نشریه علمی پدافند غیرعامل

سال دهم، شماره ۳، پاییز ۱۳۹۸، (پیاپی ۳۹): صص ۱۳-۱

مروری بر روش‌های تشخیص ناهنجاری مبتنی‌بر گراف در شبکه‌های اجتماعی

میثم میرزایی[۱]، امین‌اله مه‌آبادی[۲*]

تاریخ دریافت: ۱۳۹۷/۰۸/۱۵
تاریخ پذیرش: ۱۳۹۷/۱۲/۱۵

چکیده

استفاده از شبکه‌های اجتماعی برای برقراری ارتباط و اشتراک‌گذاری اطلاعات رشد چشم‌گیری در سال‌های اخیر داشته و در اکثر حوزه‌های آموزش، کسب و کار، سلامت و سرگرمی کاربرد دارند. حجم زیاد اطلاعات با ارزش در شبکه‌های اجتماعی آن‌ها را هدف اصلی کاربران مخرب همچون هرزنامه‌نویس‌ها و کلاه‌برداران برای انجام فعالیت‌های ناهنجار و غیرقانونی قرار داده است. رفتار نامتعارف و دور از انتظار این کاربران با استفاده از روش‌های تشخیص ناهنجاری شناسایی می‌شود. تشخیص ناهنجاری اهمیت بسزایی در جلوگیری از کلاه‌برداری، انتشار اطلاعات تقلبی و سازمان‌دهی حملات در این شبکه‌ها دارد. ناهنجاری‌ها ایستا یا پویا و با ویژگی یا بدون ویژگی هستند. در این مقاله روش‌های مختلف توسعه‌یافته برای تشخیص انواع ناهنجاری در شبکه‌های اجتماعی مورد بررسی و دسته‌بندی قرار گرفته و مروری بر تشخیص ناهنجاری، کاربردهای آن، چالش‌های موجود و موضوعات کلیدی برای پژوهش‌های آینده ارائه شده است.

کلید واژه‌ها: شبکه‌های اجتماعی، تشخیص ناهنجاری، تحلیل شبکه‌های اجتماعی

۱- دانشجوی کارشناسی ارشد هوش مصنوعی دانشگاه ایوانکی
۲- استادیار دانشگاه شاهد، (mahabadi@shahed.ac.ir)- نویسنده مسئول

۱- مقدمه

جهان امروز وارد عصر جدیدی شده است که مهم‌ترین نشانه آن، اهمیت یافتن اطلاعات و ارتباطات و تغییر در شیوه انتقال و ایجاد آن است. امروزه یکی از فراگیرترین و مهم‌ترین شیوه‌های ارتباطی، وب و فنّاوری‌های مبتنی بر آن است. فضای مجازی، عرصه‌ای بسیار وسیع برای تبادل اطلاعات و برقراری ارتباط است که به‌دلیل سرعت و تنوع در پردازش اطلاعات، بسیار مورد توجه مخاطبان قرار گرفته است [۱]. با پیشرفت شبکه‌های اجتماعی برخط[1] و دیجیتال شدن بسیاری از ارتباطات، شبکه‌های اجتماعی برخط بخش مهمی از کاربردهای اجتماعی ارتباطات را شامل شده‌اند. این شبکه‌ها برای تعامل با دوستان مورد علاقه و آشنایان و به اشتراک‌گذاری اطلاعات مورد استفاده قرار می‌گیرد [۲]. پویایی این شبکه‌ها در ساختار و تعاملات بین افراد، باعث افزایش محبوبیت و گسترش آن‌ها شده است. طبق آمار در سال‌های اخیر بیش از ۷۰ درصد از کاربران اینترنت عضو شبکه‌های اجتماعی مختلف هستند [۳]. هم‌چنین این شبکه‌های در حوزه‌های آموزش، کسب و کار، سلامت و سایر حوزه‌ها مورد استفاده قرار می‌گیرند [۴]. تبادل حجم زیاد اطلاعات و سهولت هدایت آن از طریق شبکه، علاقه کاربران مخرب[2] را نیز به خود جلب کرده است و به یک هدف کلیدی برای این افراد جهت کسب سودهای غیرقانونی یا آسیب رساندن به سایر کاربران تبدیل شده است. گمنامی نسبی و بدون نظارت تعاملات در بسیاری از سامانه‌های برخط ابزار مناسبی را برای شکارچیان جنسی برای درگیر کردن جوانان با این مسائل فراهم کرده است [۵]. خرده‌فروشی‌ها و مزایده‌های برخط هدف مناسبی برای کلاه‌برداران هستند که با نمایش خود به‌عنوان کاربران مورد اعتماد مشتریان را در دام پرداخت‌های سنگین برای کالاهایی که هرگز تحویل نخواهند شد قرار می‌دهند [۶]. طبق اعلام فیسبوک تنها در سه ماه ابتدایی ۲۰۱۸ بیش از ۵۸۳ میلیون حساب جعلی در این شبکه غیرفعال و نزدیک ۸۳۷ میلیون اسپم شناسایی شده است[3]. حملات سایبری، جرائم سازمان‌یافته، انتشار اطلاعات تقلبی و حتی حملات تروریستی برنامه‌ریزی‌شده از سایر تهدیدات شبکه‌های اجتماعی به شمار می‌آیند که ناهنجاری‌های بسیاری را ایجاد کرده است. یکی از جنبه‌های مهم شبکه‌های اجتماعی کشف ناهنجاری‌ها[4] است. ناهنجاری‌ها نمونه داده‌هایی هستند که به میزان قابل‌توجهی با سایر نمونه داده‌ها متفاوت و ناسازگار هستند [۷]. ناهنجاری‌ها هم‌چنین پرت‌ها[5]، اختلالات، مشاهدات غیرواقعی و استثنائات نیز نامیده می‌شوند [۸]. در تعریف دیگری، ناهنجاری را به‌عنوان مشاهده یا زیرمجموعه‌ای از مشاهدات می‌داند که تا حدی زیادی از دیگر مشاهدات متفاوت است [۹]. منشأ ناهنجاری‌ها می‌تواند رفتار کلاه‌بردارانه، خطای انسانی یا شکست سامانه‌ها باشد [۱۰]. افزایش استفاده از شبکه‌های اجتماعی برای مقاصد غیرقانونی، یافتن راه‌کارهای ایجاد امنیت برای کاربران عادی را به چالش تبدیل کرده است. ناهنجاری‌های شبکه‌های اجتماعی، دلالت بر رفتار نامنظم و دور از انتظار دارد [۱۱]. تحلیل رفتار کاربران در شبکه‌های اجتماعی، یک روش استاندارد برای تشخیص ناهنجاری است [۱۲]. کاربران می‌توانند ماهیت خود را با ارائه اطلاعات غلط پنهان کرده و تحلیل‌گر را فریب دهند. استفاده از داده‌هایی همانند پیوندهای[6] ایجادی توسط کاربر که امکان جعل توسط وی را ندارد می‌تواند نتایج کشف و تشخیص ناهنجاری را معتبر سازد. این داده‌ها می توانند با استفاده از گراف برای نمایش شبکه مدل شوند که گره‌ها[7] مبین افراد و یال‌ها[8] پیوندهای آنان به یکدیگر را نشان می‌دهند [۱۳]. پیوندها می‌توانند شامل طیفی از ارتباطات از قبیل دوستی، وابستگی، خانواده و مانند آن باشند. رفتار افراد در این شبکه‌ها تنها از ویژگی‌های فردی آن‌ها نشأت نمی‌گیرد بلکه متأثر از الگوی تعاملی[9] بین آن‌ها است. ازاین‌رو این رفتارها اغلب اوقات الگوهای تعاملی را نشان می‌دهند که متفاوت با رفتار کاربران قانونی است. تشخیص این الگوهای تعاملی متفاوت می‌تواند از طریق روش‌های تشخیص ناهنجاری[10] صورت گیرد.

در این مقاله ضمن بررسی مفاهیم پایه در تشخیص ناهنجاری مروری کلی بر روش‌های تشخیص ناهنجاری در شبکه‌های اجتماعی برای دست‌یابی به اهدافی از قبیل معرفی کاربردهای تشخیص ناهنجاری در شبکه‌های اجتماعی، معرفی روش‌های تشخیص ناهنجاری در این شبکه، ارائه چالش‌های نظری و عملی موجود در این حوزه و مسائل مطرح برای کارهای آتی است.

در این مقاله ابتدا در بخش ۲ کاربردهای تشخیص ناهنجاری در شبکه‌های اجتماعی را تشریح می‌شود. در بخش ۳ مفاهیم پایه و جنبه‌های مختلف تشخیص ناهنجاری معرفی می‌گردد. در بخش ۴ روش‌های توسعه‌یافته بررسی و دسته‌بندی شده و در نهایت در بخش ۵ چالش‌های موجود و کارهای آتی نتیجه‌گیری می‌گردد.

[1] Online Social Networks (OSNs)
[2] Malicious Users
[3] Https://newsroom.fb.com/news/2018/05/enforcement-numbers/
[4] Anomalies
[5] Outliers
[6] Links
[7] Nodes
[8] Edges
[9] Pattern of Interaction
[10] Anomaly Detection

۲- کاربردهای تشخیص ناهنجاری در شبکه‌های اجتماعی

ناهنجاری‌های شبکه‌های اجتماعی، دلالت بر رفتار نامنظم و دور از انتظار دارد [۱۱]. این رفتارها می‌توانند به‌صورت برنامه‌ریزی‌شده و با اهداف خاص صورت پذیرد که نتیجه آن ایجاد تهدیدات امنیتی برای کاربران عادی در شبکه خواهد بود. تشخیص ناهنجاری می‌تواند باعث شناسایی و جلوگیری از وقوع چنین تهدیداتی در شبکه شود. موارد کاربردی تشخیص ناهنجاری در شبکه‌های اجتماعی در ادامه بیان شده است.

تشخیص کلاه‌برداری[1]: کاربران مخرب تلاش برای انجام فعالیت‌های غیرقانونی مانند حملات سایبری، درگیری، ارسال اطلاعات جعلی، جرائم سازمان‌یافته و حتی حمله تروریستی برنامه‌ریزی‌شده در شبکه‌های اجتماعی دارند که این فعالیت‌های جعلی اغلب باعث خسارت و آسیب به کاربران این بستر برخط است [۱۴]. ازآنجاکه الگوهای تعاملی عاملان این فعالیت‌های غیرقانونی اغلب به‌طور قابل‌توجهی از کاربران متفاوت است، می‌توانند با استفاده از روش‌های تشخیص ناهنجاری مبتنی بر شبکه تشخیص داده شوند.

تشخیص تهدید داخلی[2]: تهدید داخلی، تهدید امنیتی به یک سازمان از طرف افراد درون آن سازمان است. این تهدید می‌تواند، کلاه‌برداری، دزدی اطلاعات حساس و تخریب یا به خطر افتادن منابع سخت‌افزاری و نرم‌افزاری باشد. روش‌های تشخیص ناهنجاری مبتنی بر شبکه می‌توانند برای شناسایی تهدیدات داخلی به‌کار گرفته شوند. برای مثال در [۱۷-۱۵] چندین روش مبتنی بر شبکه اجتماعی برای شناسایی تهدیدات داخلی در سامانه‌های اطلاعاتی مشارکتی[3] (CIS)، مانند سامانه‌های پرونده سلامت الکترونیک، بر اساس دسترسی به وقایع محیط‌های مشارکتی ارائه شده است. یک CIS، اطلاعات حساس و مهمی را در یک محیط مشارکتی و پویا مدیریت می‌کند.

تشخیص هرزنامه[4]: امروزه خواندن نظرات یا بررسی‌های نگارش شده توسط مشتریان قبل از خرید یک محصول بسیار متداول شده است. این نظرات همچنین توسط تولیدکنندگان برای شناسایی مشکلات و مسائل مرتبط با محصولاتشان استفاده می‌شود. در بسیاری از فروشگاه‌های برخط همانند آمازون، افراد مخرب با درج نظرات جعلی یا ساختگی به بدنام کردن و یا افزایش اعتبار تولیدکنندگان و فروشندگان و گمراه کردن مشتریان می‌پردازند [۱۸]. تشخیص این نظرات جعلی بسیار مهم است زیرا می‌تواند سبب از دست دادن یا به دست آوردن اعتبار مالی برای تولیدکنندگان و فروشندگان شود.

تشخیص حداکثر نفوذ[5]: علاوه بر شناسایی کلاه‌برداری و فعالیت‌های مخرب، تشخیص ناهنجاری هم‌چنین برای شناسایی افراد بانفوذ و رویدادهای نادر در شبکه مفید است. برای مثال، یک الگوی دوستی نامعمول مانند همبندی[6] ستاره در شبکه‌های اجتماعی می‌تواند مربوط به افراد مشهور یا بانفوذ آن شبکه باشد [۱۲]. این موضوع قابلیت شناسایی بازاریابی‌های ویروسی[7] را می‌دهد که در آن تولیدکنندگان به‌وسیله افراد مشهور به تبلیغ محصول خود می‌پردازند.

تشخیص کلاه‌برداری‌های مالی[8]: در این سامانه‌ها افراد کلاه‌بردار با کارهایی چون انجام معاملات صوری با خود سعی در دست‌کاری در بازار و افزایش ارزش سهام خود دارند. این فعالیت‌های غیرقانونی به‌عنوان حلقه‌های تجاری شناخته شده و با واکاوی شبکه‌های استخراج‌شده از معاملات تجاری در مرحله‌های متوالی قابل شناسایی هستند.

تشخیص کلاه‌برداری در فروش‌های برخط[9]: قابلیت‌هایی چون تنوع محصول و سهولت مقایسه باعث رشد توجه به فروشگاه‌های برخط شده است. هم‌زمان با گسترش استفاده، تلاش برای کلاه‌برداری و انجام فعالیت‌های غیرقانونی و مخرب در این سامانه‌ها نیز افزایش یافته است. فروش بدون تحویل شایع‌ترین کلاه‌برداری است که در آن فروشنده علی‌رغم دریافت پول محصول را تحویل خریدار نمی‌دهد. این فعالیت‌ها با تحلیل تعاملات و رفتار افراد در شبکه قابل شناسایی هستند.

۳- مفاهیم پایه ناهنجاری

۳-۱- انواع ناهنجاری

ناهنجاری‌ها از چند نظر قابل دسته‌بندی هستند. از نظر ماهیت، ناهنجاری‌ها به ۴ دسته نقطه‌ای[10]، جمعی[11]، زمینه‌ای[12] و افقی[13] تقسیم می‌شوند [۱۹]. زمانی که یک نمونه داده خاص الگوی

[1] Fraud Detection
[2] Insider thread detection
[3] Collaborative information systems (CIS)
[4] Spam detection

[5] Influence maximization detection
[6] Topology
[7] Viral marketing
[8] Financial Fraud Detection
[9] Online Auction fraud detection
[10] Point anomaly
[11] Collective anomaly
[12] Contextual anomaly
[13] Horizontal anomaly

معمول مجموعه داده[1] را نقض کند، ناهنجاری نقطه‌ای به وجود می‌آید. ناهنجاری جمعی رفتار نامتعارف و غیر عادی جمعی از داده‌های مشابه نسبت به سایر نمونه‌های مجموعه داده است و رفتار غیرعادی یک نمونه داده در یک زمینه خاص با سایر نمونه‌های مجموعه داده یک ناهنجاری زمینه‌ای است. تشخیص این نوع از ناهنجاری نیاز به شناخت زمینه مورد نظر دارد و به همین دلیل ناهنجاری شرطی[2] نیز نامیده می‌شود. ناهنجاری افقی مختص به شبکه‌های اجتماعی است و تفاوت شدید رفتار یک کاربر در یک شبکه نسبت به رفتار وی در سایر شبکه‌ها است. از نظر نوع شبکه، ناهنجاری‌ها به یکی از دسته‌های ایستای بدون ویژگی[3]، ایستای با ویژگی[4]، پویای بدون ویژگی[5] و پویای با ویژگی[6] تقسیم می‌شوند [۲۰]. در ناهنجاری‌های ایستای بدون ویژگی، هر اطلاعاتی راجع به نوع تعامل، مدت‌زمان آن، سن افراد درگیر و غیره نادیده گرفته می‌شود و تنها تعامل اتفاق افتاده بین افراد قابل‌توجه است. در ناهنجاری‌های ایستای با ویژگی، علاوه بر ساختار شبکه، مشخصات مرتبط با افراد و تعامل بین آن‌ها نیز در تشخیص ناهنجاری‌ها در نظر گرفته می‌شود. ناهنجاری پویا با توجه به رفتار و وضعیت قبلی شبکه قابل تعریف است که اگر تنها ساختار شبکه در زمان‌های مختلف مورد بررسی قرار گیرد ناهنجاری پویای بدون ویژگی قابل شناسایی است و اگر علاوه بر ساختار، ویژگی‌های افراد و تعاملات نیز استفاده شده باشند ناهنجاری پویای با ویژگی تعریف می‌شود.

ناهنجاری‌ها از نظر موقعیت در شبکه، عمومی[7] یا محلی[8] هستند. اگر ناهنجاری با کل شبکه مرتبط باشد ناهنجاری عمومی خواهد بود و اگر تنها به بخشی از شبکه مربوط باشد ناهنجاری محلی تعریف می‌شود. بر اساس نوع رفتار نیز ناهنجاری‌ها به دو دسته آشکار یا پنهان تقسیم می‌شوند [۲۱]. ناهنجاری‌های آشکار تفاوت قابل‌توجهی با نمونه‌های معمول دارند در حالی‌که ناهنجاری‌های پنهان کمی متفاوت هستند. معمولاً نفوذها برای جلوگیری از شناسایی به‌صورت ناهنجاری پنهان رخ می‌دهند.

۳-۲- واحدهای ناهنجار

شامل گره ناهنجار[9]، یال ناهنجار[10]، زیرگراف ناهنجار[11] و رویداد ناهنجار[12] است. کاربر یا کاربرانی که رفتار آن‌ها به مقدار قابل‌توجهی با رفتار معمول شبکه متفاوت باشد گره ناهنجار هستند. تعاملات غیرمعمول یا نامنظم بین کاربران در شبکه یال‌های ناهنجار را نشان می‌دهد [۲۲]. بخشی از شبکه که الگوی تعاملی در آن نسبت به سایر شبکه متفاوت و نامنظم باشد زیرگراف ناهنجار است [۲۳]. لحظه‌ای که در آن ساختار شبکه تفاوت قابل‌توجهی با سایر زمان‌ها داشته و نشان‌دهنده برهم خوردن نظم شبکه باشد رویداد ناهنجار است [۲۴].

۳-۳- کشف ناهنجاری

در تحلیل داده‌های پیچیده یافتن نمونه داده‌های خارج از ساختار متعارف موضوعی بسیار مهم به شمار می‌آید. شاخه‌ای از داده‌کاوی که به دنبال یافتن چنین نمونه‌هایی در مجموعه‌ای از داده‌ها باشد تشخیص ناهنجاری نامیده می‌شود. استفاده از تشخیص ناهنجاری می‌تواند باعث جلوگیری از افشای اطلاعات حساس، جلوگیری از دسترسی‌های غیرمجاز، و پیشگیری از تصمیم‌گیری‌های خطا شود. تشخیص ناهنجاری همواره به‌عنوان یک موضوع مهم در تحلیل داده‌ها مطرح بوده که از قرن ۱۹ مورد مطالعه قرارگرفته است [۲۵]. چگونگی نمایش ناهنجاری‌ها در خروجی یک مساله مهم در کشف ناهنجاری است که به‌صورت کلی به یکی از دو روش امتیاز[13] یا برچسب[14] صورت می‌پذیرد [۸]. روش‌های مبتنی بر امتیاز به هر نمونه داده یک امتیاز ناهنجاری اختصاص می‌دهند. سپس نمونه‌ها بر اساس امتیاز مرتب شده و تحلیل‌گر می‌تواند ناهنجاری‌ها را انتخاب کند و انتخاب معمولاً با استفاده از یک مقدار آستانه[15] صورت می‌گیرد. در روش‌های نوع دوم خروجی‌ها یک برچسب دودویی به معنای آن‌که هر نمونه داده هنجار یا ناهنجار است. این روش‌ها اجازه انتخاب ناهنجاری‌ها را به‌طور مستقیم به تحلیل‌گر نمی‌دهند بلکه این کار را از طریق انتخاب معیارهای خاص کنترل می‌کنند. روش‌های برچسب‌گذاری به دلیل عدم نیاز به فراهم کردن امتیاز برای هر نمونه داده از لحاظ محاسباتی کارآمدتر هستند.

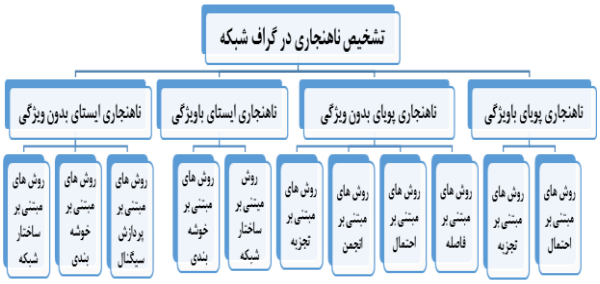
از دیدگاه داده‌کاوی و بر اساس میزان برچسب داده‌های در دسترس، روش‌های تشخیص ناهنجاری به سه دسته بانظارت[16]، نیمه نظارتی[17] و بدون نظارت[18] تقسیم می‌شوند. روش‌های بانظارت مبتنی بر استفاده از مجموعه داده‌هایی هستند که در آن‌ها نمونه‌های برچسب‌گذاری شده از داده‌های هنجار و ناهنجار

[1] Dataset
[2] Conditional anomaly
[3] Static unattributed anomaly
[4] Static attributed anomaly
[5] Dynamic unattributed anomaly
[6] Dynamic attributed anomaly
[7] Global
[8] Local
[9] Anomalous vertex
[10] Anomalous edge
[11] Anomalous subgraph
[12] Anomalous event
[13] Score
[14] Label
[15] Threshold
[16] Supervised Anomaly Detection
[17] Semi-supervised Anomaly Detection
[18] Unsupervised Anomaly Detection

وجود دارد. ایـن روش‌هـا مـدلی از داده‌هـا سـاخته و از آن بـرای تعیین نوع نمونه‌های مشاهده نشده استفاده می‌کنند. این روش‌ها با دو چالش مهم (۱) عدم تعـادل مجموعـه‌هـای داده کـه در آن تعداد نمونه‌های ناهنجار بسیار کمتر از نمونه‌های هنجـار اسـت و (۲). تعیین خروجی دقیـق بخصـوص بـرای نمونـه‌هـای ناهنجـار مواجه هستند. مجموعه داده‌های مورد اسـتفاده در ایـن روش‌هـا تنها حاوی نمونه‌های هنجار هستند. این روش‌هـا مـدلی از رفتـار هنجار را ساخته و با مقایسه نمونه‌های جدید با مـدل، ناهنجـاری را مشخص می‌کنند. روش‌های بدون نظارت نیاز به مجموعه داده-های برچسب‌گذاری شده ندارند. به این علت از روش‌هـای بـدون نظارت به‌صورت گسترده‌ای استفاده می‌شود. فرض ضمنی در این روش‌ها آن است که نمونه‌های هنجار بسیار بیشتر از ناهنجاری‌ها هستند که اشتباه بودن این فرضیه باعث افـزایش تشـخیص‌هـای نادرست خواهد شد.

۴- روش‌های تشخیص ناهنجاری

تشخیص ناهنجاری در شبکه‌های اجتماعی بـه‌منظـور شناسـایی واحدهای ناهنجار[1] در غالب گره‌ها، یال‌ها، زیرگراف‌ها یا رویدادها با توجه به ساختار شبکه صـورت مـی‌گیـرد. واحـدهای ناهنجـار می‌توانند به‌صورت منفرد باشند یا گروهی مجتمـع شـوند. تغییـر ناگهانی در متوسط ارتباط ممکن است باعث در نظر گرفتن گـره منفرد به‌عنوان گره ناهنجار شود. گروهی از افراد متقلب که بـرای افزایش شهرت خود در یک سامانه حراج برخط همکاری می‌کنند یک واحد ناهنجار گروهی در مقیاس زیر شـبکه هسـتند. در ایـن حالت ممکن است رفتار هریک از گره‌ها به تنهایی ناهنجار نباشد اما قرار گرفتن آن‌ها در کنار یکدیگر یک ناهنجاری به‌وجـود آورد [۲۳]. روش‌های بسیاری بـرای تشـخیص انـواع ناهنجـاری‌هـا در شبکه‌های اجتماعی پیشنهادشده‌اند که بر اساس نوع ناهنجاری به شش دسته مبتنی بر ساختار شبکه[2]، مبتنی بـر خوشـه‌بنـدی یـا انجمن[3]، مبتنی بر پردازش سیگنال[4]، مبتنی بر تجزیه[5]، مبتنی بر احتمال[6] و مبتنی بر فاصله[7] مطابق شکل (۱) در قالب پویا و ایستا قابل دسته‌بندی هستند.



شکل (۱): دسته‌بندی روش‌های تشخیص ناهنجاری در گراف شبکه‌های اجتماعی

در روش‌های مبتنی بر ساختار با استفاده از ویژگی‌های ساختاری گراف شبکه به دلیل آن‌که تعداد گره‌ها، تعداد یال‌ها، درجه گـره-ها، مقادیر ویژه[8]، مرکزیت بینابینی[9] و مانند آن بـه‌عنـوان الگـوی متعارف شبکه در نظر گرفته می‌شوند اگر گره‌ها یـا زیرسـاختارها انحراف قابل‌توجهی از این الگو داشته باشند به‌عنـوان ناهنجـاری کشف می‌گردند. در رویکردهای مبتنی بر خوشه‌بندی، ناهنجاری-ها بعد از خوشه‌بندی شبکه شناسایی می‌شوند که بـا اسـتفاده از خوشه‌بندی گره‌های مشابه تشکیل انجمن‌های به هـم پیوسـته را می‌دهند. در این روش‌ها فرض می‌شود کـه اعضـای یـک انجمـن روند تکاملی یکسان و رفتار شبیه به هم دارنـد و تـأثیر تغییـرات انجمن بر روی همه اعضا یکسان خواهد بود. بـا اسـتخراج حالـت معمول و مورد انتظار هر انجمـن، نمونـه‌هـایی کـه از آن پیـروی نکنند به‌عنوان ناهنجار مشخص می‌شوند. موفقیت ایـن دسـته از روش‌ها بسیار وابسته به تشخیص مناسب خوشـه‌هـا و انجمـن‌هـا است. رویکردهای مبتنی بر پردازش سیگنال از جدیدترین کارهـا در حوزه تشخیص ناهنجاری‌ها باهـدف تشـخیص زیرگـراف‌هـای کوچک ناهنجار هستند که زیرگراف‌های ناهنجار نسبتاً کوچک را به‌عنوان سیگنال‌های[10] تعبیه‌شده در یک شبکه بـزرگ بـه‌عنـوان نویز[11] پس‌زمینه در نظر می‌گیرد.

روش‌های مبتنی بر تجزیه بر پایه نمایش گراف به‌عنوان یـک بـردار تنسـور[12] و انجـام عملیـات کـاهش ابعـاد بـرای تشـخیص ناهنجاری عمل می‌کنند. این روش‌ها به دنبـال یـافتن الگـوی یـا نظم خاصی در شبکه هستند تا با استفاده از آن موارد غیرعادی و نامنظم را شناسایی کننـد. در روش‌هـای مبتنـی بـر احتمـال بـا استفاده از روش‌های آماری یک مدل از الگوی رفتار عادی یا مورد انتظار شبکه ساخته و موارد نقض این الگـو بـه‌عنـوان ناهنجـاری معرفی می‌شود. برای ساخت ایـن مـدل معمـولاً از چنـد تصـویر

[1] Anomalous units
[2] Network structure-based approaches
[3] Clustering /Community-based approaches
[4] Signal processing-based approaches
[5] Decomposition based
[6] Probability based
[7] Distance based

[8] Eigen value
[9] Betweenness centrality
[10] Signal
[11] Noise
[12] Tensor vector

لحظه‌ای شبکه در زمان‌های قبل استفاده می‌شود. مدل مورد استفاده، چگونگی ساخت آن و روش یافتن ناهنجاری باعث تمایز روش‌های مختلف مبتنی بر احتمال هستند. روش‌های مبتنی بر فاصله با در نظر گرفتن ویژگی‌های ساختاری شبکه به‌عنوان شاخص، فاصله بین گره‌ها یا یال‌ها را محاسبه و ناهنجاری‌ها را مشخص می‌کنند.

۴-۱- ناهنجاری ایستای بدون ویژگی

به دلیل بدون ویژگی بودن گره‌ها و یال‌ها در این روش‌ها، نوع تعامل، مدت آن، سن افراد و سایر مشخصات برای تشخیص ناهنجاری به‌کار گرفته نمی‌شوند و تنها ارتباط ایجادشده مورد بررسی قرار می‌گیرد. در روش ارائه‌شده توسط شیرواستاوا و همکارانش [۲۶]، با فرض اینکه اسپم و ویروس معمولاً از سوی یک کاربر مخرب منفرد به تعداد زیادی از کاربران هدف ارسال می‌شود از ساختار ستاره‌ای به‌عنوان ناهنجاری استفاده می‌کند. برای تشخیص ناهنجاری‌ها از تعدادی مثلث در هر شبکه egonet (شبکه‌ای از همسایگان درجه اول گره) برای برچسب‌گذاری کاربران استفاده می‌کند و تعداد کم مثلث نشان‌دهنده یک کاربر مخرب است. روش OddBall [۲۷] از ویژگی‌های egonet گره مانند درجه ego (گره)، وزن کلی، تعداد یال‌ها در egonet و مقادیر ویژه اصلی[1] از ماتریس مجاورت وزن‌دار egonet برای تشخیص ناهنجار استفاده می‌کند. این روش ساختارهای شبه ستاره[2] و شبه توده[3] را نشان‌دهنده رفتار مخرب و ناهنجار معرفی می‌کند. دوستاری و همکارانش روشی ارائه کرده‌اند [۲۸]که ساختارهای توده‌ای[4] را از گراف استخراج و از آن‌ها برای یافتن ناهنجاری‌ها جستجو می‌کند. این روش میزان ناهنجاری هر عضو توده را نسبت به سایرین در مشخصه‌هایی مانند تعداد مطلب اشتراکی، نسبت تعداد یال‌ها در egonet و تعداد انجمن‌های عضو محاسبه و گره‌های دارای امتیاز بالا را ناهنجار معرفی می‌کند. به دلیل پیچیدگی محاسباتی کم این روش مناسب گراف‌های بزرگ است.

روش نیمه نظارتی [۲۹]، برای استخراج ویژگی‌های گراف ارائه شده است که الگوریتم بیشینه‌سازی مورد انتظار[5] و منطق فازی را ترکیب کرده و از خصوصیات محلی مانند تعداد گره‌ها و یال‌ها برای مدل‌سازی رفتار کاربران استفاده می‌کند. یک egonet برای هر کاربر با استفاده از گره‌های متصل به آن تولید و بر اساس ویژگی‌ها، امتیاز ناهنجاری ابتدایی به گره مرکزی هر egonet داده می‌شود. الگوریتم بیشینه‌سازی مورد انتظار برای خوشه‌بندی کاربران مطابق با رفتار ناهنجار تعریف‌شده در مرحله قبل استفاده می‌شود. هر خوشه با قوانین مشترکی که کاربران آن نشان می‌دهند نمایش داده و نهایتاً از منطق فازی برای تعریف درجه بی‌نظمی استفاده می‌شود.

روش دیگری توسط میشل و همکارانش [۵] برای شناسایی پروفایل‌های جعلی مرتبط با کاربران مخرب در شبکه‌های اجتماعی ارائه شد. این کاربران برخلاف کاربران معمول دارای درجه بالا و با تعداد فراوانی از انجمن‌ها ارتباط دارند. در این روش از درجه، تعداد انجمن‌های متصل، اتصالات همسایگان گره و میانگین دوستی در هر انجمن به‌عنوان شاخص‌های ساختاری استفاده می‌شود. سپس از درخت تصمیم برای مدل‌سازی رفتار کاربران عادی و شناسایی کاربران مخرب استفاده می‌شود. روشی مشابه [۳۰]، پس از تشخیص انجمن‌ها با الگوریتم OCTracker [۳۱]، با استفاده از مشخصه‌های ساختاری انجمن‌ها به شناسایی پروفایل‌های جعلی و حساب‌های کاربری مخرب می‌پردازد. این روش بانظارت ابتدا بر اساس یک مجموعه داده برچسب‌گذاری‌شده آموزش دیده و سپس به تعیین هنجار یا ناهنجار بودن نمونه‌های جدید می‌پردازد.

AUTOPART [۳۲]، یک روش تشخیص ناهنجاری مبتنی بر خوشه‌بندی است. این روش بر پایه ایده آن‌که گره‌های دارای همسایه مشابه در یک خوشه قرار گیرند بنا شده است. یال‌هایی که به هیچ خوشه‌ای تعلق نداشته باشند و هم‌چنین یال‌هایی که خوشه‌های مختلف را به هم متصل کنند، ناهنجاری به حساب می‌آیند. همچنین گره‌هایی که اتصالات بین خوشه‌ای بالایی دارند نیز متعلق به هیچ یک از آن‌ها به حساب نمی‌آیند و ناهنجاری هستند.

الگوریتم خوشه‌بندی ساختاری شبکه‌ها (SCAN)[6] [۳۳] و الگوریتم خوشه‌بندی مبتنی بر طرح گراف (GskeletonClu)[7] [۳۴] الگوریتم‌های خوشه‌بندی شبکه مبتنی بر تراکم[8] برای شناسایی خوشه‌ها، هاب‌ها[9] و پَرت‌ها در شبکه‌های بزرگ هستند. این الگوریتم‌ها از همسایگی گره‌ها به‌عنوان معیار خوشه‌بندی به‌جای همسایگان مستقیم‌شان استفاده می‌کنند و گره‌های دارای همسایگان مشترک را در یک خوشه قرار می‌دهند. گره‌های متصل به چندین خوشه، به‌عنوان‌هاب و گره‌هایی که به هیچ خوشه‌ای تعلق ندارند به‌عنوان پرت‌ها مشخص می‌شوند.

عامل بندی ماتریس داده A به‌صورت A=X×Y+R نشان داده

[1] Principal eigen value
[2] Near-stars
[3] Near-cliques
[4] Cliques
[5] Expected Maximization Algorithm

[6] Structural Clustering Algorithm for Networks(SCAN)
[7] Graph-skeleton based clustering(GskeletonClu)
[8] Density-based network clustering algorithms
[9] Hubs

می‌شود که X و Y عامل‌های کم رتبه[1] و R ماتریس باقی‌مانده است که نشان‌دهنده ناهنجاری است. برخلاف عامل‌بندی سنتی ماتریس غیرمنفی (NMF)، که عوامل X و Y را به غیرمنفی بودن محدود می‌کند، روش مبتنی بر عامل بندی ماتریس غیرمنفی[2] [۳۵] از محدودیت‌های غیرمنفی در ماتریس باقی‌مانده برای یافتن ناهنجاری‌ها استفاده می‌کند. مایلر و همکارانش [۳۶]، چارچوبی مبتنی بر پردازش سیگنال پیشنهاد دادند که از ویژگی‌های نُرم یک بردارهای ویژه ماتریس پیمانگی[3] شبکه برای تعیین حضور زیرگراف ناهنجار استفاده می‌کند. در این چارچوب هدف تعیین آن است که آیا شبکه مشاهده‌شده از فعالیت‌های معمولی تشکیل‌شده یا آن‌که یک زیرگراف کوچک مشهود با همبندی ناهنجار درون این شبکه جاسازی‌شده است. در کاری مشابه، نویسندگان چارچوب پیشنهادی خود [۳۷] را برای مسئله تشخیص زیرگراف تهدید جاسازی‌شده در شبکه‌های اجتماعی به‌کار گرفتند. همچنین چارچوبی بر اساس فضای ویژه اصلی[4] از ماتریس مانده‌های[5] گراف شبکه [۳۸] پیشنهاد دادند که در آن شبکه مشاهده‌شده با مقدار مورد انتظارش[6] برای یافتن زیرگراف‌های ناهنجار مقایسه می‌شود. مانده‌های بزرگ‌تر نشان‌دهنده وجود زیرگراف‌های ناهنجار است.

۴-۲- ناهنجاری ایستای با ویژگی

استفاده از اطلاعات مربوط به گره‌ها ویال‌ها می‌تواند باعث بهبود تشخیص ناهنجاری شود. نویسندگان روشی [۳۹] برای شناسایی زیرساختارها و زیرگراف‌های نامتعارف در یک گراف با ویژگی پیشنهاد دادند که از روش SUBDUE [۴۰] برای شناسایی زیرساختارهای پرتکرار و فشرده‌سازی شبکه به‌وسیله آن‌ها استفاده می‌شود. در این روش ناهنجاری زیرساختارهای نادر معرفی می‌شوند که باعث افزایش مقدار فشرده‌سازی شبکه می‌شوند. هم‌چنین زیر گراف دارای تعداد کمتری از زیرساختارهای رایج شانس بیشتری برای ناهنجاربودن دارد. مشابه روش قبل، از روش SUBDUE برای شناسایی الگوهای پرتکرار در شبکه بهره برده و نویسندگان سه الگوریتم برای تشخیص سه نوع ناهنجاری در تغییرات رایج شبکه ارائه کردند [۴۱]. درج گره یا یال جدید، تغییر برچسب گره‌ها یا یال‌ها و حذف گره‌ها یا یال‌های موجود تغییرات رایج در شبکه هستند. در این روش فرض شده که ناهنجاری‌ها شامل حداقل یکی از این تغییرات هستند و باعث تغییر جزئی در زیرساختارهای متداول می‌شوند. YAGADA [۴۲] برای تشخیص ناهنجاری در گراف‌های با برچسب‌های عددی است. ابتدا گره‌ها و یال‌ها مستقل از گراف و برحسب مقادیر برچسب‌ها توسط روش نزدیک‌ترین همسایه ارزیابی می‌شوند. پس از این مرحله گره‌ها و یال‌های معمول با یک مقدار ثابت و ناهنجاری‌ها با امتیاز ناهنجاری خود در گراف قرار داده می‌شوند. سپس زیرساختارهای پرتکرار گراف توسط روش SUBDUE استخراج و پس از فشرده‌سازی امتیاز ناهنجاری زیرساختارها محاسبه و ناهنجاری‌ها معرفی می‌شوند.

در روش تشخیص زیرگراف‌های ناهنجار پیشنهادشده [۴۳]، بر اساس ساختار اتصالات، یک زیرگراف ناهنجاری تشخیص داده می‌شود. هر زیرگراف ناهنجار است اگر تعداد زیادی یال غیرمنتظره داشته باشد و یا تعداد زیادی یال مورد انتظار درون خود یا بین خود و همسایگانش را از دست بدهد. از ویژگی‌های دو موجودیت برای پیش‌بینی وجود یک یال بین آن‌ها می‌توان استفاده کرد.

Radar[7] [۴۴]، روشی مبتنی بر تحلیل ماتریس باقی‌مانده برای تشخیص ناهنجاری در گراف‌های با ویژگی بدون جهت است. برای این منظور ابتدا ماتریس ویژگی‌ها برای دستیابی به نمونه‌های نماینده بازسازی و پس از بازسازی مجدد ماتریس ویژگی‌ها ماتریس باقی‌مانده محاسبه می‌شود. ناهنجاری‌ها دارای الگوهای متفاوتی با نمونه‌های معمول در ماتریس باقی‌مانده هستند. در این روش فرض می‌شود دو گره دارای پیوند خواهند بود اگر ویژگی‌های آن‌ها شبیه به هم باشد و در واقع الگوهای مشابهی در ماتریس باقی‌مانده داشته باشند. در انتها میزان ناهنجاری هر گره بر اساس نرم۲ بردار آن در ماتریس باقی‌مانده محاسبه و گره‌های با بیشترین میزان ناهنجاری مشخص می‌شوند. SSDM [۴۵] برای شناسایی هرزنویس‌ها[8] در شبکه‌های بلاگ‌نویسی کوچک ارائه کرده‌اند که از اطلاعات ساختاری و محتوایی به‌صورت هم‌زمان استفاده می‌کند. در این روش اطلاعات شبکه با استفاده از فرمول لاپلاسین [۴۶] و اطلاعات محتوایی با استفاده از روش یادگیری خلوت [۴۷] مدل می‌شوند. تشخیص ناهنجاری در قالب یک مساله بهینه‌سازی فرمول می‌شود که پس از حل آن برچسب کاربران مشخص می‌شود.

روش دیگری [۴۸] با استفاده از اطلاعات ساختاری و ویژگی‌ها، تشخیص انجمن و ناهنجاری را در یک چارچوب جامع

[1] Low-rank factors
[2] Matrix factorization
[3] Modularity matrix
[4] Principal eigen space
[5] Residual matrix
[6] Expected value
[7] Residual Analysis for Anomaly Detection in Attributed Networks
[8] Spammer

قرار داده است تا ضمن بهبود شناسایی انجمن‌ها، ناهنجاری‌های پرمعناتری را پیدا کند. با در نظر گرفتن انجمن‌ها به‌عنوان یک زمینه، هدف یافتن ناهنجاری‌های انجمنی[1] است که الگوی رفتاری متفاوتی با سایر نمونه داده‌ها در آن انجمن دارند. فرض اصلی آن است که داده‌های معمول انجمن‌ها را تشکیل می‌دهند و ناهنجاری‌ها به‌صورت تصادفی تولید شده و از توزیع یکنواخت پیروی می‌کنند. نویسندگان روشی برای تشخیص پرت‌های انجمنی در شبکه‌های اجتماعی با استفاده از اطلاعات تعاملات و ویژگی‌های افراد ارائه کرده‌اند [۴۹] که ابتدا گره‌ها را بر اساس ویژگی‌هایشان خوشه‌بندی و در مرحله بعد از اطلاعات ساختاری برای تشخیص ناهنجاری استفاده می‌کند. فرض اصلی برای تشخیص ناهنجاری آن‌که افراد در یک خوشه تعاملات بیشتری با یکدیگر دارند. بنابراین، گره‌هایی در گراف که تعداد پیوندهای آن‌ها با خوشه‌های دیگر بیشتر از پیوندهای آن‌ها با خوشه خود باشد ناهنجارتر به شمار می‌آیند.

روشی مبتنی بر تحلیل فضای ویژگی [۵۰] برای تشخیص ناهنجاری در گراف‌های با ویژگی پیشنهاد شده است. برای این موضوع زیرگراف‌های متراکم از گره‌ها با ارتباطات درونی زیاد و شباهت درون خوشه‌ای بالا تشکیل داده می‌شوند. سپس زیرمجموعه‌هایی از ویژگی‌ها که بیشترین وابستگی را با ساختار خوشه دارند به‌دست می‌آیند که گره‌های خوشه بیشترین شباهت را به یکدیگر در آن ویژگی‌ها دارند. پس از تشکیل زیرگراف‌ها و زیرمجموعه‌ها امتیاز ناهنجاری برای هر گره محاسبه می‌شود. ایده اصلی آن است که اشیا معمول تلاش به ایجاد خوشه با بسیاری از اشیا شبیه به هم دارند. بزرگی خوشه، ابعاد خوشه، مرکزیت مقادیر ویژه[2] و چگالی پیوند محلی[3] شاخص‌های تعریف‌شده برای اندازه‌گیری میزان هنجار بودن گره‌ها هستند.

ConSub [۵۱]، یک روش آماری برای انتخاب زیرمجموعه‌های متجانس از ویژگی در گراف‌های با ویژگی است. ایده اصلی در این روش یافتن ویژگی‌هایی است که بیشترین شباهت را با ساختار گراف داشته و گره‌های گراف مقادیر مشابه در این زیرمجموعه‌ها دارند. یک گره ناهنجار است اگر اختلاف بسیار زیادی با همسایگان خود در زیر مجموعه ویژگی انتخاب شده داشته باشد. نویسندگان روشی مشابه [۵۲] برای تشخیص ناهنجاری ارائه دادند با این تفاوت که به‌جای انتخاب زیرمجموعه‌هایی از ویژگی‌ها که نسبت به کل ساختار همبستگی داشته باشند، زیرمجموعه‌ای از ویژگی‌های مناسب در همسایگی هر گره شناسایی می‌شوند. در این روش ابتدا فضای محلی هر گره بر اساس شباهت‌های ساختاری تشکیل می‌شود. سپس ویژگی‌هایی که واریانس آن‌ها در فضای محلی کمتر از کل مجموعه داده باشد مناسب تشخیص و انتخاب می‌شوند. گره‌ای ناهنجار است که دارای انحراف قابل‌توجهی در مقادیر ویژگی‌ها نسبت به سایر گره‌ها در یک فضای متراکم از گره‌های شبیه به هم باشد.

برای تشخیص ناهنجاری در شبکه وبلاگ‌نویسی‌های کوچک، چارچوبی مبتنی بر گراف‌های دوبخشی[4] و خوشه‌بندی [۵۳] پیشنهاد شده است که علاوه بر تعاملات همگن، تعاملات غیرهمگن نیز در نظر گرفته می‌شود. در این روش یک شبکه دوبخشی بین کاربران و پیام‌ها برای نمایش تعاملات همگن بین موجودیت‌های یکسان و تعاملات غیر همگن بین موجودیت‌های متفاوت ساخته می‌شود سپس یک الگوریتم خوشه‌بندی مبتنی بر سه بعد، ماتریس غیرمنفی[5] برای تشخیص کاربران و پیام‌های ناهنجار به‌کار گرفته می‌گیرد. روش Glance [۵۴] با استخراج ویژگی‌های مناسب هر انجمن، ناهنجاری‌های انجمنی را شناسایی می‌کند. ابتدا گره‌های گراف بر اساس ویژگی‌های ساختاری و با استفاده از روش خوشه‌بندی لووین [۵۵] در انجمن‌های مختلف دسته‌بندی می‌شوند. سپس ویژگی‌های مناسب هر انجمن با استفاده از روش امتیازدهی لاپلاسین [۴۶] انتخاب می‌شوند. امتیاز ناهنجاری گره‌ها بر اساس انحراف مقادیر ویژگی‌هایشان از میانگین ویژگی‌های هر انجمن محاسبه و گره‌های با بیشترین امتیاز ناهنجاری مشخص می‌شوند.

۴-۳- ناهنجاری پویای بدون ویژگی

برای تشخیص این نوع از ناهنجاری تنها تغییرات ساختار شبکه در مراحل زمانی متوالی در نظر گرفته می‌شود و ویژگی‌های گره‌ها و یال‌ها تأثیری در تشخیص ناهنجاری ندارند. روشی [۵۶] برای شناسایی مراحل زمانی در گراف‌های پویا پیشنهاد شده است که در آن‌ها رفتار تعداد زیادی از گره‌ها نسبت به رفتار معمول شبکه انحراف دارد. در این روش ابتدا مجموعه‌ای از ویژگی‌های ساختاری از قبیل درجه ورودی و خروجی است

خراج و رفتار گره‌ها به‌وسیله آن‌ها خلاصه می‌شود. برای هر پنجره زمانی یک ماتریس همبستگی[6] از ویژگی‌ها با استفاده از ضریب همبستگی پیرسون[7] و بردار ویژه اصلی[8] محاسبه می‌شود سپس یک بردار رفتار برای هر گره تشکیل شده و با بردارهای رفتاری آن در پنجره زمانی قبلی مقایسه می‌شود. اگر رفتار فعلی

[1] Community Outliers
[2] Eigen value Centrality
[3] Local Edge Density

[4] Bipartite graph
[5] Non-negative matrix tri-factorization (NMTF)
[6] Correlation matrix
[7] Pearsons correlation coefficient
[8] Principal eigenvector

گره به میزان قابل‌توجهی متفاوت از رفتارهای قبلی باشد پنجره زمانی جاری، ناهنجاری معرفی می‌شود.

Tensorsplat [۵۷] یک روش مبتنی بر تجزیه برای تشخیص ریزخوشه‌ها[1]، تغییرات و ناهنجاری‌ها است که در آن دو بعد از تانسور[2]، اطلاعات ماتریس مجاورت را دارند و از ابعاد دیگر برای موجودیت‌های اضافی و زمان استفاده می‌کند. تانسور آرایه‌ای است از اعداد که در یک جدول قرار دارند. از تانسورها برای گسترش بردارها و ماتریس‌ها به ابعاد بالاتر استفاده می‌شود. ناهنجاری با مشاهده الگوهای رفتاری ناهنجار گره‌ها در یک زمان مشخص تعریف و به‌وسیله عامل‌های تجزیه شناسایی می‌شود. افزایش ناگهانی در تعاملات گره‌ها یا رفتار مشابه ربات، نمونه‌هایی از ناهنجاری هستند. روش دیگر مبتنی بر تجزیه ماتریس [۵۸] مناطق محلی مهم تغییر را در یک جریان سریع شبکه با استفاده از تحلیل مؤلفه‌های اصلی شناسایی می‌کند. یک ماتریس همبستگی یال برای هر گره نگه‌داری می‌شود که هر سطر و ستون آن شامل یک همسایگی گره است و به‌صورت پیوسته اطلاعات تغییرات همسایگی‌های مختلف شبکه را نگاه می‌دارد. از یک الگوریتم به‌روزرسانی بردار ویژه برای ذخیره اطلاعات همبستگی محلی استفاده شده که این الگوریتم برای محاسبه امتیاز ناهنجاری هر گره در هر مرحله زمانی به‌کار می‌رود و گره‌های دارای مقدار ناهنجاری بیشتر از مقدار آستانه ناهنجار به شمار می‌آیند.

روشی برای شناسایی گره‌های ناهنجار پیشنهاد شده است [۵۹]. این گره‌ها پرت‌های انجمن تکاملی[3] هستند که در گذر زمان رفتار متفاوتی با سایر اعضای انجمن دارند. این روش از یک الگوریتم کاهش مختصات برای بهبود کارایی تطبیق انجمن و شناسایی پرت‌ها استفاده می‌کند که پس از تطبیق انجمن‌ها در چند تصویر متوالی از شبکه، امتیاز پرتی هر گره محاسبه و مبنای معرفی ناهنجاری‌ها قرار می‌گیرد. روشی مشابه [۶۰] از الگوریتم افزایشی برای شناسایی پرت‌های تکاملی محلی[4] در یک شبکه وزن‌دار استفاده می‌کند. این گره‌های ناهنجار رفتار تکاملی نامتعارفی نسبت به همسایگان محلی خود دارند. همسایگان محلی یک گره با یک زیرگراف نشان داده می‌شوند که شامل گره و همسایگان تا دو فاصله است و وزن مسیر بین آن‌ها از یک مقدار آستانه بزرگ‌تر است. پس از تشکیل این زیرگراف، بر اساس ساختار شبکه و وزن یال‌ها، امتیاز ناهنجاری با بررسی و مقایسه زیرگراف‌ها در زمان‌های متفاوت محاسبه و گره‌های دارای امتیاز بالا ناهنجار هستند.

برای تشخیص ناهنجاری‌های مبتنی بر انجمن در شبکه‌های تکاملی مشخص توسط هم‌پوشانی انجمن‌ها روشی [۲۱] معرفی شده است که از روش مبتنی بر نماینده گراف و نماینده انجمن برای شناسایی هر شش نوع ناهنجاری ممکن یعنی رشد[5]، کاهش[6]، ادغام[7]، تقسیم[8]، به‌وجود آمدن[9] و از بین رفتن[10] استفاده می‌کند. نماینده گراف مجموعه‌ای از گره‌ها هستند که در تصاویر لحظه‌ای متوالی گراف وجود دارند. نماینده انجمن گره‌ای از آن انجمن است که در تعداد کمتری از سایر انجمن‌ها در گراف حضور دارد. با مشخص شدن نمایندگان گراف و انجمن ارتباط بین انجمن‌ها برقرار شده و ناهنجاری‌ها تعیین می‌شوند. روش NetSpot [۶۱] برای شناسایی زیرگراف‌های ناهنجار در گراف‌های پویای وزن‌دار است. این زیرگراف‌ها مناطق ناهنجار قابل ملاحظه[11] نامیده می‌شوند. برای شناسایی این مناطق ناهنجار، هر یال در هر نمونه گراف بر اساس توزیع وزن‌های روی آن محاسبه می‌شود. مناطقی که مجموع ناهنجاری آن‌ها از یک مقدار آستانه بیشتر باشد به همراه زمان متناظر به‌عنوان ناهنجاری شناخته می‌شوند.

یک روش مبتنی‌بر پیش‌بینی لینک[12] [۶۲] برای شناسایی ایمیل‌های ناهنجار پیشنهاد شده است. پیش‌بینی لینک برای پیش‌بینی تعاملات آینده بین افراد بر اساس تعاملات قبلی آن‌ها استفاده می‌شود. در این روش یک الگوریتم پیش‌بینی لینک اجرا و برای هر زوج کاربر در شبکه احتمال رخداد تعامل در مرحله زمانی بعدی محاسبه می‌شود. احتمال پیش‌بینی شده سپس با تعاملات مشاهده‌شده مقایسه و تعاملات مشاهده‌شده با احتمال پیش‌بینی پائین ناهنجاری تشخیص داده می‌شوند. روش Netprobe [۲۳] برای شناسایی کلاه‌برداران در یک سامانه حراج آنلاین پیشنهاد شد. این روش کاربران و تعاملات بین آن‌ها را به‌عنوان میدان تصادفی مارکف[13] برای شناسایی زیرگراف‌های ناهنجار نشان داده و از انتشار اعتقاد[14] برای پیش‌بینی احتمال خرابکار بودن کاربران استفاده می‌کند. کاربران در این روش به سه دسته کلاه‌بردار[15]، همدست[16] و صادق[17] تقسیم می‌شوند.

[1] Micro-clusters
[2] Tensors
[3] Evolutionary community outliers
[4] Local evolutionary outliers

[5] Grown
[6] Shrunken
[7] Merged
[8] Split
[9] Born
[10] Vanished
[11] Significant Anomalous Regions (SAR)
[12] Link prediction
[13] Markov Random Field
[14] Belief Propagation
[15] Fraudster
[16] Accomplice
[17] Honest

تعاملات بین کاربران کلاه‌بردار و همدست در شبکه هسته‌های دوبخشی را تشکیل می‌دهد که با کشف آن‌ها کلاه‌برداران شناسایی می‌شوند. روش دیگری [۶۳] برای یافتن الگوها و تشخیص فعالیت ناهنجار از نمایش فشرده ساده گراف بهره می‌برد و وزن‌ها را به یال‌ها به شیوه‌ای اختصاص می‌دهد که تکرار، مدت زمان و سرعت یال‌ها را در برگیرد. این الگوریتم، الگوهای پایدار را با استخراج اجزای متصل یال‌های تکراری به دست می‌آورد. این الگوهای پایدار رفتار مورد انتظار شبکه را مشخص و برای تشخیص رفتارهای ناهنجار محلی و عمومی به‌کار گرفته می‌شوند.

یک روش احتمالی مبتنی بر بیزین[1] [۲۲] برای تشخیص گره، یال و زیرگراف ناهنجار استفاده می‌کند. در این روش دو مرحله ای ابتدا ناهنجاری هر یال با مدل‌سازی تعاملات بین هر زوج گره به‌عنوان یک فرآیند شمارشی محاسبه می‌شود. این محاسبه بر اساس مقدار p پیش‌گویانه با توجه به یادگیری بیزی توزیع تعداد است. پائین بودن این مقدار از مقدار آستانه ثابت نشان‌دهنده انحراف رفتار مدل شده قبلی است و گره‌های مرتبط به این یال به مجموعه ناهنجاری بازه زمانی متناظر افزوده می‌شوند. در مرحله دوم یک زیرشبکه بر مبنای گره‌های مشخص‌شده در مرحله قبل و گره‌های در حال تعامل با آن‌ها ساخته و از روش‌های خوشه‌بندی استاندارد برای تشخیص نواحی ناهنجار استفاده می‌کند.

۴-۴- ناهنجاری پویای با ویژگی

برای تشخیص این ناهنجاری‌ها در شبکه علاوه بر بررسی تغییرات ساختار شبکه در گذر زمان، ویژگی‌های مربوط به گره‌ها و یال‌ها نیز در نظر گرفته می‌شوند. کارهای انجام شده در حوزه شناسایی این ناهنجاری‌ها در مقایسه با سایر انواع بسیار محدودتر است. روش ParCube [۶۴]، روشی سریع مبتنی بر نمونه‌برداری و قابلیت موازی‌سازی برای تجزیه تانسورهای تنک است. این روش امکان پردازش تانسورهایی را دارد که فضای بیشتری از حافظه موجود می‌خواهند. در این روش ابتدا تانسور با ابعاد بسیار بالا نمونه‌برداری وازتانسور نمونه به‌جای تانسور اصلی استفاده می‌شود. برای کار با شبکه‌های با ویژگی، برچسب‌های شبکه به اعداد طبیعی نگاشت می‌شوند. گره‌های ناهنجار در شبکه با ردیابی الگوهای نامتعارف در عامل‌های تجزیه تانسور تشخیص داده می‌شوند. ارتباطات اجتماعی افراد بر علاقه‌مندی‌های شخصی آن‌ها و ارتباطات اجتماعی آینده تأثیر می‌گذارد. به همین دلیل نویسندگان از مفهوم آبشار ویژگی[2] بهره برده‌اند [۶۵].

[1]Bayesian
[2]Feature cascade

مفهومی که چگونگی تأثیر ویژگی‌های قبلی گره بر ویژگی‌های همسایگان آینده آن را مشخص می‌کند. این روش از دو مرحله مدل‌سازی رفتار نرمال و تشخیص ناهنجاری تشکیل می‌شود. برای مرحله اول ابتدا فرآیند به وجود آمدن، از بین رفتن و مدت حضور ویژگی‌های ساختاری پنهان به‌صورت یک توزیع نمایی[3] فرض می‌شود. سپس ویژگی‌های گره‌های مدل شده توسط مدل مخفی فاکتوریل بی‌نهایت مارکف[4] [۶۶] با در نظر گرفتن آبشار ویژگی‌ها ارزیابی می‌شوند و رابطه بین ویژگی‌های گره‌ها و تولید پیوند مدل می‌شود. پس از این مدل‌سازی، پیوندهایی که از آن پیروی نکنند به‌عنوان ناهنجاری معرفی می‌شوند.

۵- چالش‌های تشخیص ناهنجاری

با وجود کارهای انجام‌شده، این حوزه همچنان جدید و ناشناخته است و با چالش‌های نظری و عملی متعددی مواجه است که هر یک از این چالش‌ها می‌توانند مبنایی برای تحقیقات آتی در این زمینه باشند. چالش‌های مهم موجود در تشخیص ناهنجاری در شبکه‌های اجتماعی در ادامه شرح داده شده است.

پیچیدگی و مقیاس‌پذیری: تشخیص ناهنجاری در داده‌های زیاد و به‌ویژه کلان داده‌ها[5] به دلیل اندازه و ابعاد بزرگ دشوار است. اغلب الگوریتم‌های تشخیص در این داده‌ها دارای پیچیدگی بالای محاسباتی از نظر زمان و فضا هستند. برای مثال استفاده از تمام ویژگی‌های یک مجموعه داده باعث نمایی شدن زمان اجرای الگوریتم‌ها خواهد شد. بنابراین، توسعه الگوریتم‌های کارآمد و مقیاس‌پذیر برای تشخیص ناهنجاری در گراف‌های بزرگ از چالش‌های مهم به شمار می‌آید [۶۷].

ارزیابی کارایی: با توجه به عدم وجود مجموعه داده‌های واقعی با حقیقت پایه[6]، ارزیابی کارایی واقعی الگوریتم‌های توسعه داده‌شده ممکن نیست. عمده روش‌ها ارزیابی خود را روی مجموعه داده‌های ساختگی با ناهنجاری‌های تزریق‌شده انجام می‌دهند. هرچند برخی از روش‌ها نیز اقدام به تولید مجموعه‌های با حقیقت پایه کرده‌اند اما این موضوع همچنان از جدی‌ترین چالش‌های این حوزه است. از سوی دیگر هنوز معیار یا استاندارد جامعی نیز برای ارزیابی و مقایسه روش‌ها معرفی نشده است [۶۶].

استخراج ویژگی: استفاده از تمام ویژگی‌ها برای تشخیص ناهنجاری‌های با ویژگی به دلیل وجود ویژگی‌های تصادفی و

[3]Exponential distribution
[4]Infinite Factorial Hidden Markov Model
[5]Big Data
[6]Ground truths

غیرمرتبط، باعث کاهش کارایی خواهد شد. ویژگی‌هایی مناسب هستند که تمایز بین هنجار و ناهنجاری را بهتر مشخص کنند، پیچیدگی محاسباتی بالا نداشته و در مقابل نویز مقاوم باشند. بنابراین، انتخاب ویژگی‌های مناسب از موضوعات مهم است که باید مورد توجه قرار گیرد [۶۷].

تشخیص ناهنجاری گراف‌های چندلایه: شبکه‌های اجتماعی را می‌توان در قالب گراف‌های چندلایه نمایش داد [۶۸]. هر لایه متناظر با یک شبکه اجتماعی بوده و این کار موجب ذخیره بهتر تعاملات و ارتباطات بین افراد خواهد شد. با در نظر گرفتن چنین ساختاری ناهنجاری‌های معنادارتری نیز می‌توان تشخیص داد. ارائه الگوریتم‌های تشخیص ناهنجاری روی چنین گراف‌هایی موضوعی است که باید مورد توجه قرار گیرد.

۶- نتیجه‌گیری

گراف ابزاری قوی برای نمایش ساختارهای پیچیده از قبیل شبکه‌های اجتماعی است. استفاده از گراف باعث استخراج اطلاعات مفید جهت تشخیص موجودیت‌های ناسازگار خواهد شد. انتخاب یک روش مناسب برای تشخیص ناهنجاری وابسته به کاربرد و نوع ناهنجاری مورد نظر است. معرفی جنبه‌های مختلف تشخیص ناهنجاری و هم‌چنین چالش‌های آن می‌تواند در انتخاب روش مناسب مفید باشد. توسعه روش‌هایی که ضمن کارایی بر روی شبکه‌های اجتماعی بزرگ و پویا، پیچیدگی محاسباتی قابل قبول داشته و بتوانند به‌صورت بلادرنگ و یا برخط عمل کنند می‌تواند محور توسعه روش‌های جدید در این زمینه باشد.

۷- منابع

۱. بسطامی، اسماعیل، جوادزاده، محمدعلی، تحلیل مرکزیت شبکه‌های اجتماعی در فضای سایبری با رویکرد مقابله با تهدیدات نرم، فصلنامه پدافند غیرعامل، شماره ۲۳، صفحات ۷۸-۶۹، ۱۳۹۴.

2. Z. Papacharissi, "Community, and Culture on Social Network Sites," A Networked Self: Identity, New York, Routledge, 2010.

3. "Worldwide Social Networks Users," eMarketer, 2017.

4. X. Ying, X. Wu, and D. Barbara, "Spectrum based fraud detection in social networks," IEEE 27th International Conference of Data Engineering (ICDE), 2011.

5. M. Fire, G. Katz, and Y. Elovici, "Strangers intrusion detection-detecting spammers and fake profiles in social networks based on topology anomalies," ASE Human Journal, vol. 1, pp. 26-39, 2012.

6. D. H. Chau, S. Pandit, and C. Faloutsos, "Detecting fraudulent personalities in networks of online auctioneers," Knowledge Discovery in Databases: PKDD, 2006.

7. D. Toshniwal and S.Yadav, "Adaptive Outlier Detection in Streaming Time Series," International Conference on Asia Agriculture and Animal, 2011.

8. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Computing Surveys (CSUR), vol. 41, pp. 1-72, 2009.

9. V. Barnett and T. Lewis, "Outliers in Statistical Data," 3rd Edition, John Wiley & Sons, 1994.

10. V. Hodge and J. Austin, "A survey of outlier detection methodologies," Artificial Intelligence Review, vol. 22, pp. 85–126, 2004.

11. V. Chandola, "Anomaly detection for symbolic sequences and time series data," Ph.D. thesis, university of Minnesota, 2009.

12. R. Hassanzadeh, R. Nayak, and D. Stebila, " Analyzing the Effectivenees of Graph Metrics for Anomaly Detection in Online Social Networks," Web Information Systems Engineering (WISE), 2012.

13. M. E. Newman, D. J. Watts, and S. H. Strogatz, "Random Graph Models of Social Networks," National Academy of Sciences, vol. 99, pp. 2566-2572, 2002.

14. R. Yu, X. He, and Y. Liu, "Glad: group anomaly detection in social media analysis," ACM Transactions on Knowledge Discovery from Data (TKDD), 2015.

15. Y. Chen and B. Malin, "Detection of anomalous insiders in collaborative environments via relational analysis of access logs," 1st ACM conference on Data and application security and privacy, 2011.

16. Y. Chen, S. Nyemba, and B. Malin, "Auditing medical records accesses via healthcare interaction networks," AMIA Annual Symposium Proceeding, 2012.

17. Y. Chen, S. Nyemba, W. Zhang, and B. Malin, "Specializing network analysis to detect anomalous insider actions," Security informatics, vol. 1, pp. 1-24, 2012.

18. N. Jindal, B. Liu, and E. P. Lim, "Finding unusual review patterns using unexpected rules," 19th ACM international conference on Information and knowledge management, 2010.

19. R. Kaur and S. Singh, "A survey of data mining and social network analysis based anomaly detection techniques," Egyptian Informatics Journal, vol. 17, pp. 199-216, 2016.

20. D. Savage, X. Zhang, X. Yu, P. Chou, and Q. Wan, "Anomaly detection in online social networks," Social Networks, vol. 39, pp. 62-70, 2014.

21. Z. Chen, W. Hendrix, and N. F. Samatova, "Community-based anomaly detection in evolutionary networks," Intell. Inf. Syst., vol. 39, pp. 59-85, 2012.

22. N. A. Heard, D. J. Weston, K. Platanioti, D. J. Hand, and others, "Bayesian anomaly detection methods for social networks," The Annals of Applied Statistics, vol. 4, pp. 645–662, 2010.

23. S. Pandit, D. Chau, S. Wang, and C. Faloutsos "Netprobe, A fast and scalable system for fraud detection in online auction networks," 16th international conference on World Wide Web, 2007.

24. L. Akoglu, M. McGlohon, and C. Faloutsos, "Anomaly Detection in Large Graphs," School of Computer Science Carnegie Mellon University, 2009.

25. F. Y. Edgeworth, "On discordant observations," Philosophical Magazine, pp. 364-375, 1887.

26. N. Shrivastava, A. Majumder, and R. Rastogi, "Mining (social) network graphs to detect random link attacks," In Data Engineering, IEEE 24th International Conference on, 2008.

27. L. Akoglu, M. McGlohon, and C. Faloutsos, "Oddball: spotting anomalies in weighted graphs," Advances in Knowledge Discovery and Data Mining, 2010.

28. M. A. Doostari, R. Zeinali, H. Lashkari, and M. Ajamzamani, "Anomaly Detection in Cliques of Online Social Networks Using Fuzzy Node-Fuzzy Graph," Journal of Basic and Applied Scientific Research, vol. 3, pp. 614-626, 2013.

29. R. Hassanzadeh and R. Nayak, "A semi-supervised graph-based algorithm for detecting outliers in online-social-networks," 28th Annual ACM Symposium on Applied Computing, 2013.

30. S. Y. Bhat and M. Abulaish, "Communities Against Deception in Online Social Networks," Computer fraud Security, vol. 2, pp. 8-16, 2014.

31. M. Abulaish and S. Y. Bhat, "A densitybased based approach to detect community evolutionary events in online social networks," 12th Social Network Analysis and Mining, 2013.

32. D. Chakrabarti, "Autopart: parameter-free graph partitioning and outlier detection," 8th European Conference on Principles and Practice of Knowledge Discovery in Databases, 2004.

33. X. Xu, N. Yuruk, Z. Feng, and T. A. Schweiger, "Scan: a structural clustering algorithm for networks," 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2007.

34. H. Sun, J. Huang, J. Han, H. Deng, and P. Zhao, "gskeletonclu: Density-based network clustering via structure-connected tree division or agglomeration," 10th IEEE International on DATA Mining, 2010.

35. H. Tong and C. Y. Lin, "Non-Negative Residual Matrix Factorization with Application to Graph Anomaly Detection," 11th SIAM International Conference on Data Mining, 2011.

36. B. Miller, N. Bliss, and P. J. Wolfe, "Subgraph detection using eigenvector L1 norms," 24th Annual Conference on Neural Information Processing Systems, 2010.

37. B. Miller, M. S. Beard, N. T. Bliss, and Others, "Eigenspace analysis for threat detection in social networks," 14th IEEE International Conference on Information Fusion, 2011.

38. B. Miller, M. Beard, P. Wolfe, and N. Bliss, "A spectral framework for anomalous subgraph detection," Signal Processing, IEEE Transactions, vol. 63, pp. 4191–4206, 2015.

39. C. C. Noble and D. J. Cook, "Graph-based anomaly detection," 9th ACM SIGKDD international conference on Knowledge discovery and data mining, 2003.

40. L. B. Holder, D. J. Cook, S. DjokO, et al, "Substucture Discovery in the SUBDUE System," 3rd International Conference on Knowledge Discovery and Data Mining, 1994.

41. W. Eberle and L. Holder, "Anomaly detection in data represented as graphs," Intelligent Data Analysis, vol. 11, pp. 663–689, 2007.

42. M. Davis, W. Liu, P. Miller, and G. Redpath, "Detecting Anomalies in Graphs with Numeric Labels," 20th ACM International Conference on Information and Knowledge Management, 2011.

43. M. Gupta, A. Mallya, S. Roy, J. Cho, and J. Han, "Local learning for mining outlier subgraphs from network datasets," SIAM International Conference on Data Mining, 2014.

44. J. Li, H. Dani, X. Hu, and H. Liu, "Radar: Residual Analysis for Anomaly Detection in Attributed Networks," 26th International Joint Conference on Artificial Intelligence, 2017.

45. X. Hu, J. Tang, Y. Zhang, and H. Liu, "Social Spammer Detection in Microblogging," 23th International Joint Conference on Artificial Intelligence, 2013.

46. X. He, D. Cai and P. Niyogi, "Laplacian Score for Feature Selection," 18th International Conference on Neural Information Processing Systems, 2005.

47. L. Ghanoui, G. Li, V. Duong, V. Pham, and Srivasta, "Sparse machine learning methods for understanding large text corpora," Conference on Intelligent Data Understanding, 2011.

48. J. Gao, F. Liang, W. Fan, C. Wang, and Y. Sun, "On community outliers and their efficient detection in information networks," 16th ACM SIGKDD international conference on Knowledge discovery and data mining, 2010.

49. T. Ji, J. Gao, and D. Yang, "A Scalable Algorithm for Detecting Community Outliers in Social Networks," International Conference on Web-Age Information Management, 2012.

50. E. Müller, P. I. Sánchez , Y. Mülle, and K. Böhm, "Ranking outlier nodes in subspaces of attributed graphs," 29th International Conference on Data Engineering Workshops, 2013.

51. P. I. Sánchez, E. Müller, F. Laforet, and F. Keller, "Statistical Selection of Congruent Subspaces for Mining Attributed Graphs," 13th IEEE International Conference on Data Mining, 2013.

52. P. I. Sánchez, E. Müller, O. Irmler, and K. Böhm, "Local context selection for outlier ranking in graphs with multiple numeric node attributes," 26th International Conference on Scientific and Statistical Database Management, 2014.

53. W. Yang, G. W. Shen, W. Wang, L. Y. Gong, and M. Yu, "Anomaly detection in microblogging via co-clustering," Journal of Computer Science and Technology, vol. 30, pp. 1097–1108, 2015.

54. M. A. Prado-Romero and A. Gago-Alonso, "Community Feature Selection for Anomaly Detection in Attributed Graphs," Pattern Recognition, Image Analysis, Computer Vision, and Applications, 2017.

55. V. D. Blondel, J. Guillaume, R. Lambiotte, and Lef, "Fast unfolding of communities in large networks," Journal of Statistical Mechanics: Theory and Experiment, vol. 2008, 2008.

56. L. Akoglu and C. Faloutsos, "Event detection in time series of mobile communication graphs," Army Science Conference, 2010.

57. D. Koutra, E. E. Papalexakis, and C. Faloutsos, "Tensorsplat: Spotting latent anomalies in time," 16th IEEE Panhellenic Conference on Informatics, 2012.

58. W. Yu, C. C. Aggarwal, S. Ma, and h. Wang, "On anomalous hotspot discovery in graph streams," 13th IEEE International Conference on Data Mining, 2013.

59. M. Gupta, J. Gao, Y. Sun, and J. Han, "Integrating community matching and outlier detection for mining evolutionary community outliers," 18th international conference on Knowledge discovery and data mining, 2012.

60. T. Ji, D. Yang, and J. Gao, "Incremental local evolutionary outlier detection for dynamic social networks," 13th European Conference on Machine Learning and Knowledge Discovery in Databases, 2013.

61. M. Mongiovi, P. Bogdanov, R. Ranca, E. E. Papalexakis, C. Faloutsos, and A. K. Singh, "Netspot: Spotting significant anomalous regions on dynamic networks," SIAM International Conference on Data Mining, 2013.

62. Z. Huang and D. D. Zeng, "A link prediction approach to anomalous email detection," International Conference on Systems, Man and Cybernetics, 2006.

63. B. Thompson and T. Eliassi-Rad, "Discovery and analysis of patterns and anomalies in volatile time-evolving networks 1st Workshop on Information in Networks," 2009.

64. E. E. Papalexakis, C. Faloutsos, and N. Sidiropoulos, "Parcube: Sparse parallelizable tensor decompositions," Machine Learning and Knowledge Discovery in Databases, 2012.

65. Y. Yasami and F. Safaei, "A statistical infinite feature cascade-based approach to anomaly detection for dynamic social networks," Computer Communications, vol. 100, pp. 52-64, 2017.

66. J. Van Gae, L. J. The, and Z. Ghahramani, "The infinite factorial hidden Markov model," 23rd Annual Conference on Neural Information Processing Systems, 2009.

67. P. V. Bindu and T. Santhi, "Mining Social Networks for Anomalies: Methods and Challenges," Journal of Network and Computer Applications, vol. 68, pp. 213-229, 2016.

68. X. H. Dang, I. Assent, R. T. Ng, A. ZimekSchub, and Schub, "Discriminative features for identifying and interpreting outliers," 30th IEEE International Conference on Data Engineering, 2014.

A survey on graph-based anomaly detection methods in social networks

M. Mirzaee, A. Mahabadi[*]

Abstract

The use of social networks to communicate and share information has grown dramatically in recent years. These networks are nowadays used in most areas such as education, business, health and entertainment. The large amount of valuable information on social networks has made them the main target of malicious users, such as spammers and fraudsters, for carrying out abusive and illegal activities. The abnormal and unexpected behavior of these users is identified using anomaly detection methods. Detection of anomalies is important in preventing fraud, dissemination of counterfeit information and configuration of attacks in these networks. Anomalies are static or dynamic, with or without attributes. In this paper, various methods developed for anomaly detection in social networks have been investigated and categorized and an overview provided on anomaly detection, its applications, existing challenges and key areas for future research.

Key Words: *Social networks, Anomaly detection, Social networks analysis*

[*] Shahed University- (mahabadi@shahed.ac.irr) - Writer-in-Charge