علوم رایانش و فناوری اطلاعات
نشریه علمی انجمن کامپیوتر ایران
مجلد ۲۰، شماره ۱، ۱۴۰۱
صفحات ۵۴-۶۵
مقاله پژوهشی

کشف کاربران ناهنجار با ارتباطات برجسته در شبکه‌های اجتماعی چندلایه با استفاده از گراف چندگانه

سیده سودابه غفوری[۱]، محمود دی پیر[۲]*

*نویسنده مسئول، دریافت: ۱۴۰۱/۰۱/۲۱، بازنگری: ۱۴۰۱/۰۲/۲۰، پذیرش: ۱۴۰۱/۰۳/۲۸

[۱] دانش‌آموخته کارشناسی‌ارشد مهندسی فناوری اطلاعات، دانشگاه آزاد واحد علوم و تحقیقات، تهران، ایران
[۲] دانشیار، دانشکده مهندسی کامپیوتر، دانشگاه هوایی شهید ستاری، تهران، ایران

چکیده

در دنیای امروز بیشتر تعاملات بشر از طریق شبکه‌های اجتماعی برخط متفاوت انجام می‌گیرد و اطلاعات بسیاری در این بستر، گسترده شده است و همین امر باعث شده تا کاربران بدخواه به دنبال اعمال غیرقانونی و ارتباطات ناهنجار در این شبکه‌ها باشند. علمی که به تشخیص اعمال غیرقانونی در شبکه‌ها می‌پردازد تشخیص ناهنجاری نام دارد. در این تحقیق، به تشخیص ناهنجاری در شبکه‌های اجتماعی می‌پردازیم. از آن‌جایی‌که افراد از طریق شبکه‌های اجتماعی مختلفی می‌توانند با هم ارتباط برقرار کنند بنابراین بهتر است تمام این شبکه‌ها را بررسی کنیم تا نتایج بهتری حاصل شود. توجه به این امر باعث اهمیت تحلیل شبکه‌های اجتماعی چندلایه می‌شود. ناهنجاری موردبررسی در این تحقیق، ارتباط قوی برجسته بین کاربران می‌باشد که این ناهنجاری را در شبکه‌های اجتماعی چندلایه بررسی می‌کنیم. برای کشف ناهنجاری مذکور در شبکه اجتماعی چندلایه، الگوریتمی با دو فاز ارائه می‌شود که در فاز اول درجه ناهنجاری هر کاربر در هر لایه شبکه محاسبه می‌شود و در فاز دوم با توجه به اهمیت هر لایه برای هر کاربر، درجه ناهنجاری نهایی در کل شبکه محاسبه می‌شود. برای اعتبارسنجی این الگوریتم از سه مجموعه داده شبکه‌های دنیای واقعی استفاده کردیم که نتایج به‌دست‌آمده نشان‌دهنده کارایی روش پیشنهادی است.

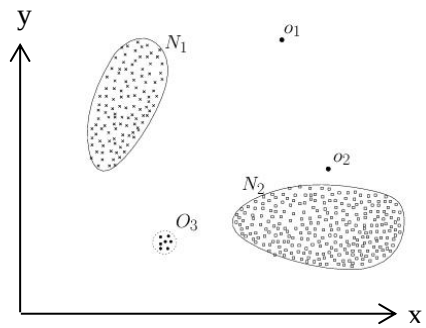
کلمات کلیدی: تشخیص ناهنجاری، کاربران ناهنجار، شبکه‌های اجتماعی، شبکه‌های چندلایه، گراف چندگانه.

۱- مقدمه

در دنیای امروز، شبکه‌های اجتماعی برخط[۱] محیطی برای افراد فراهم کرده‌اند تا بتوانند باهم در تعامل باشند. چنین محیطی که در دهه‌های اخیر بسیار فراگیر شده و استفاده از آن اجتناب‌ناپذیر است باعث شده مقدار زیادی اطلاعات در آن گسترده شود و محلی برای سوءاستفاده کاربران بدخواه[۲] شود. کشف فعالیت‌های غیرعادی یا افرادی که فعالیت‌های غیرعادی انجام می‌دهند در شبکه‌های اجتماعی با عنوان تشخیص ناهنجاری[۳] مطرح می‌شود.

ناهنجاری، تعاریف متفاوتی دارد اما تعریف جامعی که بیان شده است، تعریفی از هاوکینز در سال ۱۹۸۰ است: «ناهنجاری مشاهده‌ای است که از مشاهدات دیگر بسیار متفاوت است و شک‌برانگیز است و به‌وسیله مکانیسم متفاوتی تولید شده است» [۱]. همان‌گونه که در شکل ۱ مشاهده می‌شود،

دو ناحیه عادی N_1 و N_2 داریم که نمونه داده‌ها در آن ناحیه‌ها گرد هم آمده‌اند و نقاط O_1، O_2 و O_3 که دور از این نواحی هستند، ناهنجاری‌های این فضا می‌باشند.



شکل ۱. نمایی از ناهنجاری در مجموعه داده دو بعدی [۱۸]

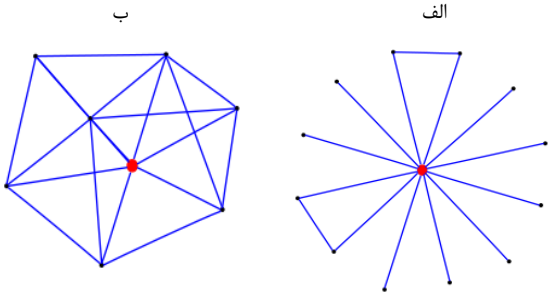
ناهنجاری در شبکه‌های اجتماعی برخط به‌صورت‌های مختلف یافت می‌شود: کاربران ناهنجار[۴]، تعاملات ناهنجار[۵]، گروه‌های ناهنجار[۶]، رویدادهای ناهنجار[۷] [۱۸]. الگوهای ساختاری کاربران یا گره‌های ناهنجاری که وجود دارند به شرح زیر هستند [۲]:

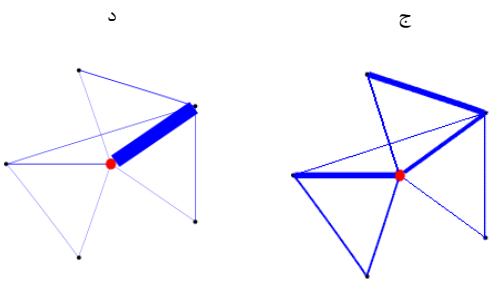
ستاره / ستاره‌مانند[۸]: در هر توپولوژی ستاره‌ای، همسایگان هر گره، هیچ ارتباطی با یکدیگر ندارند و در ستاره‌مانند همسایگان، ارتباطات بسیار کمی باهم دارند (شکل ۲- الف).

گروه / گروه‌مانند[۹]: هر گروه، یک زیرگراف کاملاً متصل در شبکه است. تعداد لبه‌ها در یک گروه با n گره، برابر با $\frac{n(n-1)}{2}$ است. اگر تعداد کمی از همسایه‌های گره، به هم متصل نباشند، گره از الگوی گروه- مانند پیروی می‌کند (شکل ۲- ب).

همسایگی قوی[۱۰]: اگر کاربری با n کاربر ارتباط داشته باشد توقع داریم تعداد تماس‌ها یا وزن نهایی، تابعی از n باشد. اگر وزن نهایی خیلی زیاد باشد همسایگی قوی وجود دارد (شکل ۲- ج).

ارتباط قوی برجسته[۱۱]: اگر ارتباط خاصی در یک همسایگی تک‌گامی[۱۲]، بسیار قوی باشد یعنی ارتباطی طولانی برقرار باشد. این ارتباط، ارتباطی مشکوک است (شکل ۲- د).

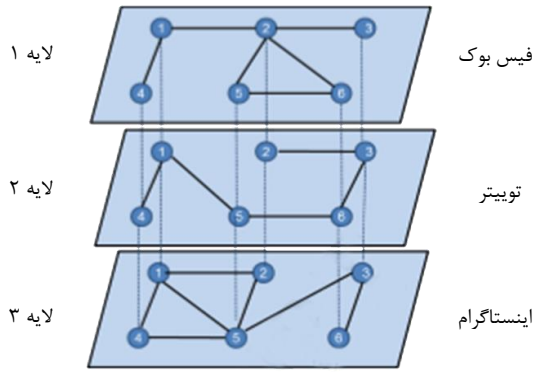




شکل ۲. الف) ستاره / ستاره‌مانند ب) گروه / گروه‌مانند ج) همسایگی قوی د) ارتباط قوی برجسته [۲]

برای کشف ناهنجاری‌ها از روش‌های تشخیص ناهنجاری استفاده می‌شود. تشخیص ناهنجاری، استراتژی‌ای است که الگوهایی از داده‌ها را پیدا می‌کند که با رفتار موردانتظار مطابقت ندارند. برای تشخیص ناهنجاری‌ها می‌توان از روش‌های مبتنی بر ساختار[۱۳]، مبتنی بر رفتار[۱۴] و مبتنی بر طیف[۱۵] استفاده کرد که بیشتر از روش‌های ساختاری استفاده می‌شود زیرا ویژگی‌های رفتاری توسط کاربران قابل پنهان‌کردن است. برای مثال کاربر می‌تواند پروفایل خود را کاملاً متفاوت از چیزی که واقعاً هست وارد کند و خود را پشت ظاهر مجازی ساخته‌شده، پنهان کند. ولی روش‌های مبتنی بر ساختار، از توپولوژی شبکه و ویژگی‌های استاندارد شبکه استفاده می‌کنند و قابل پنهان‌کردن نیستند. برای مثال ارتباطی که در شبکه وجود دارد قابل‌انکار نیست. بنابراین در روش‌های مبتنی بر ساختار، شبکه اجتماعی را به‌صورت گراف اجرا می‌کنند که گره‌ها[۱۶] بیانگر کاربران شبکه و لبه‌ها[۱۷] بیانگر تعاملات بین کاربران شبکه هستند و موجودیت‌های ناهنجار را شناسایی می‌کنند.

تعاملات بشر ذاتاً چندگانه است و افراد می‌توانند به روش‌های مختلفی با هم ارتباط برقرار کنند مثل تلفن، پیامک، شبکه‌های اجتماعی و ... . افراد در شبکه‌های اجتماعی می‌توانند به چند روش باهم تعامل داشته باشند که منجر به تشکیل شبکه‌های اجتماعی چندلایه می‌شود. افراد می‌توانند هم‌زمان در چند شبکه اجتماعی مثل فیس‌بوک، تلگرام، اینستاگرام و ... عضو باشند و با فهرست دوستان خود از طریق هرکدام از این شبکه‌ها ارتباط داشته باشند. شکل ۳، نمونه‌ای از شبکه اجتماعی سه لایه را نشان می‌دهد.



شکل ۳. نمونه‌ای از شبکه اجتماعی سه لایه [۳]

بحث ناهنجاری در شبکه‌های چندلایه نیز مطرح می‌شود. برای کشف ناهنجاری‌ها در شبکه‌های اجتماعی چندلایه، روش معمول این است که شبکه چندلایه را به شبکه تک‌لایه تبدیل می‌کنند به‌گونه‌ای که گره‌های شبکه تک‌لایه همان گره‌هایی باشد که در تمام لایه‌ها تکرار شده است و لبه‌های شبکه تک‌لایه مجموع لبه‌های موجود در همه لایه‌ها باشد و سپس از روش‌های موجود برای شبکه‌های تک‌لایه استفاده می‌کنند اما باید این مسئله را درنظر بگیریم که شبکه‌های چندلایه می‌توانند شامل اطلاعات غنی‌تری نسبت به شبکه‌های تک‌لایه باشند و ما می‌توانیم از این اطلاعات استفاده کنیم. بنابراین بهتر است از روش‌هایی استفاده کنیم که اطلاعات مهم شبکه چندلایه را از دست ندهیم.

در تحقیق انجام‌شده هدف این است که کاربران ناهنجار با الگوی «ارتباطات قوی برجسته» کشف شوند. برای کشف این ناهنجاری از الگوریتم ADOMS [۳] با تغییراتی استفاده می‌شود.

۲- پژوهش‌های پیشین

برای مرور مطالعات و پژوهش‌های گذشته در زمینه تشخیص ناهنجاری درشبکه‌های اجتماعی چندلایه، ابتدا مروری بر پژوهش‌های انجام‌شده در مورد تشخیص ناهنجاری در شبکه‌های اجتماعی داریم، سپس به مطالعاتی که در رابطه با شبکه‌های چندلایه انجام شده است می‌پردازیم و سپس به مقالات حوزه تشخیص ناهنجاری در شبکه‌های اجتماعی چندلایه می‌پردازیم. از آن‌جایی که مفهوم ناهنجاری، فاصله از هنجار است در بیشتر مقالات، یک الگو و هنجار در جامعه نمونه تعریف شده است و هر آنچه فاصله بیشتری از این الگو در جامعه یافت شده است به‌عنوان درجه ناهنجاری تعریف شده است و بر این اساس کابران ناهنجار مشخص شده‌اند.

ناهنجاری‌های ساختاری مهمی که در شبکه‌های اجتماعی مطرح می‌شود: ستاره/ ستاره‌مانند، گروه/ گروه‌مانند، همسایگی قوی و ارتباطات قوی برجسته هستند که در [۲] بررسی شده‌اند و با استفاده از الگوریتم OddBall و تمرکز روی همسایگی تک‌گامی هر گره کشف شده‌اند. این الگوریتم، گره‌های ناهنجار در گراف‌های وزن‌دار را با استفاده از ویژگی‌های گراف تشخیص می‌دهد. الگوی نرمالی برای گره‌ها تعریف شده است که میزان فاصله از این الگو، درجه ناهنجاری را نشان می‌دهد. روش استفاده‌شده در این مقاله، یک فرمول ابتکاری[۱۸] به‌اضافه روش تشخیص ناهنجاری بر مبنای تراکم شبکه است. تکنیک تشخیص بر مبنای تراکم از فاکتور دور از محل[۱۹] [۱۱] استفاده می‌کند که این فاکتور، برابر تعداد نزدیک‌ترین همسایه‌هایی که در همسایگی محلی گره قرار دارند می‌باشد. در حقیقت گره‌هایی که دور از تراکم همسایگی محلی گره باشند ناهنجار هستند. برای کشف ناهنجاری ستاره‌مانند و گروه‌مانند، از ویژگی‌های دیگری مثل: تعداد گره‌ها، تعداد لبه‌ها، مقدار واسط[۲۰]، میانگین مرکزیت بینابینی[۲۱] می‌توان استفاده کرد که در [۵] مطرح شده‌اند. در این مقاله، رویکردی در سه مرحله ارائه شده است که مرحله اول همسایگی تک‌گامی برای همه گره‌ها مشخص می‌کند. مرحله دوم ویژگی‌های گراف برای تشخیص ناهنجاری را انتخاب می‌کند و درجه ناهنجاری را محاسبه می‌کند. در مرحله سوم به گره‌ها برچسب هنجار یا ناهنجار زده می‌شود.

شبکه‌های اجتماعی اصولاً به‌صورت گراف اجرا می‌شوند. ۶ نوع ناهنجاری برای گراف می‌تواند اتفاق بیفتد [۶]: رأسی در گراف وجود دارد که انتظارش را نداریم، لبه‌ای وجود دارد که انتظارش را نداریم، برچسبی روی لبه‌ای وجود دارد که انتظارش را نداریم، برچسبی که انتظارش را داریم وجود ندارد، رأسی که انتظارش را داریم وجود ندارد، لبه‌ای که انتظارش را داریم وجود ندارد. برای کشف الگوی نرمال در گراف از اصل MDL[22] استفاده شده است تا الگوی اصلی و هنجار درگراف را تعیین کند و الگوهای مشابه از نظر ساختاری را پیدا کند و هر الگویی که از این الگوی هنجار منحرف شود ناهنجاری است. در [۴] کاربرانی که محتوای نامناسب به اشتراک می‌گذارند، کاربرانی که حساب‌های تقلبی[۲۳] دارند و کاربرانی که حساب‌هایی را هک می‌کنند بررسی شده‌اند. کاربرانی که محتوای نامناسب به اشتراک می‌گذارند عموماً دوستان کمی در لیست خود دارند ولی اطلاعات زیادی را به اشتراک می‌گذارند و درجه خروجی[۲۴] آن‌ها بالاست. کاربران قلابی، اصولاً فعالیت‌هایشان را روی کاربر خاصی انجام می‌دهند بنابراین بزرگ‌ترین درجه خروجی کاربران مهم است. کاربرانی که حسابی را هک می‌کنند اصولاً در یک دوره زمانی به‌طور ناگهانی فعالیتشان افزایش می‌یابد ولی درجه خروجی‌شان تغییر نمی‌کند. در این حالت، ارتباط بین درجه خروجی و زمان فعالیت مهم است.

در سال ۲۰۱۸ در [۱۲] روشی برای کشف گره‌های ناهنجار در شبکه‌های اجتماعی بلادرنگ که حجم وسیعی از داده‌ها را شامل می‌شوند ارائه شده است. جریان داده‌ها از شبکه‌های اجتماعی تولید می‌شود و توسط آپاچی کافکا[۲۵] ذخیره می‌شود و جریان تولیدشده را آپاچی اسپارک[۲۶] پردازش می‌کند و داده‌های پردازش شده در بانک اطلاعاتی ذخیره می‌شوند. سپس از طریق استانداردهای گراف مثل درجه، گره‌های ناهنجار تشخیص داده می‌شوند. در [۱۳] روش جدیدی بر مبنای مدل برنامه‌نویسی TALV[۲۷] با آنالیز بصری پیشنهاد شده است. سیستمی به نام GiViP طراحی شده است که داده‌های تبادل‌شده بین گره‌ها در سراسر شبکه را جمع‌آوری می‌کند و رابط بصری مناسبی برای ارتباطات در شبکه برای جستجو به تحلیل‌گر ارائه می‌دهد. مواردی که توسط این سیستم قابل مشاهده است، الگوی پیام‌های ناهنجار، بخش‌های نامتعادل گراف، ارتباطات با تأخیر بالا، واحد محاسباتی سرریز و ... می‌باشد.

در مقالات بررسی‌شده در خصوص تشخیص ناهنجاری در شبکه‌های اجتماعی، از آن‌جا که در گذشته تعداد شبکه‌های اجتماعی بسیار محدود بوده فقط یک شبکه اجتماعی بررسی شده است ولی اکنون هر کاربر در شبکه‌های بسیاری، عضو است و باید تمام ارتباطاتش در بررسی‌ها لحاظ شود تا نتایج بهتری حاصل شود. همچنین راهکارهای پیشنهادی بر مبنای ویژگی‌های گراف مانند وزن ارتباط یا تعداد ارتباط ورودی یا خروجی یک کاربر می‌باشند که به علت دقیق‌بودن روش‌ها و فرمول‌های ریاضی، این مورد بیانگر نتایج بسیار قابل‌اعتماد است.

در [۱۵] نوع نمایشی برای شبکه‌های چندلایه مطرح شده است به این ترتیب که اولاً برای نمایش هر نوع ارتباط، یک نمودار جداگانه رسم کنیم و سپس برای تجمیع این لایه‌ها به گونه‌ای عمل کنیم که هر گره یک نمودار دایره‌ای[۲۸] باشد که هر قسمتش میزان مشارکت این گره در لایه جداگانه‌ای

را نشان دهد و اندازه هر گره با توجه به میزان مرکزیت گره در گراف تنظیم شود و ارتباطات بین گره‌ها با رنگ‌های متفاوتی نمایش داده شود. به این ترتیب در شبکه چندلایه رسم‌شده، میزان مرکزیت گره‌ها، تراکم شبکه در هر لایه، میزان ارتباطات قوی و ضعیف در هر لایه و بسیاری ویژگی‌های دیگر قابل مشاهده است. در سال ۲۰۱۴ در [۷] نرم‌افزار متن بازی به نام MuxViz معرفی شده است که مجموعه الگوریتم‌هایی برای تحلیل شبکه‌های چندلایه دارد. این محیط کار، شبکه چندلایه را به‌صورت گراف‌هایی در لایه‌های جداگانه به‌صورت گراف چندگانه نمایش می‌دهد. توپولوژی شبکه، وزن ارتباطات در شبکه، مرکزیت‌های مختلف و ... در این ابزار قابل مشاهده است. در سال ۲۰۱۵ در [۱۴] نیز محیط کار دیگری به نام Net2Plan برای تحلیل شبکه‌های چندلایه معرفی شده است که بر مبنای مفاهیمی مثل گره‌ها، ارتباطات، مسیرها، ترافیک داده‌ها، لایه‌های شبکه و ... کار می‌کند. این ابزار هم روی شبکه‌های آفلاین و هم روی شبکه‌های برخط امکان تحلیل را فراهم می‌کند. توپولوژی شبکه، مسیریابی ترافیک، ظرفیت ارتباط، بخش حفاظت و موارد دیگر در این ابزار در حالتی که شبکه آفلاین را شبیه‌سازی کردیم قابل بررسی هستند. این ابزار این امکان را فراهم می‌کند که الگوریتم دلخواه را روی شبکه پیاده‌سازی کنیم.

زمانی که شبکه‌ها را به‌صورت ریاضی نمایش می‌دهیم می‌توانیم توصیف مفیدی از ساختار شبکه داشته باشیم اما شبکه‌های چندلایه به‌وسیله ماتریس مجاورت معمولی و گراف ساده قابل نمایش نیستند زیرا بین هر دو گره به‌جای یک ارتباط، چند نوع ارتباط وجود دارد بنابراین باید محیط کار و فرمولاسیون ریاضی جدیدی معرفی شود که شبکه چندلایه را با اطلاعات مخفی در آن به‌خوبی نمایش دهد که در [۸] یک محیط‌کار و چارچوب ریاضی برای شبکه‌های چندلایه به نام محیط‌کار تانسور[۲۹] معرفی شده است، از تانسور مجاورت[۳۰] برای نمایش شبکه‌های چندلایه استفاده می‌شود.

کاربران در شبکه‌های اجتماعی مختلف، اصولاً پروفایل‌های مختلفی دارند. اگر این پروفایل‌ها واضح باشند می‌توان اطلاعات تمام پروفایل‌ها در تمام لایه‌ها را با هم ترکیب کرد و به شبکه تجمیع‌شده رسید و اطلاعات کاربران را در یک بانک اطلاعاتی مرتبط ذخیره کرد. وضوح پروفایل کاربران بسیار مهم است زیرا اجازه می‌دهد لبه‌های بین‌لایه‌ای بین لایه‌ها ایجاد شود. اما به هر حال چنین وضوحی وجود ندارد. بنابراین در سال ۲۰۱۹ در [۱۶] روشی برای ترکیب تقریبی گراف‌های شبکه‌های اجتماعی چندلایه با استفاده از انواع مختلف اطلاعات مثل ویژگی‌های اولیه کاربر، فعالیت کاربر و توپولوژی ارتباطات کاربر ارائه شده است. گراف نهایی رسم‌شده شامل مجموع گره‌های تمام لایه‌هاست و لبه‌ها شامل ارتباطات بین گره‌ها در لایه‌های مختلف و لبه‌های بین‌لایه‌ای است. جدیدترین مبحثی که در شبکه‌های اجتماعی چندلایه بررسی شده است نسل جدید شبکه‌های اجتماعی برخط غیرمتمرکز[۳۱] است که هر لایه در این شبکه، مفهوم جداگانه‌ای را بیان می‌کند. در این شبکه‌ها علاوه بر این که افراد مختلف با هم ارتباط دارند، اشیای هوشمند محیط نیز می‌توانند در این ارتباطات شرکت کنند. در [۱۷] ایده جدیدی برای مدلسازی شبکه‌های اجتماعی غیرمتمرکز نظیر به نظیر چندلایه ارائه شده است. در این ایده از مفهوم اگونت استفاده شده است که هر کاربر به همراه تمام کاربران و اشیای هوشمند متصل به آن، یک شبکه اگونت محسوب می‌شود. محیط کاری که برای این مدلسازی معرفی شده است HELIOS است که دو سطح مختلف دارد: سطح اول شبکه اگونت یک گره را نشان می‌دهد و سطح دوم مدیریت لایه‌ها را انجام می‌دهد. در مقاله [۱۹] روشی بر مبنای زنجیره مارکوف برای شناسایی رفتار افراد در شبکه‌های اجتماعی ارائه شده است. در این روش ابتدا مدلی بر مبنای رفتار نرمال افراد ساخته می‌شود و سپس از این مدل برای شناسایی فعالیت‌های غیرمنتظره در شبکه‌های اجتماعی استفاده می‌گردد. اخیراً استفاده از یادگیری عمیق، کاربردهای فراوانی پیدا کرده است. از جمله این کاربردها، شناسایی اطلاعات غلط و گمراه‌کننده‌ای است که برخی از کاربران بدخواه در شبکه‌های اجتماعی مختلف مثل توئیتر و فیسبوک منتشر می‌کنند [۲۰].

طبق مقالات بررسی‌شده در زمینه شبکه‌های چندلایه، این شبکه‌ها موردتوجه بسیاری از دانشمندان قرار گرفته است و ابزارهای زیادی برای نمایش این شبکه‌ها به‌صورت بصری و گراف‌های چندگانه ایجاد شده‌اند. ابزارهای فوق، نمایانگر ویژگی‌های بسیار جالبی در ارتباطات چندگانه هستند که می‌توانند به شناسایی انواع ارتباطات کمک کنند. یکی از مواردی که در این ابزارها قابل تشخیص است ارتباطات قوی و برجسته نسبت به بقیه است.

کاربران ناهنجار با الگوهای ساختاری گروه/ گروه‌مانند و ستاره/ ستاره‌مانند در [۳] بررسی شده که از رویکرد ADOMS برای تشخیص آن‌ها استفاده شده است. این رویکرد پس از استخراج ویژگی‌های مناسب برای داده‌کاوی، الگوی نرمال شبکه را تعریف می‌کند و درجه ناهنجاری گره‌ها در هر لایه شبکه را محاسبه می‌کند و سپس با محاسبه اهمیت هر لایه، به هر گره درجه ناهنجاری نهایی اختصاص می‌یابد و گره‌ها بر اساس درجه ناهنجاری، رتبه‌بندی می‌شوند.

در [۹] نوع ناهنجاری بررسی‌شده، ناهنجاری افقی[۳۲] است که در شبکه‌های چندلایه به‌وجود می‌آید، کاربری در تمام شبکه‌ها با افراد مشخصی ارتباط دارد ولی در یک شبکه خاص کاملاً ارتباطات متفاوتی دارد. هر کاربری که در منابع مختلف رفتارهای مختلف دارد ناهنجاری نیست بلکه باید به‌صورت کیفی بررسی شود. در این مقاله از رویکرد NHAD که بر مبنای شبکه عصبی و سیستم استنتاج فازی است برای کشف، بازیابی و حذف این ناهنجاری‌ها استفاده می‌شود. برای کشف گره‌های ناهنجار در شبکه‌های چندلایه استانداردی به نام CAD [۳۳] مورد استفاده قرار می‌گیرد که در [۱۰] معرفی شده است. در این مقاله برای هر گره شبکه مرکزیت بردار ویژه، مرکزیت page-rank و مرکزیت درجه محاسبه می‌شود زیرا این مرکزیت‌ها مهم‌ترین گره شبکه را نشان می‌دهند. سپس یک مجموعه آموزشی با استفاده از خروجی هر معیار مرکزیت ایجاد می‌کنند و توزیع گاوس را برای آن محاسبه می‌کنند و سپس مقداری که کم‌ترین احتمال را دارد پیدا می‌کنند و گره‌هایی که کمترین مقدار احتمال را دارند ناهنجاری هستند.

تشخیص ناهنجاری در شبکه‌های اجتماعی چندلایه موضوعی است که بسیار کم بررسی شده و همین موضوع جذابیت آن را دو چندان کرده است.

۳- بیان مسئله

شبکه‌های اجتماعی برخط بانک عظیمی از اطلاعات کاربران عضو در آن‌ها می‌باشند. هر جایی که اطلاعاتی به‌صورت گسترده پخش می‌شود مکانی برای افراد سودجوست و همچنین بحث امنیت اطلاعات پیش می‌آید. از مباحث امنیتی بسیار مهم تشخیص ارتباطات ناهنجار، اشخاص ناهنجار و

گروه‌های ناهنجار در شبکه‌ها می‌باشد که ما به دنبال کشف ناهنجاری با الگوی «ارتباطات قوی برجسته» هستیم. کاربران در شبکه‌های اجتماعی با افراد مختلفی ارتباط دارند ولی بعضی از این ارتباطات متفاوت از بقیه هستند و ارزش بالاتری نسبت به بقیه دارند. یک ارتباط در شبکه ممکن است بسیار طولانی ادامه داشته باشد یا در یک بازه زمانی بارها تکرار شود یا ارتباط با یک شخص خاص برقرار شود. این موارد، مواردی شک‌برانگیز هستند که باید بررسی شوند. برای مثال یک کاربر می‌تواند ۱۰۰۰ ارتباط برقرار کرده باشد ولی ۸۰۰ ارتباط با شخص خاصی بوده باشد. یا شخصی ۱۰۰ مقاله در ژورنال‌های معتبر چاپ کرده ولی ۹۰ مقاله را در یک ژورنال خاص چاپ کرده است.

به‌طور کلی در این مقاله هدف، کشف ارتباطات، خاص است. ارتباطات کاربر در تمام شبکه‌های اجتماعی که عضو هستند، بررسی می‌گردد و در نهایت مشخص می‌شود که کدام ارتباط، ارتباطی قوی و متفاوت از ارتباطات دیگر است. البته کشف یک ارتباط قوی برجسته به این معنی نیست که حتماً این ارتباط کشف‌شده، یک ارتباط برای انجام فعالیت‌های بدخواهانه است. پس از اینکه ارتباطی قوی یا خاص را در شبکه چندلایه یافتیم باید این ارتباط به‌صورت میدانی بررسی شود تا مشخص گردد آیا واقعاً یک ارتباط در مسیر بدخواهی است یا خیر.

۴- مدلسازی مسئله ناهنجاری

در طرح مسئله پیشنهادی، فرض بر این است که یک شبکه اجتماعی چندلایه وزن‌دار، ایستا و غیرجهت‌دار شبیه $G=\{G^1.G^2....G^L\}$ داریم که دنباله محدودی از L لایه شبکه دارد و هر لایه، با یک نوع ارتباط کاربران متناظر است. هر لایه در شبکه چندلایه را می‌توان به خودی خود یک شبکه درنظر گرفت و l اُمین لایه شبکه چندلایه را با G^l نشان داد و V^l و E^l به‌ترتیب، مجموعه گره‌ها و مجموعه لبه‌های درون لایه l هستند. هر گره باید در همه لایه‌ها وجود داشته باشد. اگر گره‌هایی داشته باشیم که در یک لایه هستند و در لایه‌های دیگر نیستند، اجتماع گره‌ها در تمام لایه‌ها را محاسبه می‌کنیم. N تعداد گره‌ها را مشخص می‌کند. می‌توان L ماتریس همسایگی را تعریف کرد و برای هر لایه ماتریس $A_G=\{A^{[1]}.A^{[2]}....A^{[L]}\}$ را تعریف کرد که $A^{[l]}=\{a_{ij}^{[l]}\}$ و $a_{ij}^{[l]}=w$ است. یعنی ارتباط دو گره در صورتی برقرار است که اگر و تنها اگر i و j در لایه l به هم متصل باشند و وزن ارتباط برابر w باشد. نمادهای اصلی استفاده‌شده در این تحقیق، در جدول ۱ تعریف شده است.

جدول ۱. نمادها

نماد	توصیف نمادها
G	شبکه چندلایه
L	تعداد لایه‌ها در G
V	مجموعه گره‌ها در G
N	تعداد گره‌ها در G
G^l	اُمین لایه شبکه چندلایه G
E^l	مجموعه یال‌ها در Gl
A_G	مجموعه ماتریس‌های همسایگی n*n متناظر با G
$A^{[l]}$	ماتریس‌های همسایگی n*n گراف Gl با درایه‌های $a_{ij}^{[l]}$
W_i	وزن کل لبه‌ها در همسایگی تک‌گامی گره i
λ_i	مقدار ویژه ماتریس مجاورت وزن همسایگی تک‌گامی گره i
$aScore_i^l$	درجه ناهنجاری برای گره i در Gl
LR_i^l	اهمیت لایه برای گره i در Gl
$multiScore_i$	درجه ناهنجاری برای گره i در کل شبکه چندلایه

۵- رویکرد پیشنهادی

در این بخش، رویکرد پیشنهادی برای کشف ناهنجاری با الگوی ارتباطات قوی برجسته در شبکه‌های اجتماعی چندلایه ارائه می‌شود. در این تحقیق از الگوریتم ADOMS استفاده می‌شود. الگوریتم فوق برای کشف کاربران ناهنجار با الگوی ناهنجاری گروه/ گروه‌مانند و ستاره/ ستاره‌مانند استفاده شده است و ما می‌خواهیم کاربران ناهنجار با الگوی ارتباطات قوی برجسته را کشف می‌کنیم. مدل پیشنهادی، دو فاز دارد.

۵-۱- فاز ۱: پیدا کردن درجات ناهنجاری گره‌ها در تک‌تک لایه‌های شبکه

برای کشف کاربران ناهنجار در شبکه چندلایه ابتدا باید درجه ناهنجاری تمام گره‌ها در هر لایه را به‌دست آوریم. برای اختصاص درجات ناهنجاری به هر گره، باید فضای ویژگی را به گونه‌ای مؤثر شناسایی کنیم و الگوها و قوانینی پیدا کنیم که به‌وسیله بیشتر گره‌های شبکه پیروی می‌شود. داده-هایی که از این الگوها تخلف می‌کنند به‌عنوان ناهنجاری شناسایی می‌شوند. بنابراین فاز اول مدل پیشنهادی ۳ مرحله دارد:

۵-۱-۱- استخراج ویژگی‌های مناسب

رویکرد پیشنهادی یک رویکرد مبتنی بر ساختار شبکه است بنابراین از ویژگی‌های ساختاری شبکه استفاده می‌شود. هر کاربر در شبکه اجتماعی را اگو و شبکه‌ای از گره‌هایی که با این گره، ارتباط تک‌گامی دارند و تمام ارتباطات بین آن‌ها را اگونت می‌نامند. اولین گام در کشف ناهنجاری این است که اگونت‌های مربوط به هر گره را مشخص کنیم. ویژگی‌هایی که به تشخیص ناهنجاری ارتباطات قوی برجسته کمک می‌کنند و در این تحقیق استخراج می‌شوند وزن کل اگونت و مقدار ویژه ماتریس مجاورت وزن اگونت می‌باشند که در الگوریتم oddball [۲] بیان شده‌اند. علت استفاده از مقدار ویژه ماتریس این است که این مقدار، بین ارزش یال‌ها تفاوت قائل است و نشان‌دهنده ارتباط گره‌ها با گره‌های مؤثر در شبکه است.

۵-۱-۲- الگویابی در هر لایه شبکه

یکی از معضلات اصلی کشف ناهنجاری، فهم رفتار عمومی یا نرمالی است که بیشتر آیتم‌های داده‌ای درون شبکه دنبال می‌کنند. رفتار مشکوک کاربر را می‌توان با یافتن انحراف از رفتار نرمال شناسایی کرد. پس از تعیین ویژگی‌ها باید ارتباط این ویژگی‌ها مشخص گردد و یک قانون بین این ویژگی‌ها تعریف شود که بیشتر کاربران از آن پیروی کنند. وقتی با توجه به ویژگی‌های استخراجی از گراف، یک نمودار از نقاط دوبعدی به‌دست‌آمده رسم می‌کنیم و نمودار برازش تولید می‌کنیم تابعی که نقاط را به‌خوبی برازش می‌کند پیشگویی می‌شود. این تابع به‌صورت زیر است:

$$\lambda_i = \theta * w_i + c \qquad (۱)$$

در این معادلهٔ خط، θ شیب خط پیشگویی‌شده و C عرض از مبدأ خط است. برای هر لایه شبکه تک‌لایه، یک خط برازش پیشگویی می‌شود یعنی در هر لایه داده‌ها باید از قانون مربوط به آن لایه پیروی کنند.

۵-۱-۳- محاسبه درجه ناهنجاری برای هر گره در هر لایه

درجه ناهنجاری، در حقیقت فاصله از قانون است و با توجه به ویژگی‌های استخراج‌شده برای هر گره در هر لایه جداگانه محاسبه می‌شود. برای محاسبه این مقدار از فرمول فاصله، از خط برازش استفاده می‌شود.

$$a.Score_i^l = Out_line_i^l = \frac{\left|-\theta^l * w_i^l + \lambda_i^l - C^l\right|}{\sqrt{\theta^2 + C^2}} \qquad (۲)$$

* C و θ مقادیر عرض از مبدأ و شیب خط هستند که از خط برازش قابل محاسبه هستند.

* w و λ مقادیر وزن کل و مقدار ویژه اگونت هستند که ویژگی‌های استخراجی از گراف هستند.

۵-۲- فاز ۲: رتبه‌بندی گره‌ها با توجه به درجات ناهنجاری در شبکه چندلایه

در شبکه‌های چندلایه چیزی که خیلی اهمیت دارد استفاده از ویژگی‌های نهفته در لایه‌هاست. به هر گره در هر لایه شبکه وزنی به نام اهمیت لایه اختصاص می‌دهیم. پس از محاسبه این معیار، درجه ناهنجاری کل گره را محاسبه می‌کنیم و سپس درجات را مرتب می‌نماییم. فاز دوم الگوریتم نیز ۳ مرحله دارد:

۵-۲-۱- محاسبه اهمیت لایه

اختصاص اهمیت‌های مختلف به لایه‌های شبکه چندلایه با توجه به ویژگی‌های ارتباطات در محاسبه درجه ناهنجاری بسیار تأثیرگذار است. اهمیت لایه برای یک گره، اتصال همسایگی در لایه را در مقایسه با لایه‌های دیگر در شبکه چندلایه نشان می‌دهد. اگر اتصال یک کاربر در یک لایه زیاد باشد یعنی این لایه در محاسبه درجه ناهنجاری سهم بیشتری دارد. اگر اتصال یک گره در لایه، خلوت باشد، آن لایه کمتر در درجه ناهنجاری گره در شبکه چندلایه سهیم است. بنابراین برای محاسبه اهمیت لایه برای هر گره باید وزن همسایگی هر گره در یک لایه را به وزن کل همسایگی همان گره در تمام لایه‌ها طبق رابطه ۳ تقسیم کنیم.

$$LR_i^l = \frac{w_i^l}{\sum_{l=1}^{L} w_i^l} \qquad (۳)$$

w_i^l : وزن گره i در لایه l

۵-۲-۲- محاسبه درجه ناهنجاری نهایی برای هر گره در شبکه چندلایه

پس از محاسبه اهمیت لایه گره‌های درون هر لایه از شبکه، درجات ناهنجاری برای گره‌های درون شبکه چندلایه، به‌صورت ترکیب خطی درجات ناهنجاری در هریک از لایه‌های شبکه چندلایه محاسبه می‌شود. اهمیت لایه گره‌ها به‌صورت ثابت‌هایی برای ترکیب خطی در نظر گرفته می‌شوند. ارتباط خطی نشان‌دهنده میزان مشارکت یک لایه خاص در درجه ناهنجاری یک گره در شبکه چندلایه است. درجه ناهنجاری نهایی هر گره در شبکه چندلایه از ضرب درجه ناهنجاری هر گره در هر لایه در اهمیت هر لایه برای هر گره به‌صورت زیر محاسبه می‌شود:

$$multiScore_i = \sum_{l=1}^{L} LR_i^l . aScore_i^l \qquad (۴)$$

LR_i^l : اهمیت لایه l برای گره i

$aScore_i^l$: درجه ناهنجاری گره i در لایه l

۵-۲-۳- رتبه‌بندی گره‌ها بر اساس درجات ناهنجاری

در نهایت، درجات ناهنجاری گره‌ها به‌صورت نزولی مرتب می‌شوند و رتبه گره‌ها براساس شماره خانه آن گره بازگردانده می‌شود. طبق این رتبه‌بندی گره‌های با درجات بالاتر در بالای لیست قرار می‌گیرند و شناسایی می‌شوند.

جدول ۲. الگوریتم پیشنهادی

الگوریتم فاز ۱
ورودی الگوریتم: مجموعه داده بیانگر شبکه چندلایه
خروجی الگوریتم: درجه ناهنجاری هر گره در هر لایه شبکه چندلایه
۱- برای هر لایه l از گراف G
۲- برای هر گره i در لایه l

۳- اگونت گره i را استخراج کن
۴- وزن اگونت گره و مقدار ویژه ماتریس مجاورت وزن اگونت گره را محاسبه کن
۵- خط برازش برای فضای ویژگی رسم کن
۶- مقدار عرض از مبدأ و شیب خط ترسیم شده را محاسبه کن
۵- درجه ناهنجاری هر گره i را بر اساس رابطه ۲ محاسبه کن

الگوریتم فاز ۲

ورودی الگوریتم: درجه ناهنجاری هر گره در هر لایه شبکه چندلایه
خروجی الگوریتم: رتبه‌بندی گره‌های شبکه چندلایه بر اساس درجه ناهنجاری نهایی

۱- برای هر گره i
مجموع وزن یال‌ها در اگونت‌های i در تمام لایه‌ها را محاسبه کن

$$T_edge_i = \sum_{l=1}^{L} W_i^l$$

۲- برای هر لایه l
برای هر گره i
اهمیت لایه برای گره i را محاسبه کن

$$LR_i^l = W_i^l \Big/ T_edge_i$$

۳- برای هر گره i درجه ناهنجاری نهایی را طبق رابطه زیر محاسبه کن

$$multiScore_i = \sum_{i=1}^{L} LR_i^l . aScore_i^l$$

۴- درجات ناهنجاری نهایی را به‌ترتیب نزولی مرتب کن

۶- آزمایش های انجام شده

برای تحلیل شبکه‌های اجتماعی چندلایه روشی که در گذشته وجود داشته این بوده که بدون درنظرگرفتن اینکه ارتباطات در لایه‌های مختلفی گسترده شده است، شبکه تجمیع‌شده‌ای از ارتباطات رسم می‌شد و الگوریتم‌های تحلیل یا تشخیص ناهنجاری روی آن‌ها اعمال می‌شد. اما با گذشت سال‌ها دانشمندان متوجه شده‌اند که باید ویژگی چندلایه‌بودن شبکه را نیز درنظر بگیرند زیرا اطلاعات خاصی را در اختیار ما قرار می‌دهد که برای تحلیل‌ها بسیار مفید هستند. بنابراین در این تحقیق نیز از ویژگی چندلایه بودن شبکه برای کشف ناهنجاری ارتباط قوی برجسته استفاده می‌شود. این ناهنجاری تاکنون در شبکه چندلایه بررسی نشده است. ما این ناهنجاری را به دو روش بیان‌شده یعنی به‌صورت گراف تجمیع‌شده بدون درنظرگرفتن لایه‌های مختلف و به‌صورت درنظرگرفتن لایه‌های ارتباطی جداگانه با هم مقایسه می‌کنیم.

۷- محیط شبیه‌سازی

در این بخش برای شبیه‌سازی و تحلیل شبکه‌های اجتماعی چندلایه از زبان برنامه‌نویسی R که یک زبان مناسب برای فعالیت‌های تحقیقاتی در زمینه شبکه‌های اجتماعی است، استفاده می‌شود. نرم‌افزار Rstudio نسخه ۱.۱.۴۵۶ برای کدنویسی استفاده شده است و کتابخانه‌هایی که برای این کدنویسی مورد استفاده قرار گرفتند، igraph و sna هستند.

۸- داده‌های مورداستفاده[۳۴]

داده‌های بسیار کمی از داده‌های موجود برای تحلیل شبکه‌های چندلایه مربوط به شبکه‌های اجتماعی چندلایه ایستا، وزن‌دار و بدون جهت می‌شود، بنابراین داده‌هایی را انتخاب کردیم که ویژگی‌های موردنیاز برای تحلیل را دارند اما ممکن است به شبکه اجتماعی مربوط نباشد. ویژگی‌های این شبکه‌ها در ادامه بیان شده و به‌صورت خلاصه در جدول ۳ آورده شده است.

۸-۱- شبکه حمل‌ونقل چندگانه در لندن[۳۵]

این مجموعه داده را دومینیکو[۳۶] در سال ۲۰۱۳ از سایت رسمی حمل‌ونقل در لندن جمع‌آوری کرده است. گره‌ها در این شبکه، ایستگاه‌های قطار و لبه‌ها، مسیرهای بین ایستگاه‌های مختلف هستند. تعداد گره‌ها در این شبکه پیچیده ۳۶۹ تاست و تعداد مسیرها ۴۴۱ مسیر است. این شبکه شامل ۳ لایه است:

۱) شبکه مسیرهای حمل‌ونقل زیرزمینی ۲) شبکه مسیرهای حمل‌ونقل روی زمین ۳) شبکه DLR

۸-۲- شبکه نویسندگان همکار آرکایو[۳۷]

شبکه چندلایه دیگری که در این تحقیق بررسی می‌شود شبکه نویسندگان همکار در مخزن علمی آزاد آرکایو است. این شبکه شامل مقالاتی تا ماه می سال ۲۰۱۴ است که کلمه "networks" را در عنوان یا چکیده دارند. این شبکه ۱۴۴۸۹ گره دارد و ۵۹۰۲۶ لبه دارد و شامل ۱۳ لایه متناظر با دسته‌بندی‌های مختلف مجتمع آرکایو است. لیست لایه‌ها به شرح زیر است:

1) physics.soc-ph 2) physics.data-an 3) physics.bio-ph
4) math-ph 5) math.OC 6) cond-mat.dis-nn
7) cond-mat.stat-mech 8) q-bio.MN 9) q-bio
10) q-bio.BM 11) nlin.AO 12) cs.SI 13) cs.CV

۸-۳- شبکه فعالیت‌های مختلف پیر ایگر

این شبکه مربوط به وظایف کاری متفاوت پیر ایگر بین سال‌های ۲۰۱۰ تا ۲۰۱۲ می‌باشد و شامل ۵۱۴ گره و ۷۱۵۳ لبه است. همه گزارش‌ها مطابق با کلمات کلیدی و محتوای آن به ۱۶ لایه تقسیم شده‌اند.

1) Neutrinos 2) Detector 3) Enhancements
4) Anisotropy 5) Point-source 6) Mass-composition
7) Horizontal 8) Hybrid-reconstruction
9) Spectrum 10) Photons 11) Atmospheric
12) SD-reconstruction 13) Hadronic-interactions
14) Exotics 15) Magnetic16)Astrophysical-scenarios

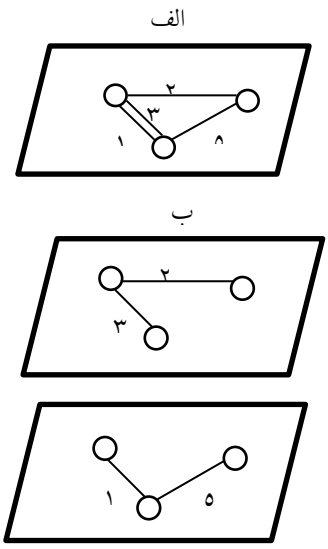
جدول ۳. ویژگی‌های مجموعه داده‌های بررسی‌شده

مجموعه داده	تعداد گره	تعداد لبه	تعداد لایه	توصیف
شبکه حمل و نقل در لندن	۳۶۹	۴۴۱	۳	شبکه مسیرهای ارتباطی بین ایستگاه‌های قطار

شبکه نویسندگان همکار آرکایو	۱۴۴۸۹	۵۹۰۲۶	۱۳	شبکه ارتباطات مختلف نویسندگان همکار
شبکه فعالیت‌های مختلف پیر ایگر	۵۱۴	۷۱۵۳	۱۶	شبکه ارتباطات فعالیت‌های مختلف پیر ایگر

۹- استفاده از گراف شبکه تجمیع شده

در گذشته، برای تحلیل شبکه‌های اجتماعی چندلایه، روش خاصی وجود نداشت. برای همین لایه‌های شبکه چندلایه را تجمیع می‌کردند، یعنی بدون اینکه درنظر بگیرند چند لایه ارتباطی داریم و هر گره در هر لایه چه نوع ارتباطی دارد، گره‌ها را با ارتباطات بینشان در یک لایه فرض می‌کردند (شکل ۴).



شکل ۴. الف) شبکه تجمیع‌شده ب) شبکه چندلایه

شبکه تجمیع‌شده به‌صورت گرافی نمایش داده می‌شود که در آن گره‌ها همان گره‌هایی است که در تمام لایه‌ها به صورت ثابت وجود دارند و ارتباطات، جمع ارتباط گره‌ها در تمام لایه‌ها است که به صورت زیر قابل نمایش است:

$$G = (V,E) \qquad E = E^1 \cup E^2 \ldots \cup E^l \qquad (۵)$$

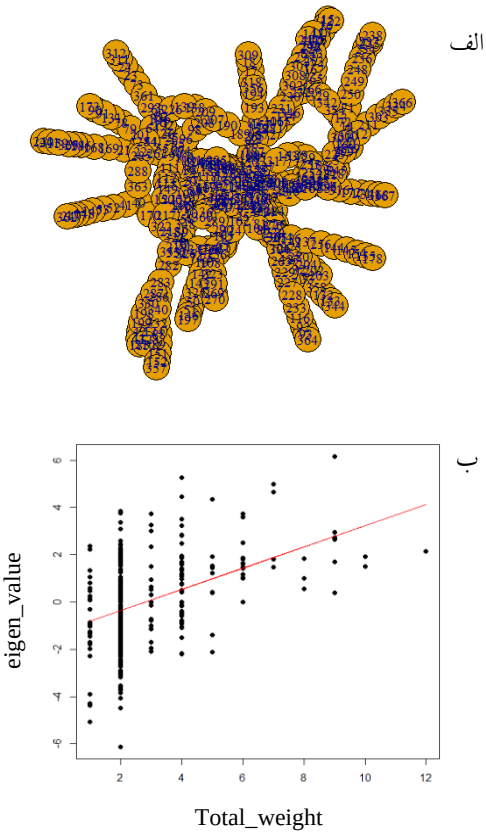
ماتریس مجاورتی که برای این گراف تجمیع‌شده قابل نمایش است به‌صورت رابطه ۶ تعریف می‌شود که در آن اگر ارتباطی بین دو گره وجود نداشته باشد مقدار آن برابر صفر و اگر ارتباطی وجود داشته باشد، مقدار آن جمع وزن‌هاست.

$$a_{ij} = \begin{cases} w & \exists 1 \le l \le L : a_{ij}^{[l]} \ne 0 \\ 0 & \text{در غیر این صورت} \end{cases} \qquad (۶)$$

برای تشخیص ناهنجاری با الگوی ارتباطات قوی برجسته در گراف تجمیع‌شده باید اگونت مربوط به هر گره را استخراج کرد و دو ویژگی وزن کل اگونت و مقدار ویژه ماتریس مجاورت اگونت را محاسبه کرد. با این فضای ویژگی دوبعدی نمودار برازش داده‌ها رسم می‌شود و درجه ناهنجاری برای هر گره با توجه به رابطه ۲ محاسبه می‌شود.

۱۰- نتایج و بحث‌ها

پس از پیاده‌سازی هر دو الگوریتم تجمیع شده و پیشنهادی، مجموعه داده‌های توصیف‌شده به‌عنوان ورودی به این الگوریتم‌ها داده شد. در الگوریتم شبکه تجمیع‌شده، یک گراف و نمودار برازش داده‌ها که توسط برنامه پیشگویی می‌شود، رسم می‌شود. نمونه‌ای از گراف و خط برازش در شکل ۵ دیده می‌شود. از خط پیشگویی‌شده، مقدار شیب و عرض از مبدأ محاسبه می‌شود و بر اساس آن فاصله از خط که همان درجه ناهنجاری است برای هر گره محاسبه می‌شود. درجه ناهنجاری برای تمام مجموعه داده‌ها به‌صورت ۱۰ گره با بالاترین درجه ناهنجاری در جدول ۴ مرتب شده‌اند.



شکل ۵. الف) گراف تجمیع‌شده شبکه حمل‌ونقل لندن ب) خط برازش داده‌ها

اما در الگوریتم پیشنهادی، یک گراف برای هر لایه و نمودار برازش داده‌ها برای هر لایه که توسط برنامه پیشگویی می‌شود رسم می‌شود. از خط پیشگویی‌شده، مقدار شیب و عرض از مبدأ محاسبه می‌شود و بر اساس آن فاصله از خط که همان درجه ناهنجاری است برای هر گره محاسبه می‌شود. سپس میزان اهمیت هر لایه برای هر گره و درجه ناهنجاری نهایی برای گره‌ها در کل شبکه چندلایه محاسبه می‌شود. درجه ناهنجاری برای تمام

مجموعه داده‌ها به‌صورت ۱۰ گره با بالاترین درجه ناهنجاری در جدول ۴ مرتب شده‌اند.

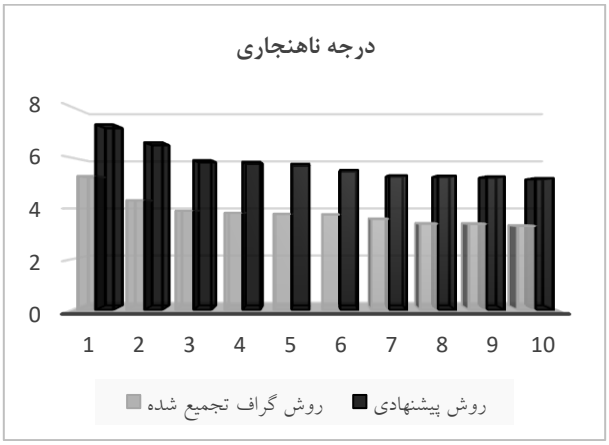
جدول ۴. ده گره اول با رتبه ناهنجاری بالا در روش تجمیع‌شده و رویکرد پیشنهادی برای مجموعه داده‌های بررسی‌شده

مجموعه داده	رتبه	روش تجمیع شده		رویکرد پیشنهادی	
		گره ناهنجار	درجه ناهنجاری	گره ناهنجار	درجه ناهنجاری
شبکه حمل‌ونقل در لندن	۱	۳۶۹	۵.۲۷۳	۲۷۱	۷.۳۰۰
	۲	۲	۴.۳۲۲	۲۷۰	۶.۵۹۶
	۳	۳۶۸	۳.۹۰۹	۲۶۸	۵.۸۹۴
	۴	۷	۳.۸۲۵	۲۶۷	۵.۸۳۲
	۵	۳۶۷	۳.۷۸۵	۲۶۹	۵.۷۵۳
	۶	۸	۳.۷۶۲	۲۶۶	۵.۵۱۳
	۷	۵	۳.۵۸۷	۲۶۴	۵.۳۰۲
	۸	۱۳	۳.۳۹۹	۲۶۳	۵.۲۸۸
	۹	۳۶۴	۳.۳۹۲	۲۶۲	۵.۲۵۹
	۱۰	۱۰	۳.۳۱۱	۲۶۱	۵.۱۹۳
شبکه نویسندگان همکار آرکایو	۱	۴۷۹	۶.۴۰	۵۴۶۵	۲۰.۹۶
	۲	۸۳	۳.۹۴	۵۴۶۴	۱۹.۸۸
	۳	۲۱۸	۳.۹۲	۳۵۰۶	۱۸.۹۵
	۴	۱۷۲	۳.۴۳	۵۴۶۳	۱۸.۹۳
	۵	۵۴	۳.۳۲	۵۴۶۲	۱۸.۹۱
	۶	۷۱۵	۳.۱۴	۵۴۶۱	۱۸.۷۶
	۷	۵۷۸	۲.۸۱	۵۴۶۰	۱۸.۷۳
	۸	۸۰	۲.۱۴	۵۴۵۹	۱۷.۹۳
	۹	۱۷۵۱	۲.۱۳	۵۴۵۸	۱۷.۹۲
	۱۰	۸۴۲	۲.۰۰	۵۴۵۷	۱۷.۶۰
شبکه فعالیت‌های مختلف پیر ایگر	۱	۱	۵.۱۹	۲۳۲	۱۷.۹۹
	۲	۲	۴.۷۷	۲۳۱	۱۷.۲۸
	۳	۳	۴.۴۶	۲۳۰	۱۷.۰۷
	۴	۴	۳.۴۵	۲۲۹	۱۷.۰۱
	۵	۵	۳.۲۶	۲۲۸	۱۶.۹۹
	۶	۶	۳.۱۶	۲۲۷	۱۶.۰۴
	۷	۸	۳.۱۱	۲۲۶	۱۵.۹۹
	۸	۷	۳.۰۲	۲۲۵	۱۵.۸۸
	۹	۵۱۴	۲.۹۳	۲۲۴	۱۵.۸۳
	۱۰	۵۱۳	۲.۸۹	۲۲۲	۱۵.۸۲

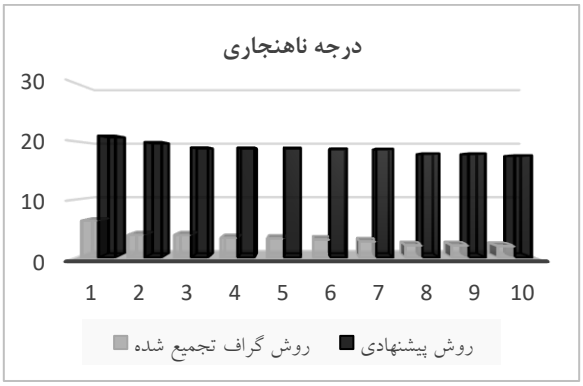
۱۰-۱-تحلیل نتایج به‌دست‌آمده از الگوریتم معیار و پیشنهادی

در این بخش به تحلیل نتایج به‌دست‌آمده از روش معیار و روش پیشنهادی می‌پردازیم. در تمام نمونه داده‌هایی که برای آزمایش الگوریتم‌ها استفاده کردیم شاهد این مسئله هستیم که در شبکه‌های چندلایه باید از ویژگی‌های نهفته در لایه‌ها نیز استفاده کنیم تا با دقت بیشتری درجات ناهنجاری را محاسبه کنیم. در روش استفاده از گراف‌های تجمیع‌شده بدون اینکه درنظر بگیریم روی شبکه چندلایه کار می‌کنیم، لایه‌ها را تجمیع می‌کنیم و با گراف به‌دست‌آمده مانند یک گراف تک‌لایه رفتار می‌کنیم. هر روش تشخیص ناهنجاری که روی گراف‌ها قابل پیاده‌سازی است روی این گراف نیز قابل پیاده‌سازی است و حتماً نتایج نشان می‌دهند که گره‌های ناهنجاری با درجات ناهنجاری مشخص وجود دارند. اما زمانی که از روش پیشنهادی استفاده می‌شود، در حقیقت لایه‌ها جداگانه بررسی می‌شوند. علاوه بر این که درجه ناهنجاری برای هر گره در هر لایه محاسبه می‌شود. یک ضریب به نام اهمیت لایه برای هر گره در هر لایه محاسبه می‌گردد و در ناهنجاری به‌دست‌آمده از هر لایه برای هر گره ضرب می‌شود. به این روش دقت الگوریتم بالا می‌رود. همان‌گونه که در جدول ۴ مشاهده می‌شود، درجات ناهنجاری به‌دست‌آمده در روش معیار اعداد کوچک‌تری نسبت به روش پیشنهادی هستند. این دو الگوریتم را از دو جنبه رتبه ناهنجاری و زمان اجرا می‌توان با هم مقایسه کرد. با توجه به مطالب مذکور در مورد درجات ناهنجاری محاسبه‌شده، دقت این الگوریتم افزایش یافته است. وقتی خروجی هر دو الگوریتم را با هم مقایسه می‌کنیم کاملاً مشهود است که درجه ناهنجاری برای هر گره در روش پیشنهادی، بزرگ‌تر از روش معیار است. نمودارهای مربوطه در شکل‌های ۶ و ۷ و ۸ آورده شده‌اند. این اشکال به‌ترتیب مربوط به مقایسه درجه‌های ناهنجاری به‌دست‌آمده برای ۱۰ گره ناهنجار در شبکه‌های حمل‌ونقل لندن، شبکه نویسندگان همکار آرکایو و شبکه فعالیت‌های پیر ایگر هستند. این نمودارها نشان‌دهنده برتری روش پیشنهادی نسبت به روش معیار در شبکه‌های چندلایه مختلف هستند زیرا در هر مورد، روش پیشنهادی درجه ناهنجاری بزرگ‌تری برای گره‌های ناهنجار به‌دست آورده است.

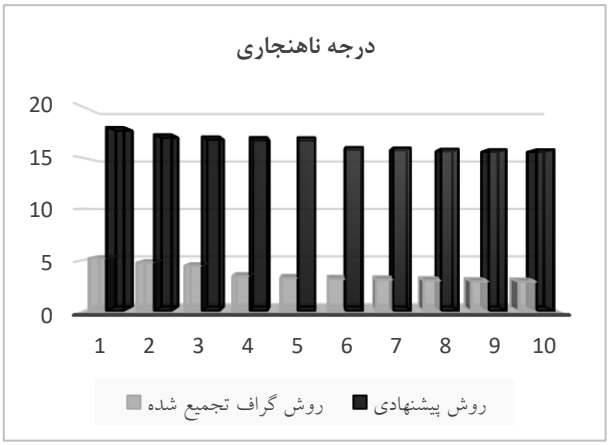
وجود اعداد بزرگ‌تر درجه ناهنجاری در الگوریتم پیشنهادی به علت دخیل بودن ضریب اهمیت یک لایه (یک شبکه اجتماعی) در فرمول محاسبه درجه ناهنجاری است و این موضوع بیانگر این است که در یک لایه خاص، یک کاربر، یک ارتباط بسیار قوی دارد که شک‌برانگیز است و این موضوع در روش گراف تجمیع‌شده پنهان بود زیرا مجموع ارتباطات کاربر در تمام لایه‌ها مد نظر قرار می‌گرفت و از آن‌جایی که جمع ارتباطات کاربر A که در یک شبکه ارتباط بسیار زیادی دارد ولی در شبکه‌های دیگر ارتباط بسیار کمی دارد، عددی را به دست می‌دهد که با مجموع ارتباطات کاربر B که در هر شبکه ارتباطات متوسطی دارد نزدیک یا یکسان باشد و در نتیجه ناهنجاری کاربر A از محاسبات مخفی می‌ماند. بنابراین روش پیشنهادی چون ارتباط کاربر در هر شبکه اجتماعی را به تفکیک بررسی می‌کند نتیجه بهتری ارائه می‌نماید.



شکل ۶. مقایسه درجات ناهنجاری به‌دست‌آمده در روش معیار و پیشنهادی در شبکه حمل‌ونقل لندن



شکل ۷. مقایسه درجات ناهنجاری به‌دست‌آمده در روش معیار و پیشنهادی در شبکه نویسندگان همکار آرکایو



شکل ۸. مقایسه درجات ناهنجاری به‌دست‌آمده در روش معیار و پیشنهادی در شبکه فعالیت‌های پیرایگر

علاوه بر این، از نظر زمان اجرا، الگوریتم گراف تجمیع‌شده زمانی را صرف تجمیع چندلایه می‌کند و پس از آن به دنبال کشف ناهنجاری می‌باشد. اما در رویکرد پیشنهادی، این زمان نیز صرف کشف ناهنجاری می‌شود.

۱۱- نتیجه‌گیری و جمع‌بندی

با گسترش شبکه‌های اجتماعی و فراگیرشدن این شبکه‌ها در کل دنیا، اطلاعات بسیاری در این محیط گسترده شده‌اند. افراد ناهنجار بسیاری در این شبکه‌ها فعالیت می‌کنند که کشف این افراد ناهنجار کار بسیار مهمی است. ناهنجاری که در این تحقیق بررسی می‌شود کاربران ناهنجار با الگوی ارتباطات قوی برجسته است. این ارتباطات اصولاً از بقیه ارتباطات مؤثرتر و مهم‌تر و با وزن بیشتر هستند. البته این ناهنجاری در گذشته در شبکه‌های اجتماعی تک‌لایه بررسی شده است و ما در این تحقیق این ناهنجاری را در شبکه‌های اجتماعی چندلایه بررسی کنیم. برای کشف ناهنجاری مذکور در شبکه چندلایه تاکنون روشی ارائه نشده است. ما دو روش را برای کشف این ناهنجاری پیاده می‌کنیم: یک روش، روش تجمیع لایه‌ها و استفاده از یک روش تشخیص ناهنجاری مناسب است و روش دیگری که ما پیشنهاد کردیم روشی مبتنی بر الگوریتم ADOMS می‌باشد. البته برای محاسبه درجه ناهنجاری در هر لایه، روش متفاوتی را استفاده کردیم و شبکه‌ای که این روش را بر آن پیاده کردیم یک شبکه چندلایه وزن‌دار است. برای اعتبارسنجی رویکرد پیشنهادی ۳ مجموعه داده از دنیای واقعی را که ویژگی چندلایه بودن، غیرجهت‌دار بودن و وزن‌دار بودن را دارا بودند، انتخاب کردیم و روی هر دو روش مورد آزمایش قرار دادیم. نتایج نشان داد روش پیشنهادی نسبت به روش تجمیع لایه‌ها، روش دقیق‌تری است.

رویکرد پیشنهادی روی شبکه‌های ایستا اعمال شده است یعنی شبکه‌هایی که تعداد مشخصی گره در هر لایه دارند. در صورتی که ما شبکه‌هایی مبتنی بر زمان داریم. شبکه‌هایی که در آن‌ها ارتباطات در زمانی برقرار است و در زمانی دیگر قطع می‌شود. می‌توان به عنوان تحقیقی در آینده روی چنین شبکه‌هایی ناهنجاری ارتباطات قوی برجسته را بررسی کرد. به عنوان جهت‌دهی دیگری از این تحقیق می‌توان این مسئله را بیان کرد که تحقیق حاضر فقط به ارتباطات میان‌لایه‌ای توجه کرده است ولی می‌توان از ارتباطات بین‌لایه‌ای نیز استفاده کرد. همچنین در روش پیشنهادی ما برای محاسبه ضریب اهمیت لایه‌ها فقط از وزن ارتباطات در هر لایه استفاده شده است. می‌توان بررسی کرد که چه پارامترهای دیگری برای محاسبه اهمیت لایه می‌تواند مورد استفاده قرار گیرد تا ارزش این پارامتر را بیش از پیش نشان دهد. برای ادامه این تحقیق و ارائه روشی بهتر در آینده، می‌توان از روش کشف زیرگراف‌های غیرنرمال [۲۱] برای کشف کاربران ناهنجار در شبکه‌های اجتماعی چندلایه استفاده کرد.

۱۲- مراجع

[1] D.M. Hawkins, Identification of outliers (Vol. 11). London: Chapman and Hall, 1980.
[2] L. Akoglu, M. McGlohon, and C. Faloutsos, “Anomaly detection in large graphs,” In In CMU-CS-09-173 Technical Report, 2009.
[3] P. V. Bindu, P. S. Thilagam, and D Ahuja, “Discovering suspicious behavior in multilayer social networks,” Computers in Human Behavior, 73, pp. 568-582, 2017.
[4] A. Rawat, G. Gugnani, M. Shastri, and P. Kumar, “Anomaly recognition in online social networks,” International Journal of Security and Its Applications, 9(7), pp. 109-118, 2015.
[5] R. Kaur, and S. Singh, “A comparative analysis of structural graph metrics to identify anomalies in online social networks,” Computers & Electrical Engineering, vol. 57, pp. 294-310, 2017.

[6] W. Eberle, and L. Holder, “Anomaly detection in data represented as graphs,” Intelligent Data Analysis, vol. 11, no. 6, pp. 663-689, 2007.
[7] M. De Domenico, M. A. Porter, and A. Arenas, “MuxViz: a tool for multilayer analysis and visualization of networks,” Journal of Complex Networks, vol. 3, no. 2, pp. 159-176, 2015.
[8] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, and A. Arenas, “Mathematical formulation of multilayer networks,” Physical Review X, vol. 3, no. 4, 041022, 2013.
[9] R. KUMAR, W. H. Cheng, M. Atiquzzaman, K. Srinivasan, and A. Zomaya, “NHAD: Neuro-Fuzzy Based Horizontal Anomaly Detection In Online Social Networks,” IEEE Transactions on Knowledge and Data Engineering, 2018.
[10] R. Mittal, and M. P. S. Bhatia, “Anomaly Detection in Multiplex Networks. Procedia Computer Science,” vol 125, pp. 609-616, 2018.
[11] M. M. Breunig, H. P. Kriegel, R. T. Ng, and J. Sander, “LOF: identifying density-based local outliers,” In ACM sigmod record. Vol. 29, No. 2, pp. 93-104, ACM, 2000.
[12] H. C. Manjunatha, and R. Mohanasundaram, “BRNADS: Big data real-time node anomaly detection in social networks,” In 2018 2nd International Conference on Inventive Systems and Control, ICISC, pp. 929-932, IEEE, 2018.
[13] A. Arleo, W. Didimo, G. Liotta, and F. Montecchiani, “Profiling distributed graph processing systems through visual analytics,” Future Generation Computer Systems, vol 87, pp. 43-57, 2018.
[14] J. L. Izquierdo-Zaragoza, J. J. Pedreno-Manresa, and P. Pavon-Marino, “Net2Plan: An integrated open-source framework for multilayer network planning and in-operation simulation,” In 2015 17th International Conference on Transparent Optical Networks, ICTON, pp. 1-4. IEEE, 2015.
[15] Rossi, Luca, and Matteo Magnani. "Towards effective visual analytics on multiplex and multilayer networks." Chaos, Solitons & Fractals 72 (2015): 68-76.
[16] G. Bergami, F. Bertini, and D. Montesi, “On Approximate Nesting of Multiple Social Network Graphs: a preliminary study,” 2019.
[17] B. Guidi, A. Michienzi, K. Koidl, and K. Kapanova, “A multilayer social overlay for new generation DOSNs.” In Proceedings of the 5th EAI International Conference on Smart Objects and Technologies for Social Good, pp. 114-119, ACM, 2019.
[18] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” ACM computing surveys (CSUR), vol 41, no. 3, 15, 2009.
[19] F. Amato, A. Castiglione, A. De Santo, V. Moscato, A. Picariello, F. Persia, and G. Sperlí, “Recognizing human behaviours in online social networks,” Computers & Security, vol. 74, pp. 355-370, 2018.
[20] M. R. Islam, S. Liu, X. Wang, and G. Xu, “Deep learning for misinformation detection on online social networks: a survey and new perspectives,” Social Network Analysis and Mining, vol. 10, 1, 1-20, 2020.
[21] Y. Sun, W. Wang, N. Wu, W. Yu, and X. Chen, “Anomaly Subgraph Detection with Feature Transfer,” In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 1415-1424, 2020.

سیده سودابه غفوری مدرک کارشناسی خود را در رشته‌ی مهندسی کامپیوتر- نرم‌افزار از دانشگاه آزاد لاهیجان دریافت کرده است. او همچنین دارای مدرک کارشناسی ارشد مهندسی فناوری اطلاعات-شبکه‌های کامپیوتری از دانشگاه علوم و تحقیقات تهران است. زمینه‌های تحقیقاتی ایشان شامل شبکه-های کامپیوتری و گراف کاوی است. او همچنین پس از دانش آموختگی تا کنون در شرکت های معتبر نرم افزاری مشغول به کار بوده است.

محمود دی پیر مدرک دکترای خود را در رشته ی کامپیوتر-سیستم های نرم افزاری و مدرک کارشناسی ارشد خود را در رشته ی کامپیوتر نرم افزار هر دو از دانشگاه شیراز دریافت کرده است. مقطع کارشناسی خود را نیز در همین رشته از دانشگاه هوایی شهید ستاری دریافت کرده است. هم اکنون عضو هیئت علمی دانشکده رایانه و فناوری اطلاعات دانشگاه هوایی شهید ستاری است. زمینه های تحقیقاتی ایشان شامل داده کاوی و امنیت فضای سایبر است و دارای مقالات متعددی در مجلات و کنفرانس های معتبر ملی و بین المللی است. نامبرده در پروژه های تحقیقاتی و صنعتی متعددی مشارکت داشته است.

[1] Online Social Networks (OSN)
[2] Malicious users
[3] Anomaly Detection
[4] Anomalous User
[5] Anomalous Interaction
[6] Anomalous Community
[7] Anomalous Events
[8] Star & Near-Star
[9] Clique & Near-Clique
[10] Heavy Vicinity
[11] Dominant Heavy Link
[12] Egonet
[13] Structure-based
[14] Behavior-based
[15] Spectral-based
[16] Nodes
[17] Edges
[18] Heuristic
[19] Local Outlier Factor (LOF)
[20] Brokerage
[21] Average Betweenness Centrality
[22] Minimum Description Length
[23] Fake Accounts
[24] Outdegree
[۲۵] یک محیط کار سریع و مقیاس پذیر تبادل پیام میان برنامه‌ها است.
[۲۶] یک محیط کار برای محاسبات توزیع شده در داده‌های حجیم است.
[27] Think-Like-A-Vertex

[۲۴] داده‌ها از سایت comunelab.com انتخاب شده‌اند.

[28] Pie graph
[29] Tensorial Framework
[30] Adjacency tensor
[31] Decentralized Online Social Networks (DOSNs)
[32] Horizontal Anomaly
[33] Cross-Layer Anomaly Detection
[35] London Multiplex Transport Network
[36] Manlio De Domenico
[37] arXiv