فرد در شبکه‌های مبتنی بر مکان مشخص شود [۱]. این در حالی است که کاربران شبکه‌های اجتماعی تمایلی به آشکارسازی اطلاعات حساس خود ندارند. مطابق تحقیقی که در سال ۲۰۱۲ انجام شده، ۹۰ درصد کاربران شبکه‌های اجتماعی در مورد آشکارسازی اطلاعات حساس خود نگران هستند [۲]. همچنین در سال ۲۰۱۶، یاهو نسخه گمنام شده‌ای از مجموعه داده [۳] را منتشر کرد که نشان از اهمیت این موضوع دارد. بنابراین، لازم است که محرمانگی داده‌های حساس کاربر قبل از انتشار تضمین شود [۱].

برخی از مدل‌ها برای گمنام‌سازی داده‌های جدولی وجود دارد، مانند k-گمنامی[۱] [۴]، l-تنوعی[۲] [۵] و t-نزدیکی[۳] [۶]. این مدل‌ها را نمی‌توان به‌طور مستقیم بر روی گراف اعمال کرد و تلاش‌هایی نیز برای اصلاح این مدل‌ها و استفاده از آن‌ها برای گمنام‌سازی گراف صورت گرفته است. در مورد گراف شبکه اجتماعی، علاوه بر ویژگی‌های کابران، ساختار گراف نیز منتشر می‌شود. بنابراین، متخاصم می‌تواند بر اساس ساختار گراف نیز به شبکه اجتماعی حمله کند [۷]. یکی از ویژگی‌های ساختار گراف، دنباله درجه گراف[۴] است. اگر برخی از درجه‌ها به ندرت در گراف رخ دهد، ممکن است کاربران مرتبط با آن‌ها شناسایی شوند. برای رسیدگی به این مشکل، مدل k-گمنامی تغییر یافت تا بتوان آن را بر روی گراف اعمال کرد [۹]. در این مدل تضمین شده است که برای هر گره v در گراف، حداقل k-1 گره با همان درجه گره v وجود دارد.

الگوریتم‌های ارائه شده برای گمنام‌سازی دنباله درجه گراف، برای گمنام‌سازی داده‌های حجیم مناسب نیستند که دلیل آن زیاد بودن پیچیدگی زمانی این الگوریتم‌ها است [۸]. علاوه بر این، این الگوریتم‌ها باید به منظور به حداقل رساندن گمشدگی اطلاعاتی بهبود یابند [۹]. از دیگر معایب این رویکردها، نادیده گرفتن پویایی گراف است [۱۰]. برخی از این روش‌ها را نمی‌توان برای گراف‌های جهت‌دار اعمال کرد. علاوه بر این، گراف‌های چندلایه، به عنوان گراف‌های چالش برانگیز در فرآیند گمنامی در نظر گرفته می‌شوند [۱۰].

به منظور حل مشکلات فوق، مطالعاتی در مورد مدل k-گمنامی انجام شد که سعی در ایجاد یک گراف گمنام دارند به گونه‌ای که این گراف از نظر ویژگی‌ها به گراف اصلی نزدیک باشد [۱۳-۴،۱۱]. اخیراً، در مرجع [۹]، الگوریتمی را به نام الگوریتم ریز-تجمع تک-متغیره برای گمنام‌سازی گراف[۵] (UMGA) پیشنهاد کردند که گراف گمنام k-درجه را ایجاد می‌کند. هدف این الگوریتم، افزایش سودمندی گراف است و دو مرحله دارد. در مرحله اول، دنباله درجه گراف گمنام‌سازی می‌شود. در مرحله دوم، فرایند ویرایش گراف برای مطابقت گراف با دنباله درجه گمنام انجام می‌شود. اگرچه این الگوریتم‌ها می‌توانند سودمندی گراف را حفظ کنند، اما می‌توانند از نظر کاهش معیار گمشدگی اطلاعاتی بهبود یابند. علاوه بر این، با استفاده از یک ساختمان داده فشرده در فرایند گمنام‌سازی گراف، می‌توان زمان اجرا و میزان مصرف حافظه الگوریتم را کاهش داد.

با این‌حال، بهبود همزمان معیار گمشدگی اطلاعاتی و زمان اجرا می‌تواند چالش برانگیز باشد. برای بهبود معیار گمشدگی اطلاعاتی، الگوریتم گمنام‌سازی باید در هر دو مرحله گمنام‌سازی دنباله درجه گراف و همچنین ویرایش گراف بهبود یابد تا کمترین تغییرات به گراف شبکه اجتماعی اعمال شود. این عملیات مستلزم دقت زیاد در فرایند گمنام‌سازی است که می‌تواند زمان اجرای الگوریتم را افزایش دهد. به منظور پرداختن به این موضوع، این مقاله از تلفیق دو الگوریتم روش گمنام‌سازی تصادفی صرفه جویی در زمان[۶] (TSRAM) [۱۱] و نافا برای گمنام‌سازی دنباله درجه گراف[۷] (NaFa4KDA) [۱۲] که توسط نویسندگان ارائه گردیده است، استفاده می‌کند که می‌تواند بطور همزمان هر دو معیار گمشدگی اطلاعاتی و زمان اجرا را بهبود دهد. روش پیشنهادی با استفاده از تلفیق این دو روش و استفاده از معیارهای مناسب برای انتخاب بهترین یال‌ها برای ویرایش، قادر است به‌طور همزمان زمان اجرا و میزان مصرف حافظ را در کل فرایند گمنام‌سازی گراف شبکه اجتماعی کاهش داده و همچنین گراف گمنامی ایجاد نماید که از نظر ویژگی‌های عمومی و ساختاری گراف به گراف شبکه اجتماعی بسیار نزدیک باشد. با توجه به اینکه

در الگوریتم‌ها، مساله زمان اجرا و همچنین دقت الگوریتم همواره در مقابل یکدیگر قرار دارند، نیاز به استفاده از روشی است که بتواند این دو معیار را بصورت همزمان بهبود دهد. استفاده از تلفیق این دو الگوریتم می‌تواند این هدف را میسر سازد، به‌طور همزمان، زمان اجرای الگوریتم را کاهش داده و سودمندی گراف گمنام را افزایش دهد. برای نیل به این هدف، باید ابتدا مشخص کرد که از کدام الگوریتم در چه مرحله‌ای از فرایند گمنام‌سازی باید استفاده شود تا بتواند بیشترین تاثیر را داشته باشد. با توجه به اینکه برای کاهش زمان اجرای فرایند گمنام‌سازی و همچنین افزایش سودمندی گراف گمنام می‌توان هر دو مرحله فرایند را به منظور بهبود در نظر گرفت، لازم است الگوریتم‌هایی در این دو مرحله مورد استفاده قرار گیرند که تلفیق آن‌ها منجر به حذف نقاط ضعف یکدیگر شده و در نتیجه بهبود گراف گمنام حاصل گردد. استفاده مجزا از هر کدام از الگوریتم‌های TSRAM و NaFa4KDA برای گمنام‌سازی گراف نمی‌تواند هر دو هدف بهبود زمان اجرا و سودمندی گراف را منجر شود که دلیل آن تمرکز این الگوریتم‌ها تنها بر روی یکی از مرحله‌ها فرایند گمنام‌سازی گراف است. به همین دلیل، در این مقاله، از تلفیقی از این الگوریتم‌ها در هر دو مرحله فرایند گمنام‌سازی گراف برای بهبود زمان اجرا و سودمندی گراف استفاده می‌شود. ساختار این مقاله به شرح زیر است. پیشینه تحقیق، در بخش دوم ارائه شده است. بخش سوم بیان مساله را توصیف می‌کند. بخش چهارم روش پیشنهادی را با جزئیات توضیح می‌دهد. بخش پنجم نتایج تجربی را نشان می‌دهد و سرانجام، در بخش ششم نتیجه‌گیری را خواهیم داشت.

۲- پیشینه تحقیق

در این بخش، بررسی مختصری از روش‌های حفظ حریم خصوصی برای انتشار گراف شبکه اجتماعی ارائه شده است. به‌طور کلی، در گراف شبکه اجتماعی سه نوع تهدید وجود دارد: افشای هویت[۸]، افشای ویژگی[۹] و افشای پیوند[۱۰]. افشای هویت به کشف هویت کاربر مربوط می‌شود که مربوط به یک گره در گراف شبکه اجتماعی است. افشای ویژگی هنگامی رخ می‌دهد که ویژگی حساس یک گره آشکار شود و افشای پیوند مربوط به کشف رابطه حساس بین دو کاربر در گراف شبکه اجتماعی است. در این مقاله، گمنام‌سازی هویت مورد توجه قرار گرفته است. زیرا عمومیت بیشتری نسبت به گمنام‌سازی ویژگی دارد. اگر هویت کاربر در گراف شبکه اجتماعی پنهان باشد، ویژگی‌های او نیز پنهان می‌شود.

چندین تحقیق در مورد گمنام‌سازی هویت کاربر در شبکه‌های اجتماعی انجام شده است. به‌طور کلی، این آثار بر روی دو نوع از گراف ساده و برچسب‌دار انجام شده است. با توجه به اینکه تمرکز این مقاله بر روی گراف‌های ساده است، کارهای انجام شده در این زمینه در این بخش مورد بررسی قرار می‌گیرند. به‌طور کلی در الگوریتم‌های گمنام‌سازی گراف شبکه اجتماعی، یال‌ها و گره‌هایی به گراف اضافه می‌شوند تا هویت و صفات حساس کاربر را در گراف شبکه اجتماعی پنهان کنند. در مرجع [۱۳]، با استفاده از برنامه نویسی پویا، الگوریتمی ارائه شده که با محاسبه تعداد گره‌های جعلی لازم برای گمنام‌سازی گراف، تعدادی گره به گراف اضافه می‌نماید. سپس، بین این گره‌های جعلی و گره‌هایی که باید درجه آن‌ها افزایش یابد یال‌هایی اضافه می‌شود. پس از آن، گمنام بودن این رئوس جعلی نیز مورد بررسی قرار می‌گیرد و در صورت لزوم این گره‌ها نیز گمنام می‌شوند. با وجود این‌که در این روش، تعداد یال‌های اضافه شده به گراف کم است، با افزایش سطح حریم خصوصی[۱۱]، زمان اجرا به سرعت افزایش می‌یابد. همچنین، یک الگوریتم دو مرحله‌ای برای گمنام‌سازی دنباله درجه گراف ارائه شده است [۱۴]. در مرحله اول این الگوریتم، درجه هدف برای هر گره در گراف پیدا می‌شود و در مرحله دوم، برخی از یال‌ها بین گره‌ها به منظور افزایش درجه آن‌ها اضافه می‌شود. افزایش درجه همه گره‌ها باعث افزایش تعداد یال‌های گراف و در نتیجه عدم حفظ ویژگی‌های آن خواهد شد. در مرجع [۱۵]، مساله گمنام‌سازی دنباله درجه گراف را توسط اضافه‌کردن رئوس و یال‌ها به گراف مورد بررسی قرار داده‌اند و در مورد پیچیدگی آن بحث شده است. همچنین افزودن رئوس به همراه یال‌ها به گراف مورد بحث قرار گرفته تا بتوانند یک گراف غیرجهت‌دار که دارای k-درجه گمنامی است ایجاد کنند. در مرجع [۹]، الگوریتمی برای گمنام‌سازی دنباله درجه گراف ارائه شده که ابتدا با استفاده از یک گراف از دنباله درجه، درجه مقصد را برای هر گره به‌دست می‌آورد و سپس از معیار مرکزیت همسایگی برای انتخاب مناسب‌ترین یال جهت حذف، اضافه و یا جابجا شدن استفاده می‌کند. این الگوریتم دارای دو مرحله گمنام‌سازی دنباله درجه گراف و تغییر گراف است. در مرحله اول این الگوریتم، دنباله درجه گراف شبکه اجتماعی طوری گمنام می‌شود که از هر

درجه در آن حداقل k عدد وجود داشته باشد. سپس در مرحله دوم، گراف شبکه اجتماعی به گونه‌ای تغییر می‌کند که دنباله درجه آن بر دنباله درجه گراف گمنام شده تطبیق یابد. پس از گمنام‌سازی دنباله درجه گراف، ویرایش گراف شبکه اجتماعی انجام می‌شود. در واقع، گراف طوری تغییر می‌یابد که دنباله درجه آن بر دنباله درجه گمنام شده در مرحله قبل تطبیق یابد. این عملیات توسط ویرایش یال‌ها و رئوس گراف شبکه اجتماعی انجام می‌شود. الگوریتم UMGA تنها از ویرایش یال برای تغییر گراف شبکه اجتماعی استفاده می‌کند که توسط سه عملیات حذف یال، اضافه‌کردن یال و سوئیچ‌کردن یال انجام می‌شود. در سال ۲۰۱۷، یک مدل گمنام‌سازی k-درجه ارائه شده که خواص ساختاری شبکه اجتماعی را حفظ می‌کند و حریم خصوصی افراد در شبکه اجتماعی را تضمین می‌نماید [۱۶]. در این الگوریتم، با توجه به اینکه اگر یالی بین دو گره با فاصله زیاد ایجاد شود، انحرافات بیشتری را در ساختار گراف ایجاد می‌کند، رئوسی برای متصل شدن به یکدیگر انتخاب می‌شوند که فاصله کمی از یکدیگر داشته باشند. در این روش، ابتدا گره‌های گراف به خوشه‌هایی تقسیم می‌شوند که داخل هر خوشه بیشترین اتصال و بین خوشه‌ها کمترین اتصال را با یکدیگر داشته باشد. در این رویکرد خوشه‌بندی، ابتدا گره‌های دارای درجه بالاتر به عنوان نماینده خوشه در نظر گرفته می‌شوند وسپس رئوس دیگر بر اساس مقدار اتصال آن‌ها با این نماینده به خوشه اضافه می‌شوند. پس از اتمام خوشه‌بندی، یال‌های جدید تنها بین رئوس موجود در یک خوشه اضافه می‌شود. سایز هر خوشه بر اساس مقدار k در مدل k-گمنامی مشخص می‌شود. در ابتدا n گره با بالاترین درجه به عنوان نماینده خوشه‌ها در نظر گرفته می‌شوند و سپس، گره‌هایی که مستقیما به این گره‌ها متصل هستند به خوشه اضافه می‌شوند. برای هر گره، ۱-همسایگی آن به‌طور تکراری بررسی می‌شود و در صورتی‌که تعداد یال‌های بین آن گره و خوشه بیشتر از حد آستانه بود، آن گره به خوشه اضافه می‌شود. این الگوریتم دارای ویژگی حفظ سودمندی گراف است با این‌حال دارای زمان اجرای زیاد و مصرف حافظه زیاد برای نگهداری گراف پارتیشن-بندی است. به همین منظور باید الگوریتمی در این زمینه ارائه شود که این چالش‌ها را آدرس‌دهی نماید. در مرجع [۱۷]، یک نمونه گسترش‌یافته از مدل k-گمنامی ارائه شده که از خوشه‌بندی برای گمنام‌سازی کاربران استفاده می‌کند و دانش پیش‌زمینه متخاصم را میزان اتصال گره‌ها به یکدیگر در نظر می‌گیرد. در این الگوریتم، براساس اتصالات بین گره‌ها در گراف شبکه اجتماعی، گروه‌هایی از گره‌ها ایجاد می‌شود. گره‌هایی با درجه یکسان، داخل یک گروه قرار می‌گیرند و گره‌هایی با درجه متفاوت در گره‌های متفاوتی قرار می‌گیرند. اگر یالی بین گره‌های موجود در یک گروه وجود نداشته باشد، متخاصم می‌تواند به این گره‌ها حمله کند. به همین دلیل، به منظور حفظ حریم خصوصی، یال‌هایی بین آن گره‌ها اضافه می‌شود. در مرجع [۱۸]، روشی برای ویرایش گراف ارائه شده که با حفظ هسته گراف طوری یال‌های گراف تغییر می‌کند که ساختار جامعه گراف حفظ شود. با وجود این‌که این الگوریتم شماره هسته رئوس را بعد از گمنام‌سازی حفظ می‌نماید، با تغییر سطح حریم خصوصی گراف، باید الگوریتم از ابتدا اجرا شود که این مساله زمان‌بر است. در مرجع [۱۹]، مروری بر روش‌های حفظ حریم خصوصی ارائه شده و روش‌های موجود را به دو دسته روش‌های حفظ حریم خصوصی گره‌ها و روش‌های حفظ حریم خصوصی یال‌ها تقسیم کرده‌اند. در مرجع [۲۰]، الگوریتمی برای گمنام‌سازی دنباله درجه گراف جهت‌دار پویا ارائه شده که از اضافه‌کردن گره‌های ساختگی به گراف استفاده می‌کند و همزمان با اضافه شدن گره، آن را گروه‌بندی نیز می‌کند. سپس گره‌هایی به هم متصل می‌شوند که ساختار پیمانه‌بندی گراف را کمتر تغییر می‌دهند. با وجود این‌که این روش در گمنام‌سازی گراف‌های پویای جهت‌دار موثر است، تعاملی نیست و با تغییر سطح حریم خصوصی، الگوریتم باید از ابتدا اجرا شود. در مرجع [۲۱]، حریم خصوصی کاربران را در سطح شبکه به جای سطح کاربر در نظر گرفته‌اند و روشی ارائه شده که حریم خصوصی گره‌ها و یال‌ها و صفات را توسط خوشه‌بندی کاربران براساس شباهت ویژگی‌های آن‌ها با هم ترکیب می‌کند. در این روش علاوه بر مدل k-گمنامی، مدل l-تنوعی نیز در نظر گرفته شده است. حفظ حریم خصوصی در نشرهای ترتیبی موضوعی است که در سال ۲۰۲۰ در مرجع [۲۲] مورد بررسی قرار گرفته است. در این روش، سه نیازمندی حریم خصوصی کاربران، حریم خصوصی گروه و حریم خصوصی یال‌ها به‌طور همزمان در نظر گرفته شده است. در مرجع [۲۳]، روشی بر اساس حریم خصوصی تفاضلی ارائه شده است. در مرجع [۲۴]، در سال ۲۰۲۱ ثابت کردند که حداقل تعداد چرخش یال در گراف به منظور گمنام‌سازی گراف شبکه اجتماعی، NP سخت[۱۲] است. در همان سال، در مرجع [۲۵]، الگوریتمی ارائه شد که توسط کمترین تعداد یال اضافه شده به گراف، گراف شبکه اجتماعی را گمنام‌سازی می‌کند. در مرجع [۲۶]، در سال ۲۰۲۱ مدل حمله براساس دوستان[۱۳] را مورد بررسی قرار داده‌اند و الگوریتمی ارائه کردند که حمله k-

درجه را نیز پشتیبانی می‌کند. سرانجام در همان سال، در مرجع [۱۲]، روشی سریع برای ویرایش گراف شبکه اجتماعی ارائه کردند که مدل k-گمنامی را در گراف تضمین می‌کند. در این روش از یک روش موثر برای شناسایی و انتخاب یال‌های مناسب برای ویرایش گراف بهره برده شد که همزمان با کاهش زمان اجرا، سودمندی گراف را نیز افزایش می‌دهد. در مرجع [۲۷]، الگوریتمی برای مساله k-گمنامی گراف شبکه اجتماعی ارائه کرده‌اند که از ساختار درخت برای گمنام‌سازی دنباله درجه گراف استفاده شده و سودمندی گراف گمنام بهبود می‌یابد. در مرجع [۲۸]، یک روش ارائه شده، که مقدار نویز اعمال شده به داده‌ها را کاهش داده و حمله به d-همسایگی گراف را کاهش می‌دهد. در این روش از گمنام‌سازی یال‌ها و رئوس استفاده شده است. در مرجع [۲۹]، الگوریتم‌های ارائه شده برای حفظ حریم خصوصی شبکه اجتماعی مورد بررسی قرار گرفته و موضوعاتی به عنوان کارهای آینده پیشنهاد شده است. در مرجع [۳۰]، روشی برای گمنام‌سازی داده‌ها ارائه شده که سودمندی داده‌ها را در حد مطلوبی نگه می‌دارد. سرانجام در مرجع [۲۷]، یک نسخه بهبود یافته از الگوریتم TSRAM [۱۱] را در سال ۲۰۲۲ ارائه داده‌اند که از اضافه‌کردن نویز به داده‌ها در زمانی که به دنباله درجه مطلوب نرسیده باشد، استفاده می‌کند. جدول (۱) خلاصه‌ای از الگوریتم‌هایی که اخیرا در این زمینه پیشنهاد شده را ارائه می‌دهد. در الگوریتم پیشنهادی، مشکل عدم وجود ویژگی تعاملی که در الگوریتم‌های قبلی وجود دارد و منجر به اسکن مجدد مجموعه داده در صورت تغییر سطح گمنامی می‌شود، آدرس‌دهی شده است. علاوه بر این، سودمندی گراف شبکه اجتماعی که گمنام‌سازی شده است نیز نسبت به الگوریتم‌های قبلی بهبود می‌یابد.

۳- بیان مساله

همان‌طور که بیان شد، الگوریتم‌های فعلی دارای چالش‌هایی هستند که از جمله آن‌ها می‌توان به زمان اجرا، میزان مصرف حافظه و کاهش سودمندی گراف گمنام اشاره نمود. به منظور آدرس‌دهی این چالش‌ها دوالگوریتم TSRAM و NaFa4KDA ارائه شده که هر کدام سعی در بهبود معیارهای بالا در یکی از مرحله‌ها فرایند گمنام‌سازی گراف شبکه اجتماعی دارند. با این‌حال، این الگوریتم‌ها دارای چالش‌های زیر هستند:

الف- با وجود آن‌که الگوریتم TSRAM که برای مرحله اول فرایند گمنام‌سازی گراف شبکه اجتماعی (مرحله گمنام‌سازی دنباله درجه گراف) ارائه گردید، زمان اجرا و میزان مصرف حافظه آن را به شدت کاهش می‌دهد، تا حدودی باعث کاهش سودمندی گراف شبکه اجتماعی می‌شود [۱۱].

ب- الگوریتم NaFa4KDA زمان اجرای برنامه را به شدت کاهش می‌دهد و سودمندی گراف گمنام را افزایش می‌دهد. با این‌حال، با توجه به ایجاد یک گراف بزرگ از رئوس در حافظه برای یافتن دنباله درجه گمنام، میزان مصرف حافظه آن زیاد است.

به منظور آدرس‌دهی چالش‌های ذکر شده برای هر کدام از این الگوریتم‌ها، در این مقاله، از تلفیق این دو الگوریتم استفاده می‌شود به گونه‌ای که مزایای هر کدام از آن‌ها بتواند معایب آن‌ها را برطرف نماید. به این منظور، ابتدا دنباله درجه گراف شبکه اجتماعی توسط الگوریتم TSRAM گمنام‌سازی شده و سپس این دنباله درجه به الگوریتم NaFa4KDA داده می‌شود تا دنباله درجه گراف شبکه اجتماعی بر این دنباله درجه تطبیق یابد.

۴- گمنام سازی گراف شبکه اجتماعی

از آنجاکه کاربران در شبکه اجتماعی فعالیت می‌کنند، اطلاعات زیادی در گراف شبکه اجتماعی در مورد کاربران ذخیره می‌گردد که باید به منظور حفظ حریم خصوصی کاربران گمنام‌سازی شوند. یکی از ویژگی‌هایی که با گمنام‌سازی آن می‌توان حریم خصوصی کاربران را به خوبی حفظ نمود، گمنام‌سازی دنباله درجه گراف است. بنابراین، تمرکز این مقاله بر روی این جنبه از گمنام‌سازی است. در این نوع از گمنام‌سازی، دنباله درجه گراف از تجمیع درجات گره‌های گراف در یک دنباله ایجاد می‌گردد و به منظور انجام عملیات، در ابتدا این دنباله درجه مرتب‌سازی می‌شود. گرافی که در این مقاله مورد استفاده قرار می‌گیرد، گرافی بدون برچسب و غیرجهت‌دار است. به منظور گمنام‌سازی دنباله درجه گراف شبکه اجتماعی، در این مقاله از مدل گمنام‌سازی k-درجه استفاده می‌گردد. در این مدل تضمین می‌گردد که از هر درجه در گراف حداقل k-1 گره با آن درجه وجود دارد. به این ترتیب، شناسایی کابران شبکه اجتماعی براساس درجه گره‌های گراف ممکن نخواهد بود.

Table (1): Some recent methods in graph modification techniques

جدول (۱): خلاصه‌ای از الگوریتم‌های پیشنهادی در زمینه حریم خصوصی

منبع	نوع ÷گراف	مدل	روش	ویژگی‌ها	مزایا	معایب
[۹]	ساده	k- دنباله درجه گمنامی	ابتدا توسط الگوریتم حریصانه درجه مقصد را برای هر گره محاسبه کرده و سپس از معیار مرکزیت همسایگی برای تغییر گراف استفاده می‌کند.	ویرایش یال	افزایش سودمندی گراف شبکه اجتماعی	- زمان اجرای زیاد و مصرف حافظه بالا - عدم وجود ویژگی تعاملی
[۱۴]	ساده	k-درجه گمنامی با ویرایش رئوس و یال‌های گراف	از یک الگوریتم حریصانه برای پارتیشن بندی گره‌ها استفاده می‌کند. همچنین عضویت گره‌ها را در جوامع برای تغییر یال‌های گراف در نظر می‌گیرد.	ویرایش گره و یال	ویژگی‌های ساختاری گراف را به خوبی حفظ می‌کند	- کاهش زیاد سودمندی گراف - عدم وجود ویژگی تعاملی
[۱۸]	ساده	k- هسته	حذف و اضافه کردن یال-هایی که شماره هسته رئوس را تغییر نمی‌دهد	ویرایش یال	شماره هسته رئوس را حفظ می‌کند	- با افزایش مقدار k زمان اجرا به سرعت افزایش می-یابد - عدم وجود ویژگی تعاملی
[۱۶]	ساده	k- دنباله درجه گمنامی	خوشه‌بندی	اتصال رئوس با فاصله کم به یکدیگر	تغییرات کمتر در گراف نسبت به روش‌های قبلی	- عدم وجود ویژگی تعاملی
[۱۷]	ساده	k- دنباله درجه گمنامی	خوشه‌بندی	یافتن گره‌هایی با ریسک حمله به آن‌ها و تضمین محرمانگی آن-ها	حفظ بیشتر سودمندی گراف	- عدم وجود ویژگی تعاملی
[۳۱]	جهت‌دار	k- دنباله درجه گمنامی	محاسبه درجه مقصد برای هر نود و سپس ویرایش یال‌های گراف	در نظر گرفتن دانش پیش زمینه جدید برای متخاصم	مناسب برای مجموعه داده-های بزرگ	- عدم وجود ویژگی تعاملی
[۳۲]	برچسب‌دار	K-موقعیت گمنامی	خوشه‌بندی	جلوگیری از تاثیر نویز بر روی گروه‌های گمنام	کاهش تضاد بین سودمندی گراف و نیازمندی‌های محرمانگی	- عدم وجود ویژگی تعاملی
[۲۰]	جهتدار و غیربرچسبدار	k-دنباله درجه گمنامی	اضافه کردن گره‌های مجازی به گراف و گروه-بندی پویای آن‌ها	افزودن راس و یال	قابلیت پردازش گراف‌های بزرگ جهتدار و پویا	- عدم وجود ویژگی تعاملی

۵- روش تی‌اس.رنک

در این قسمت، روش پیشنهادی برای آدرس‌دهی به چالش‌های مطرح شده در قسمت بیان مساله ارائه خواهد گردید. در این روش از یک گراف بدون جهت و بدون برچسب برای گمنام‌سازی استفاده می‌شود. عملیات مورد استفاده برای ویرایش گراف در روش پیشنهادی شامل عملیات حذف و عملیات اضافه‌کردن یال است. بنابراین، تعداد گره‌های گراف تغییری نخواهد کرد. هدف

این الگوریتم بهبود همزمان زمان اجرا، میزان مصرف حافظه و همچنین سودمندی گراف گمنام است. این هدف با استفاده از تلفیق دو الگوریتم TSRAM و NaFa4KDA در مورد گراف شبکه اجتماعی به‌دست می‌آید. این الگوریتم که تی‌اس‌رنک (tsRANK) نام دارد دارای دو مرحله است. در مرحله اول از این الگوریتم، گمنام‌سازی دنباله درجه گراف (گمنام‌سازی دنباله درجه گراف) صورت می‌گیرد و در مرحله دوم (مرحله ویرایش گراف)، گراف شبکه اجتماعی به گونه‌ای تغییر می‌کند که دنباله درجه آن بر دنباله درجه گمنام منطبق گردد. از آنجا که الگوریتم TSRAM از نظر زمان اجرا و میزان مصرف حافظه بهینه است، میزان زمان اجرا و حافظه مصرفی فرایند گمنام‌سازی گراف شبکه اجتماعی را کاهش می‌دهد. با این‌حال این الگوریتم می‌تواند از نظر سودمندی گراف شبکه اجتماعی بهبود یابد [۱۱]. با توجه به اینکه دنباله درجه گمنام ایجاد شده در مرحله اول الگوریتم به مرحله ویرایش گراف داده می‌شود، می‌توان در مرحله ویرایش گراف از الگوریتمی استفاده کرد که با انتخاب یال‌های مناسب برای ویرایش، سودمندی گراف شبکه اجتماعی را افزایش دهد. علاوه براین، با توجه به این‌که یکی از چالش‌های مرحله دوم الگوریتم، زمان اجرای زیاد است، الگوریتم انتخابی باید قادر به کاهش زیاد زمان اجرا باشد. به این منظور، برای ویرایش گراف از روش مورد استفاده در NaFa4KDA [۱۲] استفاده می‌گردد. هر کدام از این مرحله‌ها در ادامه توضیح داده خواهد شد.

۱-۵- گمنام‌سازی دنباله درجه گراف

اولین مرحله در فرایند گمنام‌سازی گراف شبکه اجتماعی، گمنام‌سازی دنباله درجه گراف است. در این مرحله، درجه گره‌های گراف شبکه اجتماعی به گونه‌ای تغییر می‌کند که شناسایی گره فرد قربانی براساس درجه آن گره میسر نباشد. به این منظور، در این مقاله از الگوریتم TSRAM استفاده می‌گردد. در این الگوریتم، از دو مرحله برای گمنام‌سازی دنباله درجه گراف استفاده می‌گردد که عبارتند از ایجاد یک درخت از دنباله درجه گراف و رسم یک خط برش بر روی این درخت.

۱-۱-۵- ایجاد درخت از دنباله درجه گراف

درخت ایجاد شده در این مرحله، یک درخت فشرده است که دنباله درجه گراف را در خود ذخیره می‌کند. این کار توسط ایجاد یک بردار تجمعی[۱۴] (AV) از دنباله درجه گراف انجام می‌شود. بردار تجمعی برداری است که شامل هر درجه به همراه تعداد رخداد آن درجه در گراف است. این بردار به‌صورت رابطه (۱) تعریف می‌شود که در آن AV بیانگر بردار تجمعی است و |D| تعداد رخداد درجه D در دنباله درجه گراف را نشان می‌دهد.

$$\mathrm{AV}(\mathrm{G}) = \left\{(\mathrm{D}, \mathrm{C}) \mid \mathrm{D} \in \text{ degree_sequence}(\mathrm{G}) \text{ and } \mathrm{C} = |\mathrm{D}|\right\} \quad (۱)$$

پس از آن‌که بردار تجمعی از دنباله درجه گراف ایجاد گردید، یک درخت دودویی به‌صورت پایین به بالا رسم می‌شود. عناصر بردار تجمعی، برگ‌های این درخت را تشکیل می‌دهند و از ادغام هر دو برگ در این درخت، گره‌های موجود در سطح بالاتر این درخت ایجاد می‌گردند. به منظور ادغام این گره‌ها، در هر مرحله، گره‌ای با کمترین تعداد رخداد انتخاب می‌شود و با یکی از همسایگانش ترکیب می‌شود. گره ایجاد شده، گره‌ای خواهد بود که دارای دو ویژگی درجه و تعداد رخداد است. برای به‌دست آوردن درجه گره جدید از روش مورد استفاده در مرجع [۹] استفاده می‌شود. در این روش، یک تابع احتمال برای محاسبه درجه گره به‌کار می‌رود. تعداد رخداد نیز از جمع تعداد رخداد گره‌های ترکیبی به‌دست خواهد آمد. این نکته در رابطه (۲) مشاهده می‌شود.

$$\left(\mathcal{D}(\mathrm{v_i}), \mathcal{C}(\mathrm{v_i})\right) \cup \left(\mathcal{D}(\mathrm{v_j}), \mathcal{C}(\mathrm{v_j})\right) \rightarrow (\mathcal{TD}(\mathrm{v}), \mathcal{C}(\mathrm{v_i}) + \mathcal{C}(\mathrm{v_j}) \quad (۲)$$

که در آن، v_i و v_j گره‌هایی هستند که قرار است با هم ادغام شوند و D و C به ترتیب، درجه هر گره و تعداد رخداد آن را نشان می‌دهد. TD درجه گراف جدید را بیان می‌کند. به منظور انتخاب بهترین همسایه برای ادغام شدن، در این الگوریتم از معیار زمان سفر[۱۵] [۳۳] استفاده می‌شود. در این معیار، از فاصله اقلیدسی بین دو گره برای محاسبه شباهت بین آنها استفاده می‌شود. سپس، بر اساس آن، گره‌ای با شباهت بیشتر انتخاب شده و با آن گره ادغام می‌شود. این معیار مطابق رابطه (۳) در نظر گرفته می‌شود که در آن Φ_i پتانسیل نقطه i است و به‌صورت رابطه (۴) محاسبه می‌گردد. در این رابطه، $\Phi_{ij}(r_{ij})$ پتانسیل بین دو نقطه است و توسط رابطه (۵) به‌دست می‌آید که در آن، r_{ij} فاصله اقلیدسی بین دو نقطه است و توسط رابطه (۶) محاسبه می‌گردد.

$$S_{ij} = \frac{\Phi_i - \Phi_j}{r_{ij}^2} \quad (۳)$$

$$\Phi_i = \sum_{j=1\ldots N} \Phi_{ij}\left(r_{ij}\right) \quad (۴)$$

$$Q_{ij}\left(r_{ij}\right) = \frac{1}{r_{ij}} \quad (۵)$$

$$r_{ij} = \left|\deg\left(p_i\right) - \deg\left(p_j\right)\right| \quad (۶)$$

این عملیات آنقدر تکرار می‌شود که ریشه درخت تشکیل گردد و گره‌ای برای ادغام شدن باقی نماند. مرحله‌های این الگوریتم، در الگوریتم (۱) نشان داده شده است [۱۱]. در خط اول این شبه کد ابتدا دنباله درجه گراف به‌دست می‌آید و سپس در خط دوم، بردار تجمعی (AV) آن محاسبه می‌گردد. سپس، در خط سوم، برگ‌های درخت توسط بردار تجمعی ایجاد می‌گردند. همان‌طور که قبلا توضیح داده شد، هر کدام از بردارهای تجمعی، یک برگ از درخت را تشکیل می‌دهد. پس از آن تا رسیدن به ریشه درخت، در خط ۵ گره‌ای با کمترین تعداد، انتخاب شده و در خط ۶ این گره با یکی از همسایگان آن ادغام می‌شود. سپس، در خط ۷، درجه مقصد و تعداد برای گره جدید براساس رابطه (۲) محاسبه می‌شود.

۲-۱-۵- رسم خط برش بر روی درخت دنباله درجه گراف

پس از ایجاد درخت دنباله درجه گراف، می‌توان با استفاده از آن، دنباله درجه گمنام گراف را به‌دست آورد. به این منظور، کافی است که یک خط برش بر روی درخت رسم گردد. گره‌های موجود بر روی این خط برش، دنباله درجه گمنام را تشکیل می‌دهند. برای رسم خط برش بر روی درخت دنباله درجه گراف، از ریشه درخت شروع کرده و در هر شاخه از درخت آنقدر به سمت برگ‌ها حرکت می‌کنیم که به گره‌ای برسیم که تعداد رخداد حداقل یکی از فرزندان آن کمتر از سطح گمنامی مورد نیاز (مقدار k) است. خط برش بر روی این گره از درخت در آن زیرشاخه متوقف می‌گردد.

۲-۵- ویرایش گراف

پس از ایجاد دنباله درجه گمنام ایجاد شده توسط الگوریتم TSRAM، گراف شبکه اجتماعی توسط حذف و اضافه شدن یال‌ها تغییر می‌یابد تا دنباله درجه آن بر دنباله درجه گمنام تطبیق یابد. بنابراین، در این مرحله از گمنام‌سازی گراف، دو عمل حذف و اضافه شدن یال‌ها انجام می‌شود. با توجه به این‌که هدف در این مرحله از فرایند گمنام‌سازی گراف، کاهش زمان اجرا و همچنین افزایش سودمندی گراف است، از روش پیشنهادی در الگوریتم NaFa4KDA برای حذف و اضافه‌کردن یال‌ها استفاده می‌شود. در ادامه، روش حذف و اضافه‌کردن یال‌ها توضیح داده می‌شود.

۱-۲-۵- حذف یال

به‌منظور انتخاب بهترین یال‌ها برای حذف شدن از گراف، ابتدا یال‌هایی که کاندید حذف‌شدن هستند مشخص می‌شوند. این یال‌ها به رئوسی متصل هستند که باید درجه آن‌ها کاهش یابد. سپس، برای هر کدام از این یال‌ها، وزنی محاسبه می‌شود که ارزش آن یال را مشخص می‌کند. به این ترتیب، یال‌هایی برای حذف شدن انتخاب می‌شوند که کمترین ارزش را داشته باشند. به منظور محاسبه وزن یال‌ها از ترکیب سه معیار استفاده می‌گردد که عبارتند از مرکزیت همسایگی[۱۶] [۹]، شماره هسته گره‌های انتهایی یال[۱۷] [۱۲] و تعداد مثلث‌هایی که یال در آن‌ها شرکت می‌کند [۱۱]. رابطه (۷) این معیار را بیان می‌کند.

$$\text{Remove_score}(e) = \text{NC}(e) + \text{coreness}(e) + \text{PL}(e) \quad (۷)$$

که در آن، NC، مرکزیت همسایگی یک یال، coreness، شماره هسته گره‌های انتهایی و PL، تعداد مثلث‌هایی است که یال در آن‌ها شرکت می‌کند. مرکزیت همسایگی یک یال، تعداد رئوسی است که همسایه vi یا vj هستند ولی همسایه هر دو نیستند. شماره هسته گره‌های انتهایی یال، مشخص می‌کند که یک گره چقدر منجر به اتصال رئوس دیگر به‌هم شده است.

Algorithm (1): Creating a tree structure

الگوریتم (۱): ایجاد ساختار درختی

```
Input: graph G
Output: the tree structure of graph G
1. DS = degree_sequence (G);
2. AV= aggregate_vector (DS);
3. Init_leaves (AV);
4. While (!root) {
5.         NLC = node_with_lowest_count;
6.     New_node = merge (NLC , NLC_neighbor);
7.     Compute_TD&C (new_node);
8. }
9. Return the tree structure
```

بنابراین، اگر رئوس انتهایی یال، شماره هسته کمتری داشته باشند، رئوس کمتری را به هم متصل کرده‌اند و بنابراین تاثیر چندانی در ویژگی‌های گراف از جمله میانگین کوتاه‌ترین مسیرها نخواهد داشت. سپس، یال‌ها به ترتیب صعودی براساس ارزش مرتب می‌شوند و یال‌های موجود در ابتدای لیست برای حذف شدن انتخاب می‌شوند. علاوه بر محاسبه میزان ارزش هر یال، باید معیاری وجود داشته باشد که تعداد یال‌های قابل حذف از یک راس را نشان می‌دهد. به این منظور، ابتدا به هر راس یک عدد اول اختصاص می‌یابد و سپس از معیار وزن[۱۸] برای بیان مقدار کاهش درجه هر راس استفاده می‌شود. برای اختصاص عدد اول به رئوس، از عدد ۲ شروع کرده و اعداد اول بعد از آن، به‌ترتیب به رئوس اختصاص پیدا می‌کند. این معیار در رابطه (۸) دیده می‌شود.

$$\text{Weight} = \prod_{i=1}^{n} \text{prime}_i^{n_i} \qquad (۸)$$

در این رابطه، prime عدد اول اختصاص یافته به گره i را نشان می‌دهد و n_i مقدار کاهش درجه گره i را بیان می‌کند. پارامتر n تعداد گره‌هایی است که باید درجه آن‌ها کاهش یابد. عدد اول اختصاص یافته به یک راس به تمامی یال‌هایی که به آن راس متصل است داده می‌شود تا بتوان توسط آن، یال‌های متصل به هر راس را تشخیص داد. در این مقاله به این عدد پوشش[۱۹] گفته می‌شود. به منظور انتخاب یال‌هایی برای حذف شدن از گراف، ابتدا یال‌ها براساس ارزش به‌صورت صعودی مرتب می‌شوند و سپس، از ابتدای لیست، این یال‌ها انتخاب شده و مقدار پوشش آن‌ها در هم ضرب می‌گردند. این عملیات آنقدر تکرار می‌شود که ضرب پوشش یال‌ها با مقدار وزن مساوی گردد. در این هنگام الگوریتم خاتمه می‌یابد. الگوریتم (۲) این فرایند را نشان می‌دهد [۱۲]. در الگوریتم (۲)، پارامتر پوشش کلی[۲۰] مقدار ضرب پارامتر پوشش یال‌ها را نشان می‌دهد. G گراف شبکه اجتماعی و G' گراف گمنام است.

۲-۲-۵- اضافه‌کردن یال به گراف

به منظور اضافه‌کردن یال به گراف، ابتدا رئوسی که باید درجه آن‌ها افزایش یابد مشخص می‌شود. سپس، به تعداد افزایش درجه آن‌ها در یک بردار قرار می‌گیرند. این بردار به عنوان ورودی به یک الگوریتم بهینه‌سازی داده می‌شود تا رئوسی مشخص شوند که باید توسط یال به‌هم متصل گردند. در بردار به‌دست آمده توسط الگوریتم بهینه‌سازی، یک یال بین هر دو راس متوالی در بردار به‌دست آمده توسط الگوریتم بهینه‌سازی رسم می‌گردد. برای تعریف تابع برازش، از دو معیار مرکزیت همسایگی [۹] و تعداد مثلث‌هایی که یال در آن‌ها شرکت می‌کند [۱۱] استفاده می‌گردد. بنابراین، تابع برازش به‌صورت رابطه (۹) تعریف می‌شود.

$$\text{Fitness_Function} = \sum_{i=1}^{n} (\text{NC}(e_i) + \text{PL}(e_i)) \qquad (۹)$$

که در آن NC مقدار مرکزیت همسایگی و PL تعداد مثلث‌هایی است که یال در آن‌ها شرکت می‌کند. از بین بردارهای موجود، برداری برای افزودن یال به گراف انتخاب می‌شود که کمترین مقدار تابع برازش را داشته باشد.

Algorithm (2): Evaluating and deleting graph edges [12]

الگوریتم (۲): اعتبارسنجی و حذف یال‌های گراف [۱۲]

```
Input: negative nodes, sorted remove-vector, mask-vector, weight, graph G
Output: modified graph G'
    1.  Sort edges in remove_vector by their score in ascending order
    2.  G' = G
    3.  TotalMask = 1;
    4.  TempMask = 1;
    5.  For (edge i in remove_vector)
    6.  Find the mask_i value related to the edge i  in mask vector
    7.  Multiply TempMask by mask_i; i.e. TempMask = TotalMask × mask_i
    8.  If weight % TempMask == 0
    9.          TotalMask = TempMask;
    10.         Remove the edge from G'
    11. End
    12. End
    13. Return the modified graph G'
```

۳-۵- پیچیدگی زمانی

برای ارزیابی روش پیشنهادی می‌توان از تحلیل پیچیدگی زمانی الگوریتم استفاده نمود. الگوریتمی با پیچیدگی زمانی کمتر نسبت به سایر الگوریتم‌ها بهتر خواهد بود.

تئوری- پیچیدگی زمانی روش پیشنهادی مطابق با رابطه (۱۰) است که n تعداد رئوس و m تعداد یال‌ها است. k تعداد شاخه‌های موجود در زیر ریشه درخت است.

$$O(n2+m\log m+km) \qquad (۱۰)$$

اثبات- برای تحلیل پیچیدگی زمانی الگوریتم پیشنهادی باید پیچیدگی زمانی هر دو الگوریتم مورد استفاده در دو مرحله گمنام‌سازی در نظر گرفته شود. پیچیدگی زمانی الگوریتم TSRAM که در مرحله اول فرایند گمنام‌سازی دنباله درجه گراف مورد استفاده قرار گرفت مساوی رابطه (۱۱) است [۱۱].

$$TimeComplexity_TSRAM = O(n+m\log m+km) \qquad (۱۱)$$

همچنین پیچیدگی زمانی الگوریتم NaFa4KDA برای ویرایش گراف مساوی رابطه (۱۲) است [۱۲] که n تعداد رئوس و m تعداد یال‌ها است.

$$TimeComplexity_NaFa4KDA = O(n2+m\log m) \qquad (۱۲)$$

بنابراین، پیچیدگی زمانی الگوریتم پیشنهادی با رابطه (۱۰) است.

۶- نتایج تجربی

در این بخش، برای مقایسه و ارزیابی الگوریتم پیشنهادی با الگوریتم‌های موجود از روش‌های موجود که به منظور گمنام‌سازی دنباله درجه گراف ارائه شده‌اند استفاده می‌شود. اولین در مرجع [۱۳] ارائه گردید و افزودن-راس[۲۱] (VA) نام دارد. دومین روش KDVEM (k-درجه گمنامی با ویرایش رئوس و یال‌های گراف[۲۲]) [۱۴] است. سومین الگوریتم نیز در مرجع [۳۰] در سال ۲۰۲۲ ارائه شد و تی.کی.دی.ای[۲۳] نامیده می‌شود و آخرین روش مورد استفاده، در مرجع [۹] ارائه شد و UMG-NC (ریز تجمع تک متغیره برای بی‌نام‌سازی گراف-مرکزیت همسایگی[۲۴]) نامیده می‌شود. برای ارزیابی TSRAM، سه مجموعه داده واقعی استفاده می‌شود. از آنجا که اندازه‌گیری معیارهای ارزیابی برای گراف‌های بزرگ ممکن نیست، از گراف‌های با سایز متوسط و کوچک برای ارزیابی الگوریتم‌ها استفاده می‌نماییم. همان‌طور که قبلا نیز مطرح گردید، با توجه به اینکه هدف این الگوریتم، گمنام‌سازی دنباله درجه گراف است، این گراف‌ها غیرجهت‌دار و غیربرچسب‌دار هستند. این مجموعه داده‌ها شامل دلفینز[۲۵] [۳۴]، کا.جی.آر.کیو.سی[۲۶] [۳۵]، و کا.هپ.تی.اچ[۲۷] [۳۵] است. دلفینز[۲۸] یک مجموعه داده غیرجهت‌دار است که شامل ارتباط بین ۶۲ دلفین در یک جامعه است. CA-GrQc و ca-HepTh، شبکه‌های همکاری علمی بین مقاله‌های ارائه شده است. خصوصیات این مجموعه داده‌ها در جدول (۲) نشان داده شده است. با توجه به اینکه کارایی الگوریتم گمنام‌سازی از نظر زمان اجرا در زمان اعمال بر روی گراف‌های بزرگ مشخص می‌شود، به منظور ارزیابی الگوریتم پیشنهادی در این مقاله از نظر زمان اجرا، سه مجموعه

داده واقعی و بزرگ استفاده می‌شود که در جدول (۳) نشان داده شده‌اند. به دلیل اینکه هدف الگوریتم UMGA و الگوریتم‌های پیشنهادی تنها گمنام‌سازی دنباله درجه گره‌های گراف است و دیگر ویژگی‌های گراف در نظر گرفته نمی‌شود، از گراف‌های غیرجهت‌دار و بدون برچسب برای ارزیابی این الگوریتم استفاده می‌شود. ایمیل-انرون[۲۹] یک شبکه است که ارتباطات را در یک مجموعه داده با حدود نیم میلیون ایمیل پوشش می‌دهد [۳۶]. گره‌های داخل این شبکه، آدرس‌های ایمیل هستند و یال‌ها ارسال و دریافت ایمیل را نشان می‌دهند. دی.بی.ال.پی[۳۰] (DBLP) یک مجموعه داده است که فهرستی از مقالات پژوهشی در علوم رایانه را فراهم می‌کند [۳۷]. یوتیوپ[۳۱]، یک وب سایت اشتراک‌گذاری ویدیو است و شامل روابط دوستی در یک شبکه اجتماعی است [۲۴]. همچنین، از معیارهای عمومی و ساختاری گراف برای ارزیابی tsRANK استفاده می‌شود. معیارهای عمومی گراف، شامل میانگین طول کوتاه‌ترین مسیرها[۳۲] (APL) [۱۲]، انتقال‌پذیری[۳۳] (T) [۱۲]، میانگین ضریب خوشه‌بندی[۳۴] (ACC) [۱۲]، معیار پیمانه‌ای[۳۵] [۱۲] و اختلاف بین گراف اصلی و گراف گمنام شده از نظر تعداد یال‌ها است که در این مقاله، برای نشان دادن این معیارها از کلمات مخفف داخل پرنتز برای هر کدام از معیارها استفاده می‌شود. همچنین، از رابطه (۱۳) برای نشان دادن فاصله بین مقادیر ارزیابی شده برای گراف اصلی و گمنام‌شده استفاده می‌گردد [۱۲]. در این معادله، ATTRIB یک معیار مورد ارزیابی است.

$$\text{Error} = \frac{1}{\text{num}} \sum_{i=1}^{\text{num}} \left| \text{ATTRIB}_i - \text{ATTRIB} \right| \qquad (۱۳)$$

که در آن، num تعداد مقادیر یک ویژگی برای یک مقدار k است، ATTRIBi مقدار یک ویژگی گراف برای یک مقدار k است و ATTRIB مقدار معیار مورد ارزیابی برای گراف اصلی است. معیارهای ساختاری، معیارهای دیگری برای ارزیابی الگوریتم‌های پیشنهادی هستند. معیارهای خوشه‌بندی در نظر گرفته شده در این بخش عبارتند از: واریانس اطلاعات[۳۶] (VI) [۳۸]، اطلاعات متقابل نرمال[۳۷] (NMI) [۳۸]، فاصله انفصال-اتصال[۳۸] [۳۹]، اندیس تصادفی[۳۹] [۴۰] و اندیس تصادفی تعدیل‌شده [۴۰][۴۱]. این معیارها ساختار خوشه‌بندی دو گراف را با هم مقایسه می‌کند. هر چقدر مقدار واریانس اطلاعات کمتر باشد، ساختار خوشه‌ای گراف به گراف اصلی نزدیک‌تراست. شاخص فاصله انفصال-اتصال، تعداد گره‌هایی را مشخص می‌کند که باید از یک خوشه منتقل شوند تا خوشه دیگر را ایجاد کند. مقدار کمتر انفصال-اتصال نشان‌دهنده نزدیکی ساختار گراف گمنام به گراف اصلی است. بنابراین، مقدار ایده‌آل فاصله انفصال-اتصال نیز ۰ است. مقدار ۰ برای اطلاعات متقابل نرمال، اندیس تصادفی و اندیس تصادفی تعدیل‌شده نشان می‌دهد که گره‌های گراف به صورت تصادفی در ساختار خوشه‌بندی گراف قرار دارند. در بهترین حالت ساختار خوشه‌بندی، این مقادیر برابر با ۱ و ساختار دو گراف یکسان هستند. به منظور تسهیل درک نتایج، از معکوس اطلاعات متقابل نرمال، اندیس تصادفی و اندیس تصادفی تعدیل‌شده استفاده می‌شود. بنابراین، هر چقدر مقادیر این معیارها کم‌تر باشد، گراف گمنام‌شده سودمندی بیشتری دارد. با توجه به این‌که VA و KDVEM در بعضی مواقع، تعداد رئوس گراف را افزایش می‌دهند، بردارهای عضویت گراف‌های گمنام و اصلی دارای طول‌های مختلف هستند. بنابراین، این دو بردار را نمی‌توان مقایسه کرد. درنتیجه، در این موارد، الگوریتم‌ها نادیده گرفته می‌شوند. از آنجا که UMGA، الگوریتمی است که اخیرا در زمینه گمنام‌سازی دنباله درجه گراف ارائه شده است و در این مقاله، بهبود داده شده است، کافی است که الگوریتم پیشنهادی با آن مقایسه شود.

۱-۶- نتایج ارزیابی زمان اجرای تی.اس.رنک

نتایج ارزیابی الگوریتم ارائه شده در این مقاله در جدول (۴) دیده می‌شود. سطر آخر این جدول، مقدار میانگین زمان اجرا برای این الگوریتم‌ها است. همان‌طور که در سطر آخر این جدول دیده می‌شود، به‌طور میانگین، زمان اجرای الگوریتم tsRANK نسبت به UMGA خیلی کمتر است. کاهش زمان اجرا در این الگوریتم، نتیجه استفاده از ساختار درختی فشرده در مرحله اول الگوریتم و همچنین کاهش تعداد پویش یال‌های گراف برای انتخاب بهترین یال‌ها به منظور اضافه یا حذف شدن است. این نتایج نشان می‌دهد که روش‌های ارائه شده در این مقاله برای کاهش زمان اجرای الگوریتم UMGA موثر است.

۲-۶- نتایج ارزیابی تی.اس.رنک از نظر معیارهای عمومی گراف

نتایج ارزیابی الگوریتم پیشنهادی بر روی مجموعه داده دلفینز در شکل (۱) مشاهده می‌شود. مقادیر خطای نشان داده شده در این شکل، میانگین مقادیر به‌دست آمده برای مقادیر مختلف k است. همان‌طور که در این شکل‌ها دیده می‌شود، به‌طور کلی، الگوریتم tsRANK بر روی این مجموعه داده، مقدار خطای کمتری از سایر الگوریتم‌ها دارد. این نتایج نشان از موثر بودن روش‌های ارائه شده برای بهبود مرحله اول و دوم الگوریتم گمنام‌سازی دنباله درجه گراف دارد. استفاده از پارامترهای NC و شماره هسته گره‌ها در هنگام انتخاب یال برای حذف و یا اضافه شدن یال، منجر به حفظ ویژگی‌های گراف شبکه اجتماعی می‌شود. این معیارها یال‌هایی را برای حذف و اضافه شدن انتخاب می‌کنند که تاثیر کمی در ساختار گراف داشته باشند و درنتیجه در اثر ویرایش گراف سودمندی آن حفظ گردد.

Table (2): Attributes of the Benchmark Data Sets.

جدول (۲): ویژگی‌های مجموعه داده‌ها

مشخصات / مجموعه داده‌ها	گره‌ها	لبه‌ها
دلفینز [۳۴]	۶۲	۱۵۹
کا.جی.آر.کیو.سی [۳۵]	۵۲۴۲	۱۴۴۸۴
کا.هپ.تی.اچ [۳۵]	۹۸۷۷	۲۵۹۹۸

Table (3): Attributes of the benchmark data sets

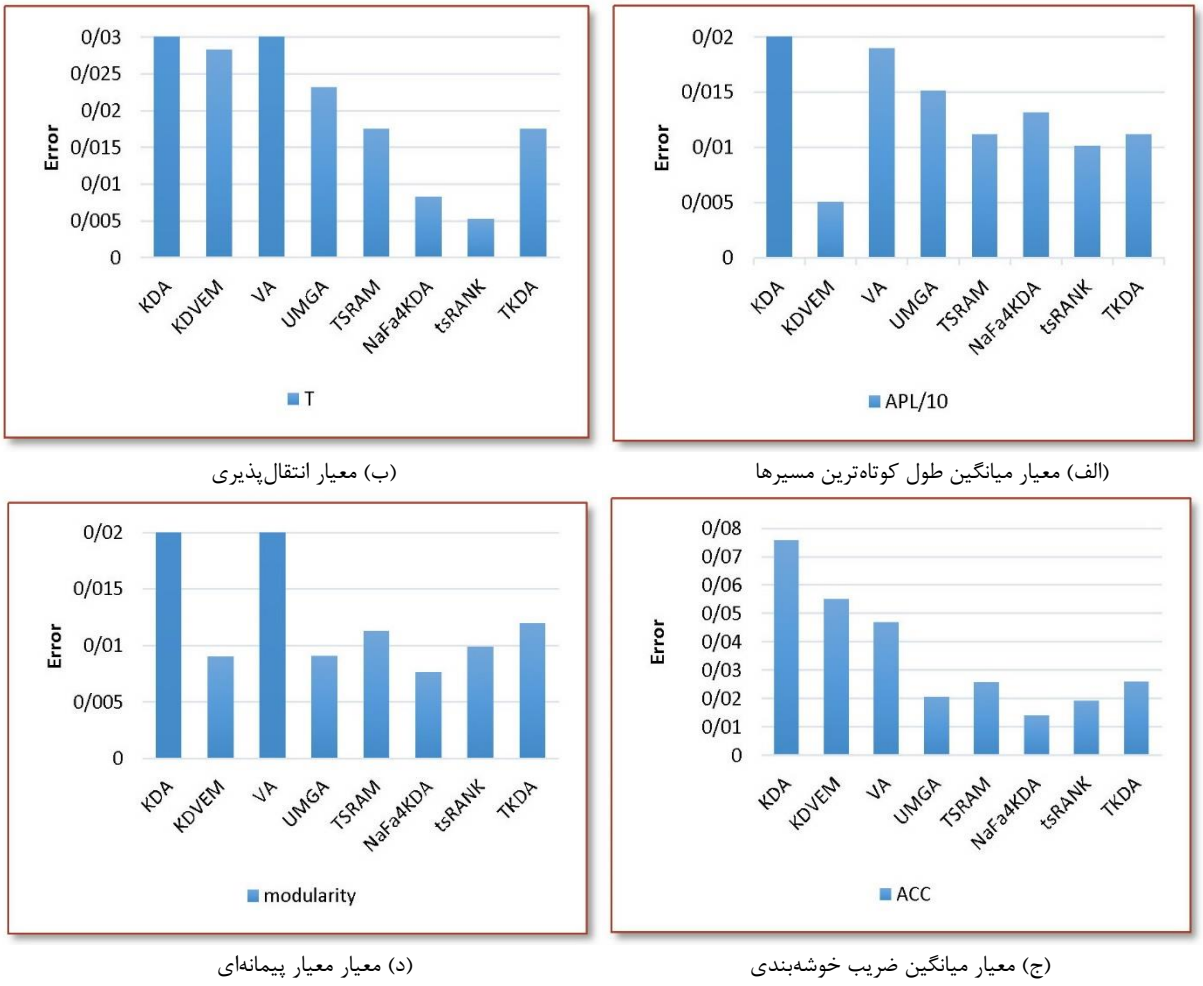
جدول (۳): ویژگی‌های مجموعه داده‌ها

مشخصات / مجموعه داده‌ها	گره‌ها	لبه‌ها
ایمیل-انرون [۳۹]	۳۶۶۹۲	۱۸۳۸۳۱
دی.بی.ال.پی [۴۰]	۰۸۰۳۱۷	۸۶۶۰۴۹۱
یوتیوب [۲۴]	۸۹۰۱۳۴۱	۶۲۴۹۸۷۲

Table (4): Comparison of tsRANK and UMGA with Respect to Runtime.

جدول (۴): مقایسه دو الگوریتم تی.اس.رنک و یو.ام.جی.آ از نظر زمان اجرا

مجوعه داده	k	UMGA (ثانیه، دقیقه، ساعت)	TSRAM (ثانیه، دقیقه، ساعت)	NaFa4KDA (ثانیه، دقیقه، ساعت)	TKDA (ثانیه، دقیقه، ساعت	tsRANK (ثانیه، دقیقه، ساعت)
ایمیل-انرون	۲۵	۲۰:۰۴:۰۰	۱۳:۰۱:۳۵	۱۰:۰۰:۴۰	۱۳:۰۱:۳۸	۱۹:۰۰:۰۰
	۵۰	۳۶:۰۴:۰۰	۳۱:۰۱:۵۶	۲۰:۰۰:۴۸	۳۱:۰۱:۵۸	۲۱:۰۰:۰۰
	۱۰۰	۴۱:۰۵:۰۰	۳۲:۰۳:۰۲	۲۲:۰۰:۵۹	۳۲:۰۳:۰۵	۳۰:۰۰:۰۰
	۲۰۰	۵۰:۱۵:۰۰	۴۵:۱۰:۱۵	۴۰:۰۱:۲۶	۴۵:۱۰:۲۰	۳۱:۰۰:۰۰
	۵۰۰	۵۶:۱۷:۰۰	۵۲:۱۵:۲۳	۴۵:۰۱:۵۴	۵۲:۱۵:۳۰	۴۶:۰۰:۰۰
دی.بی.ال.پی	۲۵	۱۲:۰۰:۰۰	۱۰:۰۰:۰۸	۰۸:۰۰:۰۶	۱۰:۰۰:۱۸	۱۰:۰۰:۰۰
	۵۰	۱۸:۰۰:۰۰	۱۳:۰۰:۱۷	۰۹:۰۰:۱۵	۱۳:۰۰۱۵	۱۲:۰۰:۰۰
	۱۰۰	۴۱:۰۰:۰۰	۳۶:۰۰:۳۶	۲۳:۰۰:۲۰	۳۶:۰۰:۲۰	۱۸:۰۰:۰۰
	۲۰۰	۴۲:۰۱:۰۰	۴۰:۰۰:۴۳	۳۷:۰۰:۳۲	۴۰:۰۰:۳۷	۳۱:۰۰:۰۰
	۵۰۰	۵۳:۰۷:۰۰	۵۰:۰۶:۰۵	۴۶:۰۱:۵۸	۵۰:۰۶:۱۰	۵۶:۰۰:۰۰
یوتیوب	۲۵	۱۳:۱۲:۰۴	۱۰:۰۵:۰۱	۰۸:۱۸:۴۲	۱۰:۰۵:۱۱	۱۴:۱۰:۰۰
	۵۰	۲۰:۳۰:۰۵	۱۴:۰۰:۳۷	۱۰:۴۷:۰۶	۱۴:۰۰:۴۰	۵۶:۱۰:۰۰
	۱۰۰	۳۴:۰۱:۰۶	۳۲:۰۱:۵۴	۲۸:۱۷:۱۸	۳۲:۰۲:۰۰	۱۲:۴۵:۰۰
	۲۰۰	۱۲:۳۰:۱۰	۱۰:۴۶:۰۳	۰۸:۰۲:۰۵	۱۰:۴۶:۰۶	۰۶:۳۸:۰۱
میانگین		۲۹:۱۰:۰۲	۲۴:۰۷:۲۳	۲۲:۰۹:۵۱	۲۴:۰۷:۲۲	**۱۰:۰۷:۰۰**



(ب) معیار انتقال‌پذیری

(الف) معیار میانگین طول کوتاه‌ترین مسیرها

(د) معیار معیار پیمانه‌ای

(ج) معیار میانگین ضریب خوشه‌بندی

شکل (۱): مقدار متوسط خطا در مجموعه داده دلفینز به ازای مقادیر مختلف k برای معیارهای مختلف

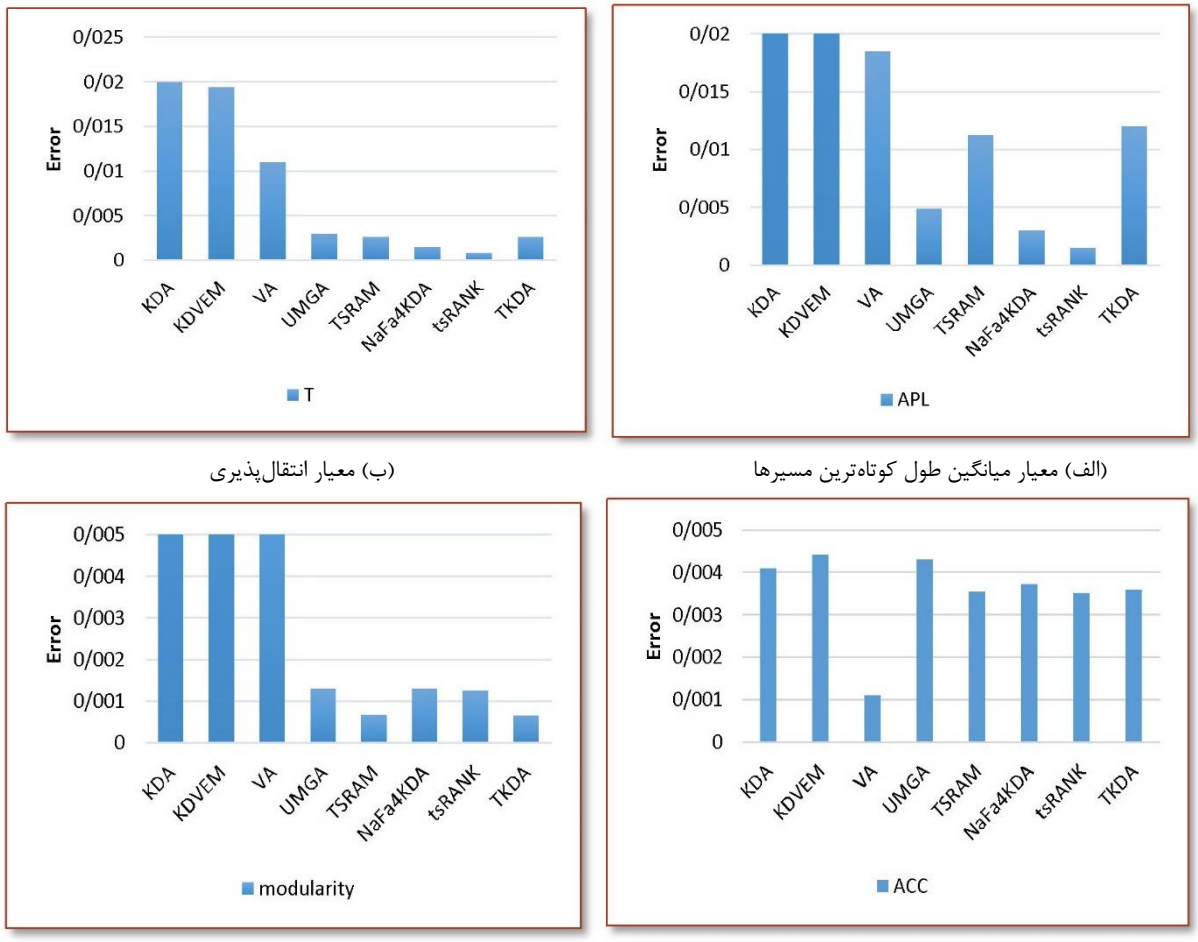
Figure (1): Average value of error in the Dolphins dataset for different values of k for criteria, a) APL, b) T, c) ACC, d) Modularity

همچنین در مقایسه الگوریتم tsRANK با دو الگوریتم TSRAM و NaFa4KDA، در همه موارد این الگوریتم از TSRAM بهتر است که دلیل آن، حفظ بیشتر ساختار گراف شبکه اجتماعی توسط ترکیب این الگوریتم است و می‌توان نتیجه گرفت که ترکیب این دو الگوریتم بهتر از تنها استفاده از الگوریتم TSRAM است. در مقایسه الگوریتم tsRANK با الگوریتم NaFa4kda نیز این الگوریتم در اکثر مواقع بهبود بیشتری داشته است. بنابراین، به‌طور کلی استفاده از ترکیب دو الگوریتم TSRAM و NaFa4KDA نسبت به استفاده تنها از یکی از این دو الگوریتم منجر به بهبود بیشتر ویژگی های عمومی گراف می‌شود. نتایج ارزیابی الگوریتم پیشنهادی بر روی مجموعه داده CA-HepTh در شکل (۲) نشان داده شده است. مقادیر خطای نشان داده شده در این شکل‌ها، میانگین مقادیر به‌دست آمده برای مقادیر مختلف k است. نتایج ارزیابی نشان می‌دهد که الگوریتم پیشنهادی، نسبت به UMGA خطای بیشتری در مورد معیار APL دارد که دلیل آن، اتصال رئوس با درجه بالا به یکدیگر است. از آنجاکه به گره‌هایی با درجه بالا گره‌های زیادی متصل است، در اثر برقراری یک یال بین آن‌ها تعداد زیادی مسیر بین گره‌های گراف اضافه می‌گردد و این امر منجر به کاهش مقدار APL می‌گردد. این کاهش در مقدار این معیار خطا را در مورد آن افزایش می‌دهد. با این‌حال در مقایسه با سایر الگوریتم‌ها، خطای APL برای الگوریتم پیشنهادی کمتر است. در مورد سایر معیارها نیز الگوریتم پیشنهادی نسبت به UMGA دارای کاهش خطا است. بنابراین، به‌طور کلی، الگوریتم پیشنهادی بر روی این مجموعه داده می‌تواند ویژگی‌های عمومی گراف را نسبت به الگوریتم‌های موجود بیشتر حفظ نماید. در مقایسه بین الگوریتم پیشنهادی با الگوریتم‌های TSRAM و NaFa4KDA همان‌طور که در شکل‌های زیر دیده می‌شود، به‌طور کلی، ترکیب این دو الگوریتم منجر به بهبود ویژگی‌های عمومی گراف گمنام شده است. بنابراین، استفاده از ترکیب این دو الگوریتم برای گمنام‌سازی گراف شبکه اجتماعی نسبت به

استفاده از هر کدام از این الگوریتم‌ها ویژگی‌های عمومی گراف را بهبود می‌دهد. نتایج ارزیابی ترکیب دو الگوریتم TSRAM و NaFa4KDA بر روی مجموعه داده CA-GrQc در شکل (۳) نشان داده شده است. مقادیر خطای نشان داده شده در این شکل، میانگین مقادیر به‌دست آمده برای مقادیر مختلف k است. همان‌طور که در این شکل‌ها دیده می‌شود، در مورد اکثر معیارهای عمومی گراف، الگوریتم پیشنهادی خطای کمتری نسبت به UMGA دارد. تنها معیاری که در آن، الگوریتم‌های پیشنهادی بهینه نیست، ACC است. همان‌طور که قبلا توضیح داده شد، این نتیجه، حاصل استفاده از شماره جامعیت گره‌ها در گروه‌بندی درجه گره‌ها است. کاهش خطای نتایج ارزیابی در الگوریتم پیشنهادی حاصل استفاده همزمان از معیارهای شماره هسته گره‌ها و سطح مشارکت یال و الگوریتم بهینه‌سازی NaFa در مرحله ویرایش گراف است. شکل (۴) دید کلی نسبت به نتایج ارزیابی ارائه شده در این بخش را فراهم می‌کند. همان‌طور که در این شکل دیده می‌شود، در همه موارد به جز ACC، الگوریتم پیشنهادی نسبت به سایر الگوریتم‌ها خطای کمتری دارد. این نتایج بیانگر موثر بودن روش ارائه شده برای ویرایش گراف است. معیارهای معرفی شده برای انتخاب بهترین یال‌ها به منظور حذف و یا اضافه شدن توانسته‌اند ویژگی‌های عمومی گراف را حفظ نمایند.

۳-۶- نتایج الگوریتم تی.اس.رنک از نظر معیارهای ساختاری

نتایج ارزیابی الگوریتم tsRANK از نظر حفظ ساختار جامعه بر روی مجموعه داده دلفینز در شکل (۵) مشاهده می‌شود. این نتایج، در اکثر موارد کاهش مقدار خطا را در الگوریتم پیشنهادی نسبت به سایر الگوریتم‌ها نشان می‌دهد. مقادیر خطای نشان داده شده در این شکل، میانگین مقادیر به‌دست آمده برای مقادیر مختلف k است.



(ب) معیار انتقال‌پذیری

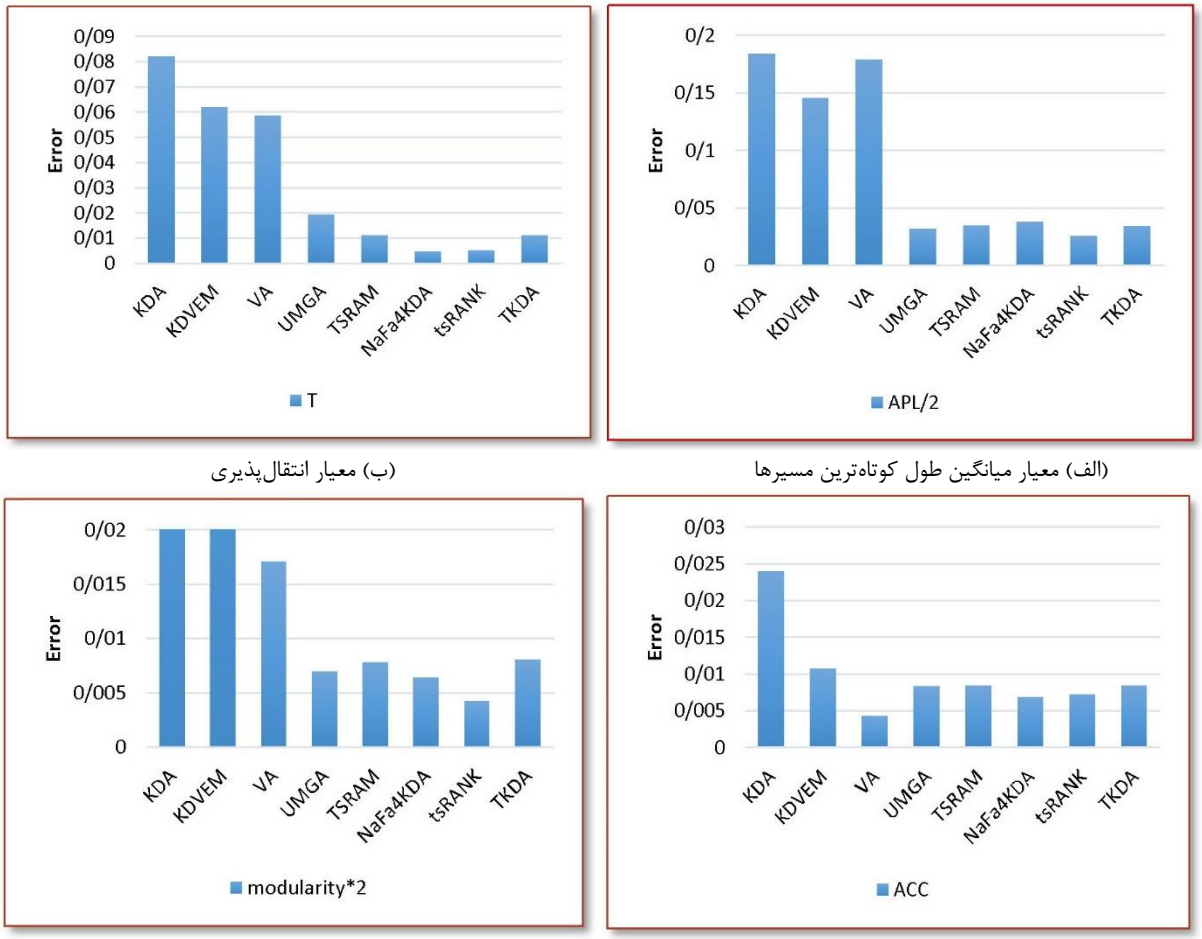
(الف) معیار میانگین طول کوتاه‌ترین مسیرها

(د) معیار معیار پیمانه‌ای

(ج) معیار میانگین ضریب خوشه‌بندی

شکل (۲): مقدار متوسط خطا در مجموعه داده کا.هپ.تی.اچ به ازای مقادیر مختلف k برای معیارهای مختلف

Figure (2): Average value of error in the CA-HepTh dataset for different values of k for criteria, a) APL, b) T, c) ACC, d) modularity



(الف) معیار میانگین طول کوتاه‌ترین مسیرها

(ب) معیار انتقال‌پذیری

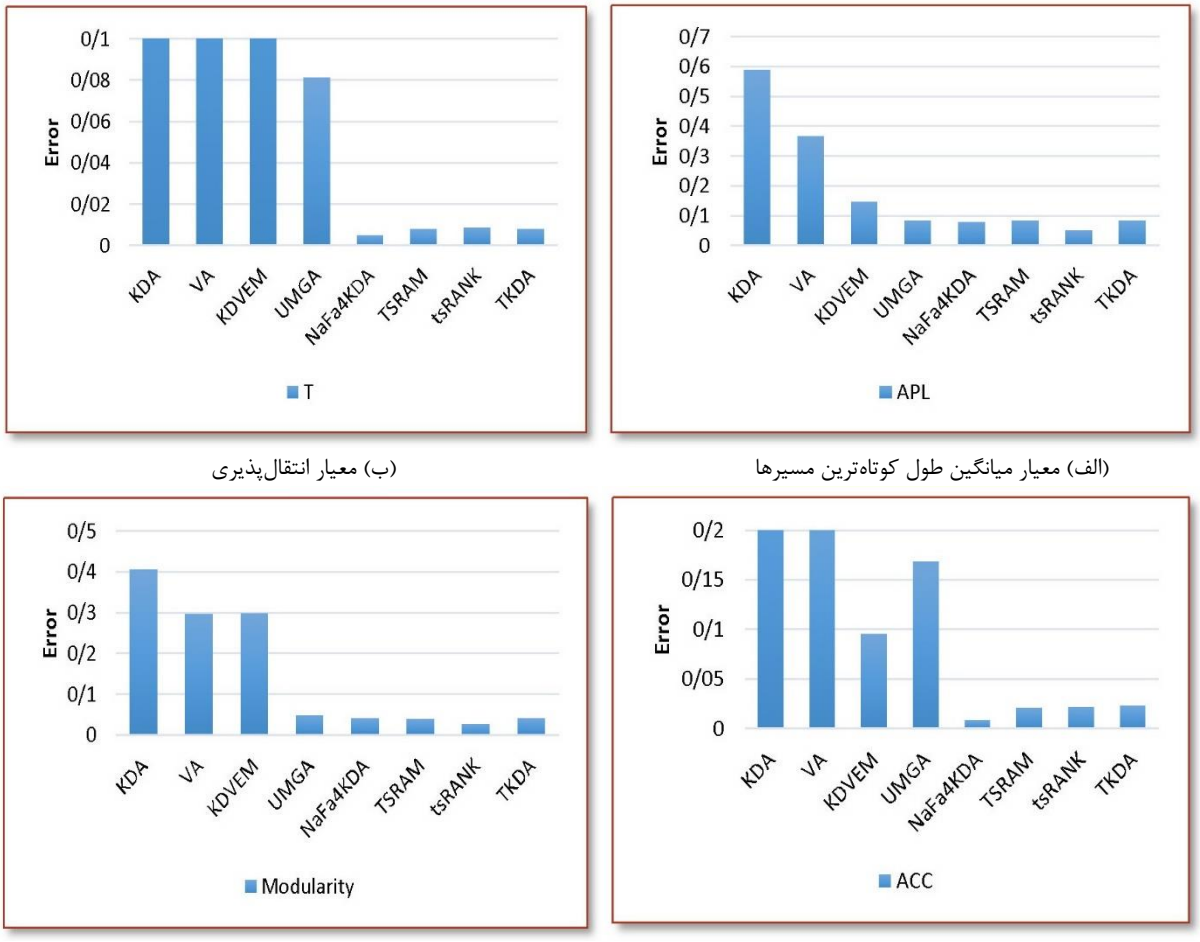
(ج) معیار میانگین ضریب خوشه‌بندی

(د) معیار معیار پیمانه‌ای

شکل (۳): مقدار متوسط خطا در مجموعه داده کا.جی.آر.کیو.سی به ازای مقادیر مختلف k برای معیارهای مختلف

Figure (3): Average value of error in the CA-GrQc dataset for different values of k for criteria, a) APL, b) T, c) ACC, d) Modularity

ترکیب معیارهای استفاده شده برای انتخاب بهترین یال‌ها و همچنین استفاده از الگوریتم NaFa برای یافتن بهترین یال‌های قابل افزایش به گراف منجر به گرافی می‌شود که از نظر ساختاری به گراف اصلی نزدیک‌تر است. نتایج ارزیابی الگوریتم NaFa4KDA بر روی مجموعه داده CA-HepTh در شکل (۶) مشاهده می‌شود. مقادیر خطای نشان داده شده در این شکل، میانگین مقادیر به‌دست آمده برای مقادیر مختلف k است. نتایج به‌دست آمده از ارزیابی الگوریتم پیشنهادی بر روی این مجموعه داده نیز مشابه نتایج ارزیابی بر روی مجموعه داده دلفینز است. در این نتایج نیز کاهش مقدار خطا در معیارهای خوشه‌بندی مشاهده می‌شود که نشان از بهبود الگوریتم پیشنهادی نسبت به سایر الگوریتم‌ها دارد. این کاهش خطا در معیارهای خوشه‌بندی نتیجه تغییرات کم در ساختار گراف شبکه اجتماعی توسط استفاده از معیارهای پیشنهادی در این مقاله است. همان‌طور که قبلا توضیح داده شد، دو معیار پیشنهادی، شماره هسته گره‌های انتهایی و سطح مشارکت یال هستند که ارزش یال‌های گراف را مورد سنجش قرار می‌دهند. از آنجا که یال‌هایی برای حذف و اضافه شدن انتخاب می‌شوند که تاثیر کمی در ساختار گراف دارند، سودمندی گراف شبکه اجتماعی حفظ می‌شود. نتایج ارزیابی الگوریتم پیشنهادی در این مقاله بر روی مجموعه داده CA-GrQc در شکل (۷) نشان داده شده است. مقادیر خطای نشان داده شده در این شکل، میانگین مقادیر بدست آمده برای مقادیر مختلف k است. همان‌طور که در این شکل دیده می‌شود، در مقایسه با الگوریتم UMGA الگوریتم tsRANK در حفظ ساختار گراف بهتر عمل می‌کند. اگر این الگوریتم‌ها با سایر الگوریتم‌ها مقایسه شود نیز در اکثر مواقع، الگوریتم پیشنهادی در این مقاله بهتر می‌تواند ویژگی‌های خوشه‌بندی گراف شبکه اجتماعی را حفظ کند. به منظور نتیجه‌گیری کلی از نتایج ارزیابی به‌دست آمده در این بخش، از میانگین مقادیر نرمال شده به ازای هر کدام از معیارهای خوشه‌بندی استفاده می‌شود.



(الف) معیار میانگین طول کوتاه‌ترین مسیرها

(ب) معیار انتقال‌پذیری

(ج) معیار میانگین ضریب خوشه‌بندی

(د) معیار معیار پیمانه‌ای

شکل (۴): میانگین مقادیر خطای معیارهای عمومی بر روی همه مجموعه داده‌ها

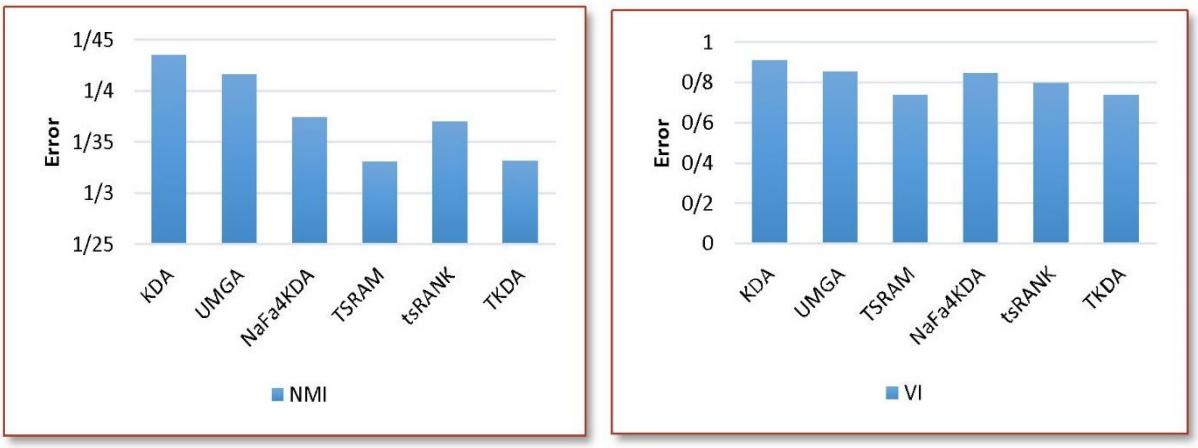
Figure (4): Average error values of general criteria on all data sets for criteria, a) APL, b) T, c) ACC, d) Modularity

شکل (۸) میانگین مقادیر نرمال شده را نشان می‌دهد. این نتایج بهبود الگوریتم پیشنهادی در این مقاله را نسبت به الگوریتم UMGA از نظر حفظ ویژگی‌های خوشه‌بندی گراف نشان می‌دهد و گراف گمنام از نظر ساختاری به گراف اصلی نزدیک‌تر است. به‌طور کلی، نتایج ارائه شده در این بخش نشان می‌دهد که الگوریتم tsRANK پیشنهادی علاوه بر این‌که ویژگی‌های عمومی گراف را حفظ می‌کند، گراف گمنامی ایجاد می‌کند که از نظر ویژگی‌های عمومی و ساختاری نیز به گراف اصلی شباهت بیشتری دارد.

۷- نتیجه‌گیری

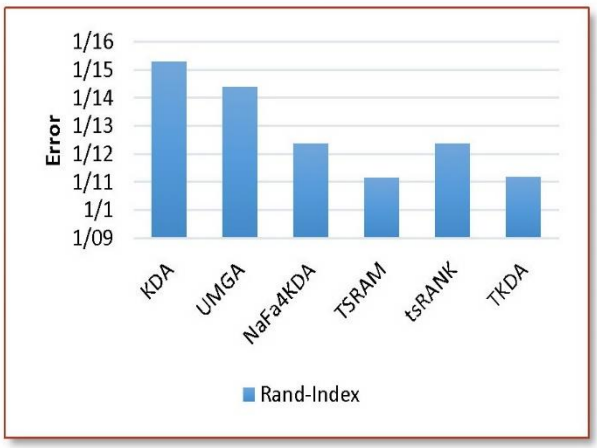
روش پیشنهادی در این مقاله به‌طور همزمان زمان اجرا، میزان مصرف حافظه و سودمندی گراف شبکه اجتماعی را نسبت به الگوریتم‌های موجود بهبود می‌دهد. در این روش برای هر کدام از مرحله‌ها فرایند گمنام‌سازی گراف از دو الگوریتم TSRAM و Nafa4KDA استفاده شده که مزایای آن‌ها می‌تواند معایب هم‌دیگر را خنثی نماید و به بهبود کلی فرایند گمنام‌سازی بینجامد. در مرحله اول فرایند، الگوریتم TSRAM زمان اجرا و میزان مصرف حافظه را کاهش می‌دهد در حالی‌که در مرحله دوم، الگوریتم Nafa4KDA زمان اجرا و سودمندی گراف گمنام را بهبود می‌دهد. از آنجا که استفاده از TSRAM تا حدود کمی سودمندی گراف گمنام را کاهش می‌دهد، با استفاده از الگوریتم Nafa4KDA در مرحله دوم فرایند، این اشکال جبران شده و منجر به افزایش سودمندی گراف نیز می‌گردد. نتایج ارزیابی انجام شده بر روی مجموعه داده‌های مختلف نشان می‌دهد که ترکیب این

دو الگوریتم نسبت به استفاده از تک تک هر کدام از این الگوریتم‌ها به‌طور مجزا، کل فرایند گمنام‌سازی گراف شبکه اجتماعی را از نظر زمان اجرا، میزان مصرف حافظه و سودمندی گراف بهبود می‌دهد.

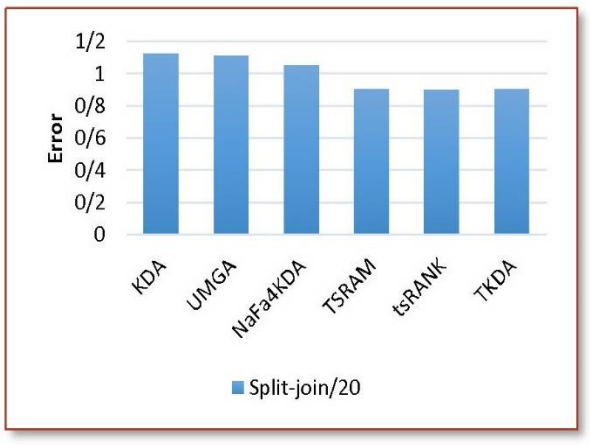


(ب) معیار اطلاعات متقابل نرمال

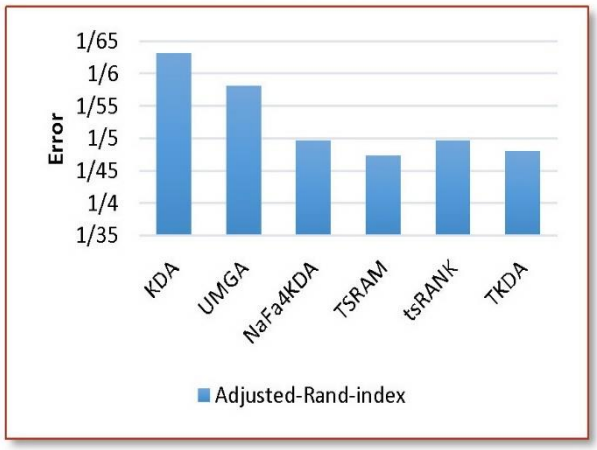
(الف) معیار واریانس اطلاعات



(د) معیار اندیس تصادفی



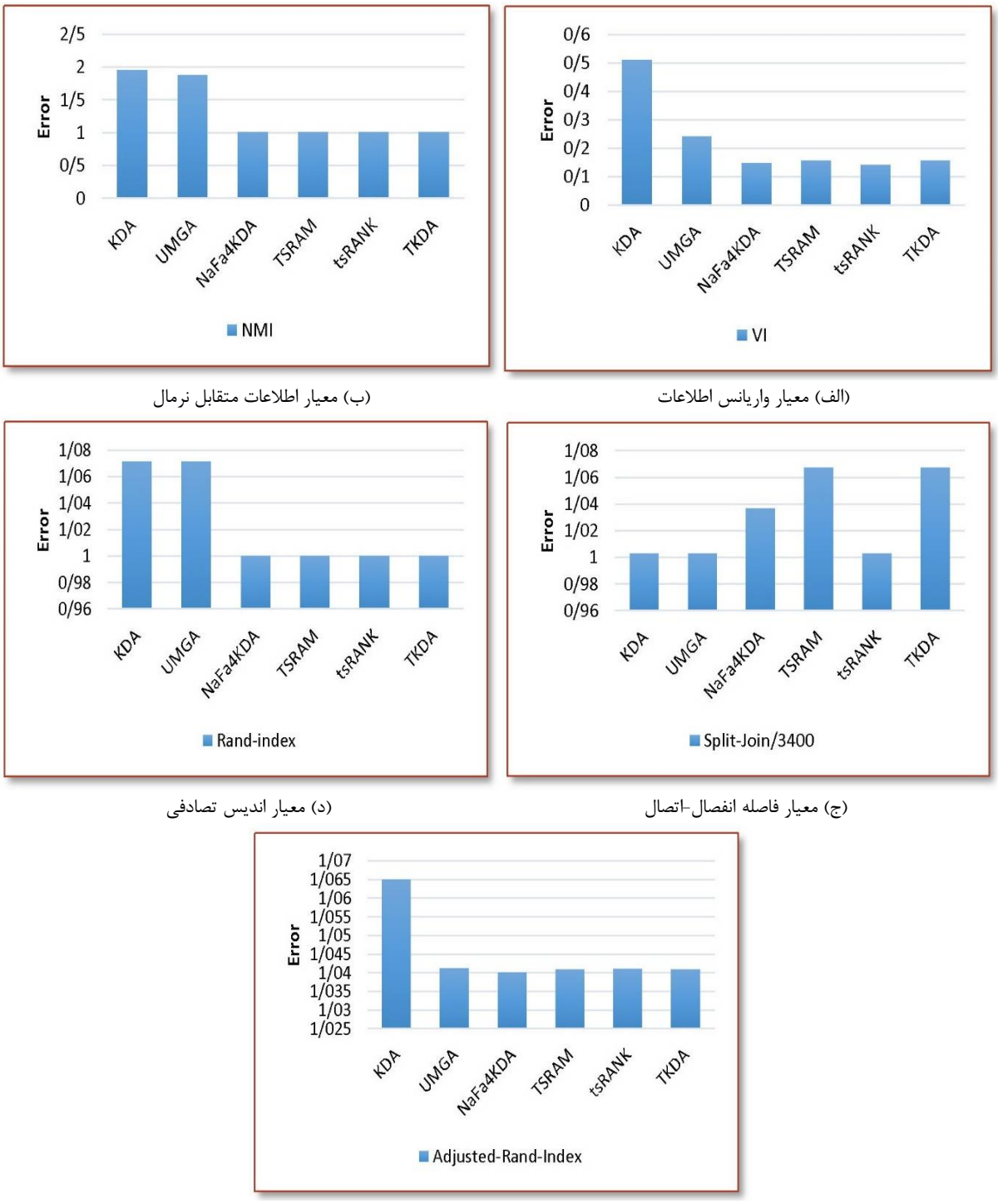
(ج) معیار فاصله انفصال-اتصال



(ه) معیار اندیس تصادفی تعدیل شده

شکل (۵): مقدار متوسط خطا در مجموعه داده دلفینز به ازای معیارهای مختلف

Figure (5): Average value of error in Dolphins data set for different values of k for criteria, a) VI, b) NMI, c) split-join, d) Rand-index, e) Adjusted-Rand-Index



(الف) معیار واریانس اطلاعات

(ب) معیار اطلاعات متقابل نرمال

(ج) معیار فاصله انفصال-اتصال

(د) معیار اندیس تصادفی

(ه) معیار اندیس تصادفی تعدیل شده

شکل (۶): مقدار متوسط خطا در مجموعه کا.هپ.تی.اچ به ازای معیارهای مختلف

Figure (6): Average value of error in CA-HepTh data set for different values of k for criteria, a) VI, b) NMI, c) split-join, d) Rand-index, e) Adjusted-Rand-Index

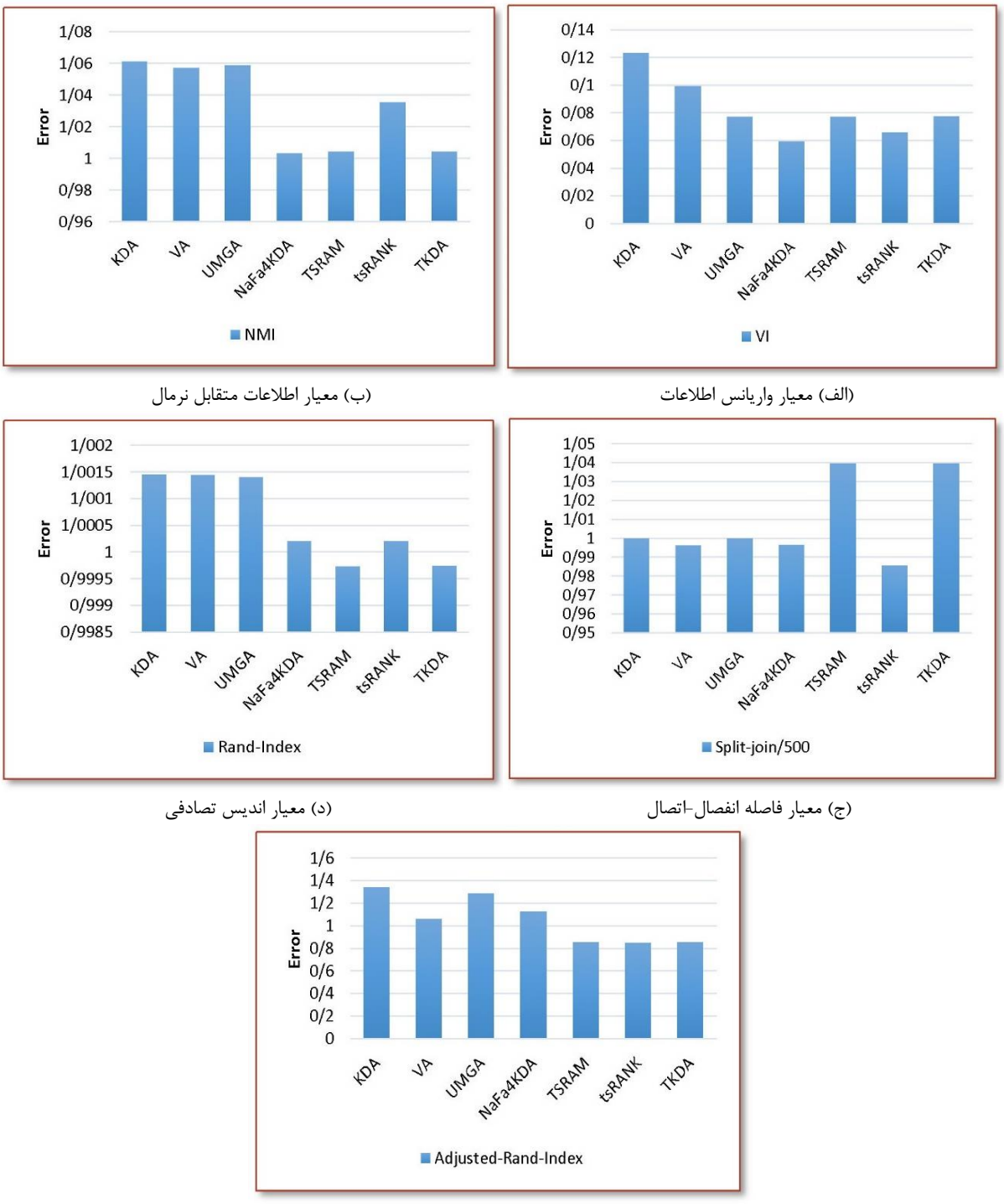
References

مراجع

[1] D. Yang, B. Qu, "Privacy-preserving social media data publishing for personalized ranking-based recommendation", IEEE Trans. on Knowledge and Data Engineering, vol. 31, no. 3, pp. 507-520, March 2019 (doi: 10.1109/TKDE.2018.2840974).

[2] F. Ferri, P. Grifoni, T. Guzzo, "New forms of social and professional digital relationships: the case of

Facebook", Social Network Analysis and Mining, vol. 2, no. 2, pp. 121–137, June 2012 (doi: 10.1007/s13278-011-0038-4).
[3] Q. Wang, C. Zeng, W. Zhou, T. Li, S.S. Iyengar, L. Shwartz, G.Y. Grabarnik, "Online interactive collaborative filtering using multi-armed bandit with dependent arms", IEEE Trans. on Knowledge and Data Engineering, vol. 31, no. 8, pp. 1569–1580, Aug. 2019 (doi: 10.1109/TKDE.2018.2866041).
[4] L. Sweeney, "k-anonymity: A model for protecting privacy", International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 05, pp. 557–570, Oct. 2002 (doi: 10.1142/S0218488502001648).



(ب) معیار اطلاعات متقابل نرمال

(الف) معیار واریانس اطلاعات

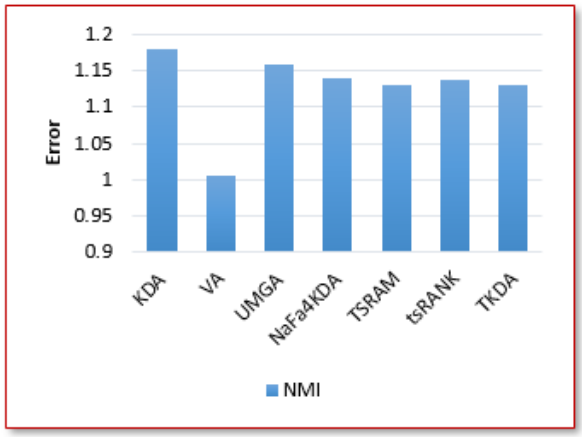
(د) معیار اندیس تصادفی

(ج) معیار فاصله انفصال-اتصال

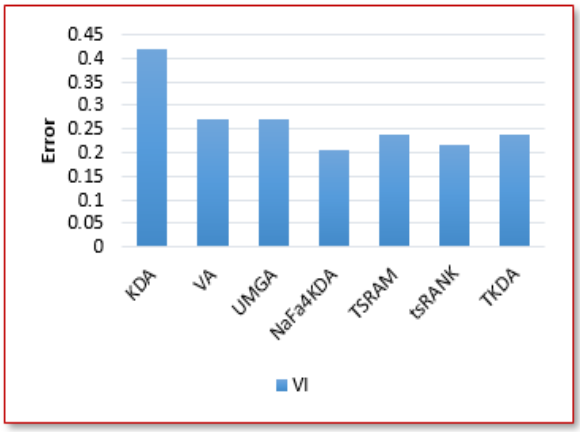
(ه) معیار اندیس تصادفی تعدیل شده

شکل (۷): مقدار متوسط خطا در مجموعه داده کا.جی.آر.کیو.سی ازای مقادیر مختلف

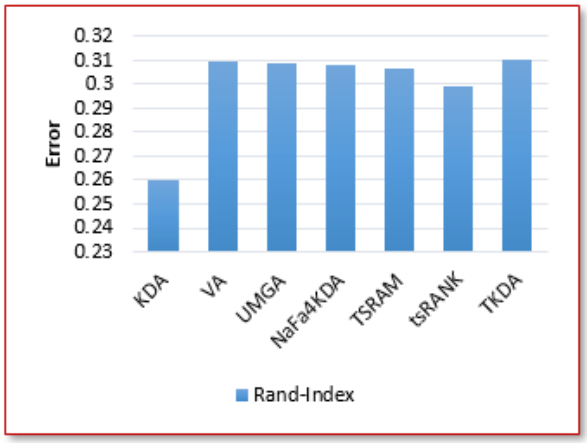
Figure (7): Average value of error in CA-GrQc data set for different values of k for criteria, a) VI, b) NMI, c) split-join, d) Rand-index, e) Adjusted-Rand-Index



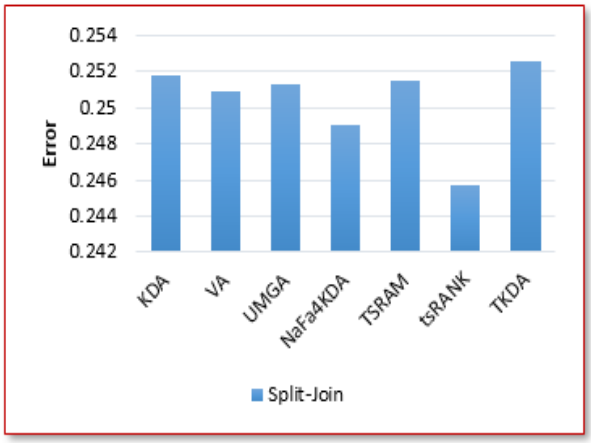
(ب) معیار اطلاعات متقابل نرمال



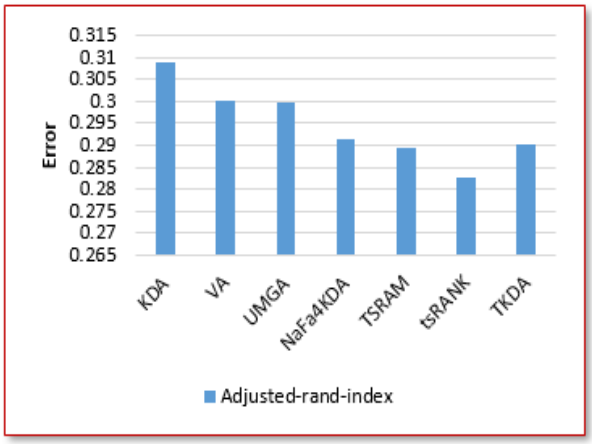
(الف) معیار واریانس اطلاعات



(د) معیار اندیس تصادفی



(ج) معیار فاصله انفصال-اتصال



(ه) معیار اندیس تصادفی تعدیل شده

شکل (۸): میانگین مقادیر خطای معیارهای ساختاری بر روی همه مجموعه داده‌ها برای معیارهای مختلف

Figure (8:) Average error values of structural criteria on all data sets for criteria, a) VI, b) NMI, c) Split-join, d) Rand-index, e) Adjusted-rand-index

[5] A. Machanavajjhala, J. Gehrke, D. Kifer, M. Venkitasubramaniam, "L-diversity: Privacy beyond k-anonymity", Proceeding of the IEEE/ICDE, pp. 1-12, Atlanta, GA, USA, April 2006 (doi: 10.1109/ICDE.2006.1).

[6] N. Li, T. Li, S. Venkatasubramania, "t-Closeness: Privacy beyond k-anonymity and -diversity", Proceeding of the IEEE/ICFE, pp. 106–115, Istanbul, Turkey, April 2007 (doi: 10.1109/ICDE.2007.367856).

[7] M. Yuan, L. Chen, P. S. Yu, T. Yu, "Protecting sensitive labels in social network data anonymization", IEEE Trans. on Knowledge and Data Engineering, vol. 25, no. 3, pp. 633–647, March 2013 (doi: 10.1109/TKDE.2-011.259).

[8] J. Casas-Roma, J.C.J. Herrera-Joancomartí, V. Torra, "A survey of graph-modification techniques for privacy-preserving on networks", Artificial Intelligence Review, vol. 47, no. 3, pp. 341–366, May 2017 (doi: 10.1007/s10462-016-9484-8).
[9] J. Casas-Roma, J. Herrera-Joancomartí, V. Torra, "k-Degree anonymity and edge selection: improving data utility in large networks", Knowledge and Information Systems, vol. 50, no. 2, pp. 447–474, April 2017 (doi: 10.1007/s10115-016-0947-7).
[10] J. Abawajy, M. I. Ninggal, T. Herawan, "Privacy preserving social network data publication", IEEE Communications Surveys and Tutorials, vol. 18, no. 3, pp. 1974-1997, March 2016 (doi: 10.1109/COMST.2-016.2533668).
[11] E. Zheleva, L. Getoor, "Preserving the privacy of sensitive relationships in graph data", Lecture Notes in Computer Science, vol. 4890, pp. 153–171, Berlin, Heidelberg, 2008 (doi: 10.1007/978-3-540-78478-4_9).
[12] A. Campan, T.M. Truta, "Data and structural K-anonymity in social networks", Lecture Notes in Computer Science, vol. 5456, pp. 33–54, Berlin, Heidelberg, 2009 (doi: 10.1007/978-3-642-01718-6_4).
[13] S. Chester, B. Kapron, G. Ramesh, G. Srivastava, A. Thomo, S. Venkatesh, "κ-Anonymization of social networks by vertex addition", Proceeding of the CEUR, vol. 789, pp. 107–116, Vienna, Austria, Jan. 2011.
[14] M. Kiabod, M.N. Dehkordi, B. Barekatain, "TSRAM : A time-saving k-degree anonymization method in social network", Expert Systems with Applications, vol. 125, pp. 378–396, July 2019 (doi: 10.1016/j.eswa.20-19.01.059).
[15] M. Kiabod, M. Naderi, B. Barekatain, "A fast graph modification method for social network anonymization", Expert Systems with Applications, vol. 180, pp. 1-19 Article Number: 115148, Oct. 2021 (doi: 10.1016/j.eswa.2021.115148).
[16] S. Chester, B.M. Kapron, G. Ramesh, G. Srivastava, A. Thomo, S. Venkatesh, "Why waldo befriended the dummy? k-Anonymization of social networks with pseudo-nodes", Social Network Analysis and Mining, vol. 3, no. 3, pp. 381–399, Sept. 2013 (doi: 10.1007/s13278-012-0084-6).
[17] T. Ma, Y. Zhang, J. Cao, J. Shen, M. Tang, Y. Tian, A. Al-Dhelaan, M. Al-Rodhaan, "KDVEM: a k-degree anonymity with vertex and edge modification algorithm", Computing, vol. 97, no. 12, pp. 1165–1184, Dec. 2015 (doi: 10.1007/s00607-015-0453-x).
[18] R. Bredereck, V. Froese, S. Hartung, A. Nichterlein, R. Niedermeier, N. Talmon, "The complexity of degree anonymization by vertex addition", Theoretical Computer Science, vol. 607, pp. 16–34, Nov. 2015 (doi: 10.1016/j.tcs.2015.07.004).
[19] K.R. Macwan, S.J. Patel, "k-degree anonymity model for social network data publishing", Advances in Electrical and Computer Engineering, vol. 17, no. 4, pp. 117–124, Jan. 2017 (doi: 10.4316/AECE.2017.04-015).
[20] A. Sharma, S. Pathak, "Enhancement of k - anonymity algorithm for privacy preservation in social media", International Journal of Engineering and Technology, vol. 7, pp. 40–45, June 2018 (doi: 10.14419/ijet.v7i2-.27.11747).
[21] F. Rousseau, J. Casas-Roma, M. Vazirgiannis, "Community-preserving anonymization of graphs", Knowledge and Information Systems, vol. 54, no. 2, pp. 315–343, May 2018 (doi: 10.1007/s10115-017-1064-y).
[22] M. Siddula, L. Li, Y. Li, "An empirical study on the privacy preservation of online social networks", IEEE Access, vol. 6, pp. 19912–19922, April 2018 (doi: 10.1109/ACCESS.2018.2822693).
[23] X. Zhang, J. Liu, J. Li, L. Liu, "Large-scale dynamic social network directed graph k-in&out-degree anonymity algorithm for protecting community structure", IEEE Access, vol. 7, pp. 108371-108383, Aug. 2019 (doi: 10.1109/ACCESS.2019.2933151).
[24] M. Siddula, Y. Li, X. Cheng, Z. Tian, Z. Cai, "Anonymization in online social networks based on enhanced equi-cardinal clustering", IEEE Trans. on Computational Social Systems, vol. 6, no. 4, pp. 809–820, July 2019 (doi: 10.1109/TCSS.2019.2928324).
[25] S. Bourahla, M. Laurent, Y. Challal, "Privacy preservation for social networks sequential publishing", Computer Networks, vol. 170, Article Number: 107106, April 2020 (doi: 10.1016/j.comnet.2020.107106).
[26] M. Bi, Y. Wang, Z. Cai, X. Tong, "A privacy-preserving mechanism based on local differential privacy in edge computing", China Communications, vol. 17, no. 9, pp. 50–65, March 2020 (doi: 10.23919/JCC.2020.0-9.005).
[27] C. Bazgana, P. Cazalsa, J. Chlebíková, "Degree-anonymization using edge rotations", Theoretical Computer Science, vol. 873, pp. 1–15, June 2021 (doi: 10.1016/j.tcs.2021.04.020).
[28] R. Al-asbahi, "Structural anonymity for privacy protection in social network", International Journal of Scientific and Research Publications, vol. 11, no. 6, pp. 102–107, June 2021 (doi: 10.29322/IJSRP.11.06.2-021.p11414).
[29] A. Singh, M. Singh, D. Bansal, S. Sofat, "Optimised K-anonymisation technique to deal with mutual friends and degree attacks", International Journal of Information and Computer Security, vol. 14, no. 3–4, pp. 281-299, April 2021 (doi: 10.1504/IJICS.2021.114706).

[30] N. Xiang, X. Ma, "TKDA: An improved method for k-degree anonymity in social graphs", Proceeding of the IEEE/ISCC, pp. 1-6, Rhodes, Greece, June/July 2022 (doi: 10.1109/ISCC55528.2022.9912964).
[31] X. Ren, D. Jiang, "A personalized α,β,l,k-anonymity model of social network for protecting privacy", Wireless Communications and Mobile Computing, vol. 2022, pp. 1-11, Jan. 2022 (doi: 10.1155/2022/7187528).
[32] A. Majeed, S. Khan, S.O. Hwang, "A comprehensive analysis of privacy-preserving solutions developed for online social networks", Electronics (Switzerland), vol. 11, no. 13, pp. 1–37, June 2022 (doi: 10.3390/electronics11131931).
[33] A. Karimi-Rizi, M. Naderi-Dehkordi, N. Nematbakhsh, "Publishing health information without distortion while balancing desired privacy-preserving and utility", Journal of Intelligent Procedures in Electrical Technology, vol. 13, no. 50, pp. 47-66, Sept. 2022 (dor: 20.1001.1.23223871.1401.13.50.3.4).
[34] J. Casas-roma, J. Salas, F. D. Malliaros, "k-Degree anonymity on directed networks", Knowledge and Information Systems, vol. 61, pp. 1743–1768, Dec. 2019 (doi: 10.1007/s10115-018-1251-5).
[35] L. Zheng, H. Yue, Z. Li, X. Pan, M. Wu, F. Yang, "k-anonymity location privacy algorithm based on clustering", IEEE Access, vol. 6, pp. 28328–28338, Dec. 2018 (doi: 10.1109/ACCESS.2017.2780111).
[36] Y. Lu, X. Hou, X. Chen, "A novel travel-time based similarity measure for hierarchical clustering ustering", Neurocomputing, vol. 173, pp. 3–8, Jan. 2016 (doi: 10.1016/j.neucom.2015.01.090).
[37] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, S. M. Dawson, "The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations: Can geographic isolation explain this unique trait?", Behavioral Ecology and Sociobiology, vol. 54, no. 4, pp. 396–405, June 2003 (doi: 10.1007/s00265-003-0651-y).
[38] J. Leskovec, J. Kleinberg, C. Faloutsos, "Graph evolution: densification and shrinking diameters", ACM Trans. on Knowledge Discovery from Data, vol. 1, no. 1, Article Number: 2, April 2006 (doi: 10.1145/121729-9.1217301).
[39] J. Leskovec, K.J. Lang, M.W. Mahoney, "Community structure in large networks : natural cluster sizes and the absence of large well-defined clusters", Internet Mathematics, vol. 6, no. 1, pp. 29--123, Nov. 2009 (doi: 10.1080/15427951.2009.10129177).
[40] J. Yang, J. Leskovec, "Defining and evaluating network communities based on ground-truth", Proceeding of the IEEE/ICDM, pp. 745–754, Brussels, Belgium, Oct. 2012 (doi: 10.48550/arXiv.1205.6233).
[41] L. Danon, A. Díaz-Guilera, J. Duch, A. Arenas, D. Albert, J. Duch, "Comparing community structure identification", Journal of Statistical Mechanics: Theory and Experiment, vol. 09008, no. 9, pp. 219–228, Sept. 2005 (doi: 10.1088/1742-5468/2005/09/P09008).

زیرنویس‌ها

1. K-anonymity
2. l-diversity
3.t-closeness
4. Degree sequence
5. Univariate micro-aggregation for graph anonymization
6. Time-saving random anonymization method (TSRAM)
7. NaFa for k-degree anonymization
8. Identity disclosure
9. Attribute disclosure
10. Link disclosure
11. Privacy level
12. NP-hard
13. Friendship attack
14. Aggregate vector
15. Travel Time
16. Neighborhood centrality
17. End nodes coreness
18. Weight
19. Mask
20. TotalMask
21. Vertex Addition
22. K-degree anonymity with vertex and edge modification (KDVEM)
23. TKDA
24. Univariate micro-aggregation for graph anonymization- neighbourhood centrality
25. Dolphins
26. CA-GrQc
27. Ca-HepTh
28. Dolphins
29. Email-enron
30. Digital bibliography and library project
31. Youtube
32. Dolphins
33. Transitivity
34. Average clustering coefficient
35. Modularity
36. Variance information
37. Normalized mutual information
38. Split join
39. Rand index
40. Adjusted rand index