Journal of Electronical & Cyber Defence
Vol. 9, No. 1, 2021, Serial No. 33

Anomaly Detection in Dynamic Social Networks Based on Behavioral Measurements

S. Ghaderi Piraghum, M. Sakhaei-nia*, M. Mansoorizadeh

*Computer Engineering Department, Bu-Ali Sina University

(Received: 28/04/2020, Accepted: 26/10/2020)

ABSTRACT

Since the detection of anomalies in dynamic social networks takes place in a sequence of graphs over time, in addition to the storage management challenge, the detection process is difficult due to the slow evolution of graphs. A number of graphs are selected in the specified time frame, and by examining the changes of these graphs, the possible anomalies are detected. Therefore, choosing the number of time points (graphs) in the sequence of graphs is an important challenge in the detection of anomalies. In this paper, a novel method is proposed to detect anomalies based on structural data extracted from dynamic social network graphs. By extracting the centrality indicators from the network graph and their normalized mean, the activity criterion for each individual has been defined. Over time, changes in the activity criterion for each individual are measured and marked as the possibility of normal or abnormal behavior. If the individual's behavior measure exceeds a certain threshold, it is reported as an anomaly. The results show that the proposed method detects more anomalies with the accuracy and recall of 64.29 and 81.82 respectively, for the VAST 2008 data set. It also, detects more anomalies by selecting different number of time points in the graph sequence.

Keywords: Social Networks, Anomaly Detection, Dynamic Graph, Behavioral Measurement

* Corresponding Author Email: sakhaei@basu.ac.ir

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Publisher: Imam Hussein University © Authors

نشریه علمی "پدافند الکترونیکی و سایبری"
سال نهم، شماره ۱، بهار ۱۴۰۰، ص ۱۲۳-۱۱۵

علمی – پژوهشی

تشخیص ناهنجاری‌ها در شبکه‌های اجتماعی پویا بر اساس سنجش رفتاری

سعید قادری پیراقوم[۱]، مهدی سخائی‌نیا[۲*]، محرم منصوری‌زاده[۳]

۱- کارشناسی‌ارشد، ۲ و ۳- استادیار گروه مهندسی کامپیوتر، دانشکده مهندسی، دانشگاه بوعلی سینا، همدان

(دریافت: ۱۳۹۹/۰۲/۹، پذیرش: ۱۳۹۹/۰۸/۰۵)

چکیده

از آنجایی که شناسایی ناهنجاری در شبکه‌های اجتماعی پویا، در یک دنباله‌ای از گراف‌ها در طول زمان صورت می‌گیرد، علاوه بر چالش ذخیره‌سازی، فرآیند شناسایی به دلیل آهستگی تکامل گراف‌ها دشوار است. چند گراف در بازه زمانی مشخص انتخاب می‌گردد و با بررسی تغییرات این گراف‌ها ناهنجاری احتمالی شناسایی می‌گردد. بنابراین انتخاب تعداد گراف(نقاط زمانی) در دنباله گراف‌ها به‌عنوان یک چالش مهم در شناسایی ناهنجاری‌ها در شبکه‌های اجتماعی پویاست در این مقاله روشی نوینی برای شناسایی ناهنجاری مبتنی بر داده‌های ساختاری مستخرج از گراف پویای شبکه اجتماعی ارائه گردیده است. با استخراج شاخص‌های مرکزیت از گراف شبکه و میانگین نرمال شده آن‌ها، معیار فعالیت برای هر فرد تعریف‌شده است. با گذر زمان، تغییرات معیار فعالیت برای هر فرد موردسنجش قرارگرفته و به‌عنوان امکان رفتار هنجار یا ناهنجار علامت‌گذاری می‌گردد. درصورتی‌که شاخص سنجش رفتار فرد از آستانه معینی بیشتر گردید به‌عنوان ناهنجاری گزارش می‌گردد. نتایج نشان داد که روش ارائه‌شده بر روی مجموعه داده VAST 2008 تعداد ناهنجاری بیشتری را با دقت ۶۴/۲۹ و فراخوان ۸۱/۸۲ یافته است. همچنین با انتخاب تعداد مختلف نقاط زمانی در دنباله گراف تعداد ناهنجاری بیشتری را شناسایی نموده است.

کلیدواژه‌ها : شبکه‌های اجتماعی،تشخیص ناهنجاری، گراف پویا، سنجش رفتاری

۱- مقدمه

شبکه‌های اجتماعی برخط[۱] امکان ارتباط افراد مختلف با علاقه‌مندی‌های گوناگون را فراهم نموده است. سادگی در ایجاد یک شبکه از افراد، سبب رشد روزافزون شبکه‌های اجتماعی گردیده است. عدم محدودیت در شبکه‌های اجتماعی برخط، این امکان را فراهم می‌آورد که کاربران حجم عظیمی از اطلاعات را درباره کاربران دیگر به دست آورند. متأسفانه این اطلاعات می‌تواند در شبکه توسط کاربران خرابکار[۲] مورد سوءاستفاده قرار گیرد. دنیای امروزی با جرائمی نوین مرتبط با شبکه‌های اجتماعی مانند سرقت هویت[۳]، جرائم سازمان‌یافته، هرزه‌نگاری[۴] انتشار اطلاعات جعلی[۵]، و حتی حملات تروریستی برنامه‌ریزی‌شده مواجه هست [۱-۳]. بنابراین شناسایی ناهنجاری[۶] در شبکه‌های اجتماعی هم‌زمان با رشد این شبکه‌ها از اهمیت فراوانی برخوردار است.

ناهنجاری[۷] الگویی بین داده‌هاست که رفتار آن تطابقی با رفتار عادی و مورد انتظار اکثریت داده‌ها ندارد [۴]. در حوزه شبکه‌های اجتماعی ناهنجاری، یک رفتار از یک کاربر یا گروهی از کاربران است که رفتارشان در مقایسه با رفتار عادی کاربران در شبکه غیرمعمول باشد [۵]. در شبکه‌های اجتماعی اشیا مستقل از هم نبوده و با یکدیگر ارتباط دارند. این وابستگی درونی اشیا سبب شده که استفاده از روش‌های مبتنی بر گراف برای شناسایی ناهنجاری‌ها در شبکه‌های اجتماعی امری ضروری و اجتناب‌ناپذیر باشد [۶].

شناسایی ناهنجاری‌ها در شبکه‌های اجتماعی بر روی گراف‌های ایستا و پویا صورت می‌گیرد. برای شناسایی ناهنجاری در گراف ایستا، در یک تصویر لحظه‌ای از گراف شبکه اجتماعی، گره‌ها یا یال‌ها یا زیر ساختارهایی که کم هستند و متفاوت‌اند یا انحراف قابل‌توجهی از الگوهای مشاهده‌شده در گراف دارند، به‌عنوان ناهنجاری شناسایی می‌گردند. اما شناسایی ناهنجاری در گراف پویا، در یک دنباله از گراف‌ها در طول زمان صورت می‌گیرد. در گراف‌های پویا هدف پیدا کردن گره‌ها، یال‌ها، بخشی از گراف یا

* رایانامه نویسنده مسئول: sakhaei@basu.ac.ir

[1] On-line social networks
[2] Malicious users
[3] Identity theft
[4] Spamming
[5] Fraudulent information dissemination
[6] Anamoly detection
[7] Anomaly

* این مقاله یک مقاله با دسترسی آزاد است که تحت شرایط و ضوابط مجوز Creative Commons Attribution (CC BY) توزیع شده است.

ناشر: دانشگاه جامع امام حسین (ع)

© نویسندگان

بازه‌های زمانی حاوی تغییر یا رویداد خاص به‌عنوان ناهنجاری است [۶, ۷].

به دلیل ماهیت پویای شبکه‌های اجتماعی و اهمیت آن، روش‌های تشخیص ناهنجاری پویا در سال‌های اخیر موردتوجه قرارگرفته است. تعمیم روش‌های ایستا برای روش‌های پویا به دلیل چالش‌های ویژه و متفاوت داده‌های پویا مشکل و ناکارآمد است[۸]. نوع جدیدی از ناهنجاری در گراف‌های پویا مطرح می‌گردد که مربوط به تکامل گراف‌ها در طول زمان است مانند تقسیم یا ناپدید شدن ارتباطات. در روش‌های پویا دنباله‌ای از گراف‌ها که در طول زمان به‌روز می‌گردند نیاز به ذخیره‌سازی و تحلیل دارند که زیادشدن این گراف‌ها حجم داده‌ها را زیاد کرده و مدیریت ذخیره و بازیابی آن‌ها را مشکل می‌سازد. از سوی دیگر بعضی ناهنجاری در چندین مرحله تکامل گراف(چندین گراف در دنباله گراف‌ها) و به آهستگی رخ می‌دهد که شناسایی آن را دشوار می‌نماید. بنابراین انتخاب تعداد نقاط زمانی در دنباله گراف‌ها برای تحلیل به‌عنوان یک چالش از اهمیت فراوانی برای شناسایی ناهنجاری‌ها برخوردار خواهد بود.

اغلب روش‌های پویای تشخیص ناهنجاری به‌طورمعمول در دو مرحله انجام می‌گردد [۸]:

- استخراج و خلاصه‌سازی ویژگی‌های داده‌ای با نگاشت داده‌های خاص یک دامنه و گراف به داده‌های عمومی مانند اعداد حقیقی.
- شناسایی ناهنجاری‌ها بر اساس خروجی مرحله اول.

دو نوع اطلاعات می‌توان از گراف نشان‌دهنده شبکه اجتماعی برخط استخراج نمود. یکی اطلاعات رفتاری است که ناظر به رفتار پویای کاربر در استفاده از شبکه اجتماعی، مانند مدت‌زمان استفاده یا چت کاربر، است. دیگری اطلاعات ساختاری است که بر اساس ساختار شبکه‌ای کاربر است. اطلاعات مربوط ساختار، شامل مواردی مانند تعداد اتصال‌های کاربر با دیگر کاربران یا خصوصیات این اتصال‌ها هست. در تحلیل شبکه‌های اجتماعی اطلاعات ساختاری از اهمیت بیشتری برخوردار هستند. اطلاعات رفتاری، برخلاف اطلاعات ساختاری، قابل اعتماد نیستند؛ زیرا کاربر می‌تواند با تغییر عمدی رفتار، الگوی رفتاری حقیقی خود را پنهان نماید و همچنین فناوری مانند سرعت اینترنت یا دستگاهی که از آن استفاده می‌نماید، می‌تواند بر این نوع اطلاعات اثرگذار باشد [۹].

در این مقاله روشی نوینی برای شناسایی ناهنجاری مبتنی بر داده‌های ساختاری استخراج‌شده از گراف پویای شبکه اجتماعی ارائه گردیده است. ابتدا معیار فعالیت برای هر فرد(معادل یک گره در گراف) بر اساس داده‌های ساختار گراف شبکه اجتماعی تعریف گردیده است. بدین منظور شاخص‌های توزیع شبکه شامل شاخص‌های مرکزیت از گراف شبکه اجتماعی استخراج‌شده [۱۰] و میانگین نرمال شده این شاخص‌ها به‌عنوان معیار فعالیت گره‌ها یا کاربران در نظر گرفته‌شده است. سپس با گذر زمان، این معیار (بر روی دنباله‌ای از گراف‌ها) برای هر فرد محاسبه می‌گردد. تغییرات معیار فعالیت موردسنجش قرارگرفته و بر اساس الگوریتمی به‌عنوان امکان رفتار هنجار یا ناهنجار علامت‌گذاری می‌گردد. درصورتی‌که شاخص سنجش رفتار فرد از یک آستانه بیشتر گردید به‌عنوان ناهنجاری گزارش می‌گردد. نتایج نشان داد که روش ارائه‌شده بر روی داده آزمون خاص تعداد ناهنجاری بیشتری را یافته که سبب بهبود شاخص‌های دقت و فراخوان گردیده است. همچنین با انتخاب تعداد مختلف نقاط زمانی(تعداد مختلف گراف) در دنباله گراف تعداد ناهنجاری بیشتر را یافته است.

ساختار مقاله در ادامه به این شرح است. در بخش دوم، مفاهیم بنیادین به‌منظور درک بهتر مفاهیم مقاله ارائه‌شده است. کارهای مرتبط در بخش سوم بررسی گردیده است. در بخش چهارم روش پیشنهادی تشریح گردیده و در بخش پنجم نتایج ارزیابی روش پیشنهادی بیان گردیده است. در پایان مقاله با نتیجه‌گیری و کارهای آتی پایان پذیرفته است.

۲- مفاهیم بنیادین

در این بخش مفاهیم و اصطلاح‌های به‌کاربرده شده در مقاله تشریح می‌گردد.

نماد عمومی برای بازنمایی گراف به‌صورت$G = (V, E)$ است که V مجموعه گره‌ها یا کاربران شبکه اجتماعی است و E مجموعه یال‌های گراف یا اتصال‌های(ارتباط‌ها) کاربران است. در بازنمایی گرافی از یک شبکه اجتماعی، گره‌ها کنش‌گران[1] شبکه اجتماعی محسوب می‌گردند و یال‌ها، روابط بین فردی این کنش‌گران را نشان می‌دهد. در این خصوص شاخص‌های مرکزیت[2] نشان‌دهنده روابط ساختاری یک کنش‌گر در شبکه اجتماعی هستند [۱۱]. رایج‌ترین این شاخص‌ها درجه، نزدیکی و بینابینی هستند. ارتباط‌های افراد در شبکه اجتماعی متفاوت هستند و بعضی نقش محوری‌تری در شبکه اجتماعی بازی می‌کنند که با شاخص‌های

[1] Actors
[2] Centrality

مرکزیت قابل‌شناسایی هستند. در این مقاله از شاخص‌های مرکزیت به‌عنوان معیار فعالیت استفاده گردیده است که در ادامه این شاخص‌ها تشریح گردیده است [۱۲, ۱۳].

۲-۱- مرکزیت درجه

در مرکزیت درجه گراف، اهمیت یک گره بر اساس تعداد گره‌های که به آن گره متصل هستند، تعیین می‌گردد. این شاخص تعداد اتصال مستقیم به سایر کنش‌گران در شبکه را نشان می‌دهد. البته ارتباط‌های غیرمستقیم در نظر گرفته نمی‌شود لذا این شاخص فقط اهمیت محلی گره را نشان می‌دهد. این شاخص به‌صورت رابطه (۱) محاسبه می‌گردد.

$$C_D(v) = deg(v) \quad (۱)$$

مقدار مرکزیت درجه با تقسیم بر بیشترین درجه ممکن نرمال می‌گردد. در یک گراف ساده با n گره این مقدار برابر n-1 خواهد بود. برای یک گراف چندگانه این مقدار ممکن است که بیشتر از n-1 باشد.

۲-۲- مرکزیت نزدیکی

مرکزیت نزدیکی[1]، میانگین طول کوتاه‌ترین مسیرها مابین گره v و تمام گره‌های دیگر در گراف است. این معیار نشان می‌دهد که یک گره برای ارتباط با گره‌های دیگر به چه تعداد گره واسط نیازمند است. مرکزیت نزدیکی برای درک انتشار اطلاعات در شبکه‌ها مهم است؛ ازاین‌جهت که فاصله بین یک کنشگر خاص و دیگران بر چگونگی ارسال یا دریافت اطلاعات از طریق این کنشگر به سایرین اثر دارد. مرکزیت نزدیکی، $C_C(v)$ ، بر اساس رابطه (۲) محاسبه می‌گردد:

$$C_C(v) = \frac{n-1}{\sum_{u=1}^{n-1} d(u,v)} \quad (۲)$$

که $d(u,v)$ کوتاه‌ترین فاصله بین u و v است و n تعداد گره‌هایی است که از v در دسترس هستند.

۲-۳- مرکزیت بینابینی

معیار مرکزیت بینابینی[2] در گراف‌ها که بیانگر تعداد مسیرهای کوتاه مابین گره‌های مختلف در گراف است که از گره v عبور کرده‌اند. این شاخص میزان اهمیت یک کنشگر را به‌عنوان اتصال‌دهنده بین شبکه‌های مختلف بیان می‌کند. به‌عبارت‌دیگر مرکزیت بینابینی نشان می‌دهد که یک گره خاص برای ارتباط چه تعدادی از زوج گره‌های موجود در گراف نقش واسط را دارد. گره با مرکزیت بینابینی زیاد کنترل‌کننده جریان اطلاعات بوده و پلی بین گروهی گره‌ها با گره‌های دیگر است. این شاخص به‌وسیله رابطه (۳) محاسبه می‌گردد:

$$C_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)} \quad (۳)$$

که v مجموعه‌ای از گره‌ها است، $\sigma(s,t)$ تعداد کوتاه‌ترین مسیرها بین s و t است و $\sigma(s,t|v)$ تعداد مسیرهای بین s و t است که از گره v عبور می‌کند.

۳- کارهای مرتبط

روش‌های مختلفی و زیادی برای تشخیص ناهنجاری‌ها در شبکه‌های اجتماعی وجود دارد. علاوه بر روش‌های ایستا، به دلیل ماهیت تغییرپذیری شبکه‌های اجتماعی امروزی روش‌های مبتنی بر گراف پویا نیز بسیار گسترش یافته‌اند. دسته‌بندی‌های مختلفی برای شناسایی ناهنجاری‌ها در حوزه گراف‌های پویا ارائه گردیده است [۶-۸, ۱۴]. مقایسه و انتخاب یکی از این روش‌ها به عواملی مختلفی مانند نوع کاربرد، نوع داده‌ها و هدف مرتبط است.

دریکی از دسته‌بندی‌ها [۶]، روش‌های مبتنی بر گراف پویا به چهار دسته تقسیم شده است. اولین دسته روش‌های مبتنی بر ویژگی[3] بر اساس این ایده است که گراف‌های مشابه احتمالاً دارای ویژگی‌های مشترکی مانند خصوصیات خاص ساختاری هستند. بنابراین با خلاصه‌سازی مناسب تصویر لحظه‌ای از گراف و مقایسه این گراف‌های متوالی با استفاده از یک معیار فاصله یا شباهت، درصورتی‌که این فاصله از یک حد آستانه بزرگ‌تر بود، تصویر لحظه‌ای به‌عنوان ناهنجاری تعیین می‌گردد. دومین دسته روش‌های مبتنی بر تجزیه[4]، ناهنجاری‌های زمانی را به‌وسیله مرتب‌سازی تجزیه ماتریس یا تانسور[5] گراف‌های دارای زمان تشخیص می‌دهند و بردارهای ویژه، مقادیر ویژه و منفرد انتخابی را به‌طور مناسب تفسیر می‌کنند. سومین دسته، روش‌های مبتنی بر پنجره[6]، در یک پنجره زمانی محدود الگوها و رفتارهای ناهنجار در دنباله گراف ورودی کشف می‌نماید. در ابتدا، تعدادی از نمونه‌های قبلی برای مدل‌سازی رفتار عادی استفاده می‌شوند. سپس گراف داده شده با آن‌ها مقایسه می‌شود تا تشخیص داده شود که ناهنجار است یا هنجار. آخرین دسته، روش‌های مبتنی بر

[1] Closeness centrality
[2] Betweenness centrality

[3] Feature-based
[4] Decomposion-based
[5] Tensor
[6] Window-based

خوشه‌بندی یا جامعه[1]، به‌جای نظارت بر تغییرات در کل شبکه، خوشه‌ها یا جوامع گراف در هر زمان را پایش کرده و رویداد زمانی تغییرات ساختاری یا محتوایی در هرکدام از آن‌ها را گزارش می‌نماید. برای هریک از چهار دسته فوق چندین روش ذکر شده است و مقایسه این روش‌ها با یکدیگر کار سختی است و وابسته به شاخص‌های متفاوت مانند مجموعه داده آزمون یا بزرگی گراف برای تحلیل می‌توانند بررسی گردند.

روش تشخیص ناهنجاری بیزین[2][۱۵] یکی از روش‌های مبتنی بر خوشه‌بندی است. این روش بر روی تشخیص نواحی ناهنجار در شبکه‌های اجتماعی با استفاده از روش بیزین دومرحله‌ای تمرکز دارد. در مرحله اول این روش، ناهنجاری هر یال با مدل کردن ارتباطات بین هر جفت گره به‌عنوان فرآیند شمارش محاسبه می‌شود. هم‌چنین در هر گراف نمونه، شاخص p بر اساس آموزش بیزین توزیع‌های شمارش برای هر یال موجود محاسبه می‌شود تا تصمیم گرفته شود که آیا ناهنجار است یا نه. الگوریتم، دنباله گراف را به‌عنوان یک جریان در نظر می‌گیرد و تغییرات در گراف‌های جدید را بر اساس تاریخچه(آنالیز ترتیبی) تشخیص می‌دهد. به‌صورت هم‌زمان، این الگوریتم تاریخچه را هم به‌روز می‌کند. به‌طور اساسی روش‌های خوشه‌بندی را بر روی زیر گراف‌های کوچک شامل گره‌ها و یال‌های ناهنجار مرحله اول اعمال می‌کند و بنابراین نواحی ناهنجار محلی کشف می‌شوند. این روش با بررسی ناهنجاری در یال‌ها به شناسایی گره‌های ناهنجار در گراف پویا می‌پردازد که روش ارائه‌شده در این مقاله نیز به شناسایی گره‌های ناهنجار در گراف می‌پردازد. پژوهش دیگری هم برای شناسایی ناهنجاری‌ها از روش بیزین الهام گرفته و مانند آن از دو مرحله تشکیل شده است [۱۶]. این روش ضمن تغییر تابع توزیع احتمال به‌کار گرفته شده در مرحله اول روش بیزین، تقسیم‌بندی داده‌های آزمایشی را نیز تغییر داده تا بر محدودیت روش بیزین در این خصوص غلبه نماید. دقت نتایج حاصل از این پژوهش نسبت به روش بیزین اصلی کمتر بوده است. این در حالی است که روش ارائه‌شده در این مقاله از روش بیزین دقت بالاتری داشته و بر محدودیت ذکر شده نیز غلبه نموده است.

به‌هرحال در تمامی روش‌های مبتنی بر گراف پویا باید یک خلاصه مناسب از ویژگی‌های ساختاری ارائه گردد که بیانگر رفتارهای گره‌ها است. بااین‌حال، ویژگی‌های ساختاری متنوع و گسترده‌ای در شبکه‌های اجتماعی وجود دارد و دستیابی به یک خلاصه ویژگی خوب از آن‌ها چالش برانگیز است [۸]. در این مقاله از شاخص‌های توزیع شبکه[3] استفاده گردیده است. شاخص توزیع شبکه مربوط به نحوه توزیع گره‌ها و یال‌ها در یک شبکه است و نشان داده شده است که این شاخص‌ها نسبت به روش‌های دیگر مؤثرتر هستند. قبلاً نیز از این شاخص‌ها در پژوهش‌ها استفاده گردیده است [۹, ۱۷, ۱۸] که در گراف‌های ایستا به کار گرفته شده و در روش‌های پویا استفاده نشده‌اند. در [۹] با مدل کردن شبکه اجتماعی با گراف و با استفاده از شاخص‌های گراف مانند تعداد گره‌های همسایه و متوسط مرکزیت بینابینی، ناهنجاری ارتباطات کاربران شناسایی گردیده است. در این مقاله با لحاظ نمودن ارتباط بین تعداد گره‌های همسایه و لبه‌های آن با متوسط مرکزیت بینابینی نشان داده است که این شاخص‌ها برای شناسایی ناهنجاری‌ها در گراف ایستا مؤثر هستند. در پژوهش دیگر [۱۷] با در نظر گرفتن مرکزیت نزدیکی و ارتباط آن با تعداد گره‌ها و لبه‌ها نشان داد که می‌توان تعداد مثبت‌ها و منفی‌های کاذب ناهنجاری‌ها را کاهش داد. در [۱۸] اثر به‌کارگیری این شاخص‌ها و ارتباط آن‌ها با یکدیگر بر شناسایی ناهنجاری‌ها مقایسه گردیده است.

۴- روش پیشنهادی

درروش ارائه‌شده در این مقاله ویژگی‌های ساختاری از گراف شبکه اجتماعی برای هر گره به‌صورت مستقل استخراج شده است. بر اساس این ویژگی‌ها، یک معیار فعالیت تعریف گردیده است که نشان‌دهنده رفتار هر گره است. با استفاده از الگوریتم سنجش رفتاری ارائه‌شده در این مقاله، اگر رفتار گره، در طول زمان، نسبت به رفتار قبلی آن تغییر داشت، بررسی لازم صورت گرفته و بر اساس یک حد آستانه این رفتار ناهنجار گزارش می‌گردد. در بخش۴-۱ معیار فعالیت و در بخش ۴-۲ تقسیم‌بندی رفتارها تشریح گردیده است. در بخش ۴-۳ الگوریتم سنجش رفتار بیان گردیده است.

۴-۱- معیار فعالیت

برای تحلیل شبکه‌های اجتماعی نیاز است که میزان فعالیت هر کنش‌گر در شبکه محاسبه گردد. همان‌طور که در بخش قبلی بیان شد در این مقاله از شاخص‌های توزیع شبکه برای تعیین میزان فعالیت هر کنش‌گر استفاده گردیده است و از آن به‌عنوان معیار فعالیت[4] یاد شده است. این معیار برای گره v به‌صورت رابطه (۴) محاسبه می‌گردد:

$$AC_v = \frac{C_B(v)+ C_C(v)+ C_D(v)+N(v)+d(v)}{5} \quad (۴)$$

[1] Community/cluster-based
[2] Bayesian anomaly detection method
[3] Network ditribution
[4] Activity Criteria(AC)

که $C_B(v)$ مرکزیت بینابینی، $C_C(v)$ مرکزیت نزدیکی و $C_D(v)$ مرکزیت درجه بوده و بر اساس روابط (۱) تا (۳) قابل محاسبه است. $N(v)$ تعداد همسایگان و $d(v)$ درجه گره v است. نیاز به توضیح هست که در رابطه (۴) از تابع میانگین به‌عنوان یک عملگر ترکیب استفاده شده است. بدیهی است که از عملگرهای مشابه و رایج دیگر مانند min و max نیز می‌توان به همین منظور استفاده نمود. در اینجا بر اساس بررسی‌های اولیه، میانگین برگزیده شد.

معیار فعالیت نقش مهمی در شناخت رفتارهای مختلف افراد در شبکه‌های اجتماعی دارد. با استفاده از این معیار، الگوریتم سنجش رفتاری می‌تواند تغییرات رفتاری یک فرد را زیر نظر گرفته و بر اساس آن تصمیمات خاصی را اتخاذ نماید.

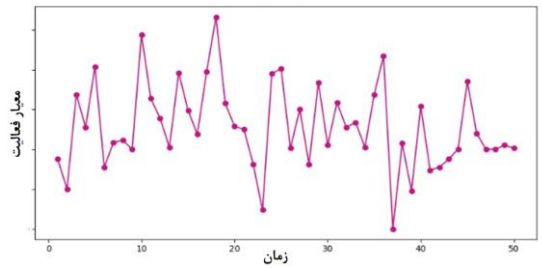
معیار فعالیت برای هر گره در طول زمان در بازه‌های زمانی گسسته $t_0 < t_1 < t_2 < t_3 < \cdots < t_T$ به‌صورت مستقل محاسبه می‌شود. بر اساس تغییر معیار فعالیت در طول زمان رفتار هر گره ارزیابی گردیده و بر اساس آن ناهنجاری‌ها شناسایی می‌گردد.

۴-۲- انواع رفتارها

رفتار یک کنش‌گر در شبکه اجتماعی برحسب تغییر معیار فعالیت آن کنش‌گر در طول زمان در بازه‌های گسسته سنجیده می‌شود. تغییر در رفتار یک کنش‌گر در بازه زمانی کوتاه می‌تواند یک ناهنجاری باشد. یکی از چالش‌های شناسایی ناهنجاری‌ها در گراف‌های پویا، تحلیل بر اساس گره‌های حاضر در گراف است. در این مقاله با تقسیم‌بندی رفتار گره‌ها، تلاش گردیده رفتار گره‌هایی که به شبکه اضافه می‌شوند و یا از شبکه خارج می‌گردند نیز تحلیل گردند. ازاین‌رو، چهار نوع رفتار برای کنش‌گران تعریف گردیده است.

- **رفتار نوع اول: کنشگران فعال**

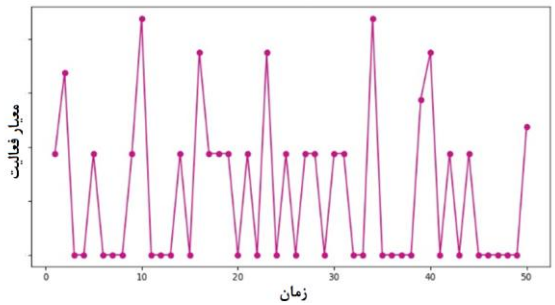
رفتار نوع اول به افرادی مربوط می‌شود که در طول زمان‌های مختلف اکثراً فعالیت داشته‌اند. بر اساس طبیعت رفتاری افراد در شبکه‌های اجتماعی، یک فرد ممکن است ارتباطات جدیدی ایجاد کند یا برخی از ارتباطات قبلی خود را خاتمه دهد. افرادی که در این گروه قرار می‌گیرند غالباً افراد با فعالیت زیاد هستند که در بیشتر زمان‌ها حضور فعالی در شبکه داشته‌اند. بنابراین، اگر این افراد در بازه‌ای از زمان فعالیت چشمگیری نداشته باشند، از رفتار طبیعی خود فاصله گرفته‌اند. در ۰نمونه‌ای از تغییر معیار فعالیت یک فرد با رفتار نوع اول نشان داده شده است که یک کنش‌گر با رفتار هنجار محسوب می‌گردد.



شکل (۱): تغییر معیار فعالیت در رفتار نوع اول

- **رفتار نوع دوم: کنشگران عادی**

رفتار نوع دوم مربوط به افراد عادی در شبکه‌های اجتماعی می‌شود. افراد عادی، افرادی هستند که در طول زمان‌های مختلف گاهی در شبکه حضورداشته‌اند و گاهی حضور نداشته‌اند. به این معنی که ممکن است در یک بازه زمانی نمودار رفتار آن‌ها فعالیت و در بازه بعدی بدون فعالیت را نشان دهد و تا آخرین بازه همین روند تکرار شود. افرادی که در این گروه قرار می‌گیرند افرادی با فعالیت متوسط هستند که تقریباً در نیمی از زمان‌ها در شبکه فعال و در نیمی دیگر غیرفعال بوده‌اند. بنابراین اگر این افراد در بازه‌ای فعالیت چشمگیر نداشته باشند و یا بالعکس، در بازه‌ای فعالیت چشمگیر داشته باشند از رفتار طبیعی خود فاصله گرفته‌اند. اما این نکته در این نوع رفتارها باید موردتوجه قرار گیرد که این نوع رفتارها با رفتارهایی که برای مدتی طولانی فعالیت داشته‌اند و در ادامه غیرفعال می‌شوند (یا بالعکس) تفاوت چشم‌گیری دارند. تغییرات معیار فعالیت نمونه‌ای از این کنش‌گران که هنجار محسوب می‌شوند، در ۰ مشاهده می‌گردد.

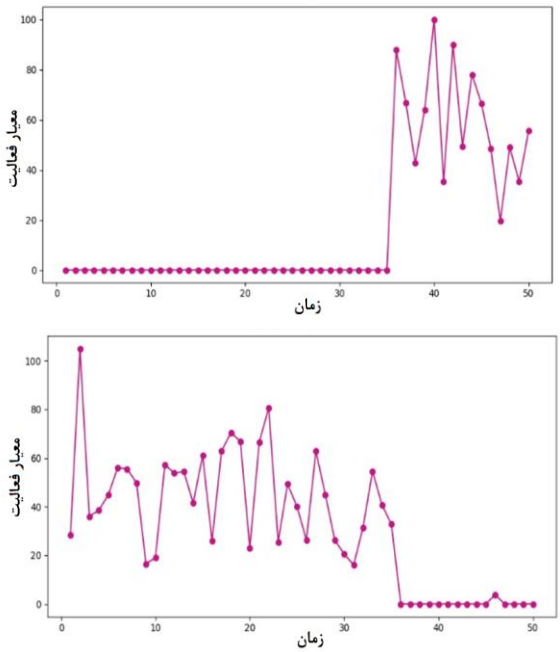


شکل (۲): تغییر معیار فعالیت در رفتار نوع دوم

- **رفتار نوع سوم و چهارم: کنشگران غیرعادی**

رفتارهای نوع سوم و چهارم مربوط به افرادی در شبکه‌های اجتماعی هستند که در کل زمان‌ها یک نوع رفتار ثابتی نداشته‌اند. دسته سوم افرادی هستند که در بازه‌ای از زمان هیچ‌گونه فعالیتی

از خود در شبکه نشان نداده‌اند، ولی در بازه‌های بعدی شروع به فعالیت زیاد کرده‌اند. دسته چهارم افرادی هستند که در شبکه حضورداشته و فعالیت داشته‌اند اما در بازه‌های زمانی طولانی بعدی از شبکه خارج‌شده‌اند. تغییرات معیار فعالیت این دو نوع رفتار در ۰نشان داده شده است. افرادی که در این دو گروه دسته‌بندی می‌شوند افرادی هستند که در یک بازه زمانی خاص یا از فعالیت خود دست کشیده‌اند یا شروع به فعالیت کرده‌اند. درواقع این نوع افراد با شدت بیشتری نسبت به نوع‌های دیگر رفتار خود را تغییر داده‌اند و از هر نوع رفتار دیگری ناهنجارتر به شمار می‌آیند. اگر معیارهای فعالیت این دسته از افراد بالاتر از سایر گره‌ها در شبکه باشد ممکن است ساختار گراف شبکه را نیز به‌شدت تغییر دهند. به همین دلیل تشخیص این نوع افراد که چنین رفتاری از خود در شبکه نشان می‌دهند بسیار مهم است و ناهنجار محسوب می‌شود.



شکل (۳): تغییر معیار فعالیت در رفتار نوع سوم و چهارم

۴-۳-الگوریتم سنجش رفتاری

سنجش رفتاری، یک روش منعطف برای ردگیری تغییرات رفتاری هر گره در شبکه‌های اجتماعی در طول زمان است. این روش با استفاده از یک الگوی امتیازدهی بر اساس تغییرمعیار فعالیت گره در طول زمان عمل می‌کند. ایده کلی این الگوریتم مبتنی بر تشویق و تنبیه[1] است. به این معنی که الگوریتم در طول زمان تغییرات معیار فعالیت را موردتوجه قرار می‌دهد و هرکجا که این معیار مانند رفتار طبیعی مورد انتظارش باشد به شاخص سنجش رفتار[2] وی افزوده می‌شود و هرکجا که خلاف رفتار طبیعی مورد انتظارش باشد، از شاخص سنجش رفتاری وی کاسته می‌شود. میزان افزایش و کاهش شاخص سنجش رفتاری یکسان نیستند. افزایش شاخص سنجش رفتاری به‌صورت خطی و کاهش آن با ضریب بیشتری در نظر گرفته شده است. الگوریتم سنجش رفتار مسکو در شبه کد ۰ بیان شده است.

در خطوط ۱ تا ۴ از شبه کد ۵۰ معیار فعالیت مرکزیت بینابینی (Bc)، مرکزیت نزدیکی (Cc)، مرکزیت درجه (Dc)، تعداد همسایگان (N) و درجه (d) هر گره در هر زمان $t_k \in [0, T]$ محاسبه شده است. میانگین این مقادیر،$AC(v_i, t_k)$ ، برای هر گره v_i در زمان t_k، نیز محاسبه گردیده است. با توجه به اینکه بازه مقداری هر یک از این متغیرها با یکدیگر متفاوت بود، نرمال‌سازی بر روی متغیرها قبل از محاسبه میانگین صورت گرفته است. کل زمان T به بازه‌هایی مساوی تقسیم گردد و در انتهای هر بازه، t_k ، وضعیت فعالیت یک گره بررسی می‌شود و رفتار گره بر اساس تغییر فعالیت آن در زمان‌های مشخص شده سنجیده می‌شود.از خط (۵) تا خط (۹) مقادیر اولیه شامل: امتیاز سنجش رفتار (BAS)، لیست امتیاز سنجش برای هر گره(BASList)، امتیاز تشویق (EScore) ، امتیاز تنبیه (PScore) و نوع رفتار فرد (Behavior) در شبکه اجتماعی بر اساس مشاهده تعداد بازه‌های فعالیت و تعداد بازه‌های بدون فعالیت با توزیع طبیعی آن‌ها در کل بازه‌های زمانی تعیین گشته است. نوع رفتار فرد در ادامه این بخش موردبررسی قرار گرفته است. از خط (۱۰) تا خط (۲۰) الگوریتم تشویق و تنبیه برای نوع رفتارهای ناهنجار، رفتارهای نوع سوم و چهارم، اعمال می‌شود. الگوریتم تشویق و تنبیه بر اساس رفتار طبیعی یک فرد که ممکن است فعالیت کردن یا فعالیت نکردن باشد، امتیاز تشویق را به‌صورت خطی افزایش داده و با امتیاز سنجش در هر بازه جمع می‌کند و امتیاز تنبیه را به‌صورت خطی افزایش می‌دهد و در هر بازه تنبیه، امتیاز سنجش تقسیم‌بر امتیاز تنبیه می‌گردد. به‌این‌ترتیب این روند تا آخرین بازه زمانی برای هر گره به‌صورت مستقل انجام می‌گردد. اگر رفتار یک شخص در تمام بازه‌ها به‌صورت هنجار، رفتارهای نوع اول و دوم، باشد، معیار سنجش رفتار برای این افراد بر اساس خط (۲۱) و (۲۲) محاسبه می‌گردد. امتیاز سنجش رفتاری در زمان t_k در یک لیست BASList نگهداری می‌گردد(خط ۲۳). در خطوط (۲۵) تا (۲۷) اگر در بازه‌ای امتیاز سنجش از حد آستانه پایین‌تر شود آن گره به لیست گره‌های ناهنجار شبکه افزوده می‌شود. نتیجه شبه کد الگوریتم سنجش رفتاری یک لیست از گره‌های ناهنجار در شبکه‌های اجتماعی است که در حالت‌های مختلفی از رفتار طبیعی خود در طول زمان فاصله گرفته‌اند (لیست AN).

[1] Encouragement and Punishment

[2] Behavior Assesment Score(BAS)

Algorithm- Behavioural Measurement

Input:

Graph: G
Node: $v_i \in V$
Time Series: $t_k \in [0, T]$

Output:

List of anomalous nodes: AN[]

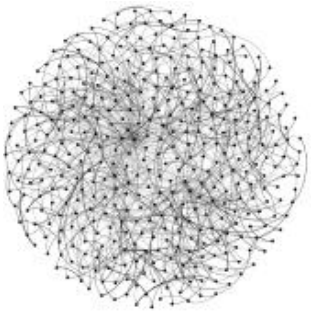
Method:

(1) For each v_i in V:
(2) For t_k in [0, T]:
(3) Compute mean of $Bc(v_i, t_k)$, $Cc(v_i, t_k)$, $Dc(v_i, t_k)$, $N(v_i, t_k)$, $d(v_i, t_k)$ to $AC(v_i, t_k)$
(4) End
(5) BAS = Median($AC(v_i, t_k)| t_k \in [0, T]$)
(6) EScore = 1
(7) PScore = 2
(8) BASList [] = [16]
(9) Choose of Behavior for v_i in B1, B2, B3, B4
(10) For t_k in [0, T]:
(11) If Behavior is B3, B4:
(12) If Behavior in t_k is NORMAL:
(13) BAS = BAS + (EScore/2)
(14) EScore += 1
(15) PScore /= 2
(16) If Behavior in t_k is ABNORMAL:
(17) BAS = BAS / PScore
(18) PScore += 1
(19) EScore /= 2
(20) Else
(21) iBAS=Median($AC(v_i, t_k)| t_k \in [0, T]$)
(22) BAS = (iBAS + $AC(v_i, t_k)$)/2
(23) BASList[k]= BAS
(24) End
(25) BAS = Min(BASList)
(26) If BAS ≅ 0 :
(27) Add v_i to Anomaly List AN[]
(28) End
(29) Final Anomaly Nodes = B

شکل (۴): الگوریتم سنجش رفتاری

۵- نتایج

برای انجام ارزیابی الگوریتم سنجش رفتاری از مجموعه داده VAST 2008[1] استفاده‌شده است. این مجموعه داده شامل داده‌های تماس تلفنی بین گروهی متشکل از ۴۰۰ نفر در طول ۱۰ روز است. طبق مشاهداتی که بروی داده‌ها انجام‌شده است، در این داده‌ها افراد مجرمی وجود دارند که از قبل شناسایی‌شده‌اند و شناسه آن‌ها در این مجموعه داده ، شناسه‌های ۰ ، ۱ ، ۲ ، ۳ ، ۵ ، ۲۰۰ ، ۳۰۰ ، ۳۰۶ ، ۳۰۹ ، ۳۶۰ و ۳۹۷ می‌باشند. هدف از انجام ارزیابی‌ها یافتن بیشترین افراد مجرم (ناهنجار) از این مجموعه داده است. به این نکته باید توجه داشت که برای پیدا کردن این افراد تنها اطلاعاتی که از آن‌ها در مجموعه داده وجود دارد اطلاعات تماس‌های تلفنی آن‌ها است. گراف ایستای این مجموعه داده در ۰آورده شده است. در ارزیابی الگوریتم سنجش رفتاری نسبت به سایر روش‌ها، مدنظر قرار دادن تعداد تقسیم‌بندی هرروز از اهمیت ویژه‌ای برخوردار است. در یک مجموعه داده مانند VAST 2008 تقسیم‌بندی هرروز به ۵ بازه منطقی به نظر می‌رسد ولی در یک مجموعه داده مربوط به ارسال پست الکترونیکی، در نظر گرفتن روزانه فعالیت‌ها منطقی‌تر به نظر می‌رسد. از همین رو الگوریتم تشخیص ناهنجاری باید بتواند با این تغییرات تقسیم‌بندی بازه‌ای خود را تطبیق دهد و کارایی خود را در تشخیص ناهنجاری‌ها از ندهد. اگر الگوریتم تشخیص ناهنجاری فاقد این ویژگی باشد، نمی‌توان آن را در حوزه‌های دیگر استفاده نمود و محدود می‌شود به یک حوزه خاص با تقسیم‌بندی خاص.



شکل (۵): گراف ایستای VAST 2008

روش ارائه‌شده با روش بیزین [۱۵] مقایسه گردیده است. به این منظور کد برنامه توسعه داده شده توسط ارائه‌دهندگان روش بیزین در MATLAB تغییر داده شده تا نتایج موردنظر به دست آید. برای پیاده‌سازی روش پیشنهادی در این مقاله از زبان برنامه‌نویسی پایتون به همراه توابع کتابخانه‌ای NetworkX استفاده گردیده است.

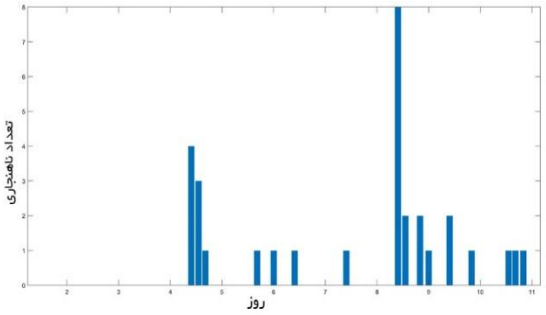
در پژوهش انجام‌شده با روش بیزین، هرروز به ۵ بازه و درمجموع ۱۰ روزبه ۵۰ بازه زمانی تقسیم گردیده است. درروش بیزین برای هر بازه تعداد گره را به‌عنوان ناهنجار مشخص می‌نماید. سپس در هر بازه‌ای که تعداد گره‌های ناهنجار بیشتری وجود داشته باشد آن بازه را به‌عنوان نقطه زمانی رخداد ناهنجاری مشخص کرده و گره‌های شناسایی‌شده در آن بازه را به‌عنوان خروجی ارائه می‌نماید. در ۰تعداد گره‌های ناهنجار شناسایی شده و همچنین گره‌هایی که دو روش نتوانسته‌اند آن‌ها را شناسایی نمایند نشان داده شده است.

[1] http://www.cs.umd.edu/hcil/VASTchallenge08

در ۰تعداد گره‌های ناهنجار شناسایی شده توسط روش بیزین در هر یک از بازه‌های زمانی نمایش داده شده است. همان‌طور که ذکر شد درروش بیزین بازه زمانی دارای بیشترین تعداد ناهنجاری، را به‌عنوان نقطه زمانی رخداد ناهنجاری انتخاب و گره‌های شناسایی شده را به‌عنوان گره‌های ناهنجار انتخاب می‌نماید.

جدول (۱): مقایسه گره‌های شناسایی شده توسط روش پیشنهادی و بیزین در ۵۰ بازه زمانی

	تعداد گره‌های ناهنجار	گره‌های شناسایی نشده
روش بیزین	۸	۰، ۳ و ۲۰۰
روش پیشنهادی	۹	۰ و ۲۰۰



شکل (۶): گره‌های ناهنجار شناسایی شده در هر بازه زمانی با روش بیزین

همان‌طور که در ۰ نشان داده شده است، بیشترین تعداد ۸ گره است که کل این گره‌ها در تقسیم‌بندی فعلی به‌درستی گره ناهنجار تشخیص داده شده‌اند. اما کل گره‌های شناسایی شده در ۵۰ بازه زمانی با حذف گره‌های تکراری برابر ۲۲ گره است. در ۰ تعداد گره‌های ناهنجار شناسایی شده در هر بازه زمانی توسط روش پیشنهادی مشاهده می‌گردد. درروش پیشنهادی با حذف گره‌های تکراری، ۱۴ گره به‌عنوان ناهنجار شناسایی شده‌اند که از بین این ۱۴ گره، ۹ گره به‌درستی به‌عنوان ناهنجار شناسایی شده است. در ۰ این دو روش بر اساس شاخص فراخوانی[1] (R) و دقت[2] (P) مقایسه شده‌اند. این دو معیار به‌صورت زیر تعریف می‌شوند.

$$P = TP/(TP+FP)$$

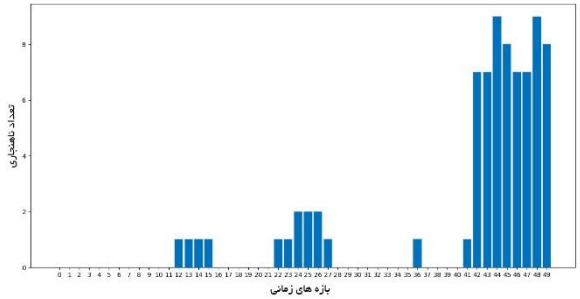
$$R = TP/(TP+FN)$$

که در آن TP مثبت صحیح[3]: تعداد گره‌هایی که به‌درستی هنجار تشخیص داده‌شده‌اند، TN منفی صحیح[4]: تعداد گره‌هایی که به‌درستی ناهنجار تشخیص داده‌شده‌اند، FP مثبت کاذب[5]: تعداد گره‌هایی که به‌اشتباه هنجار تشخیص داده‌شده‌اند و FN منفی کاذب[6]: تعداد گره‌هایی که به‌اشتباه ناهنجار تشخیص داده‌شده‌اند، است.

همان‌طور که در ۰ نشان داده شده است دقت و فراخوانی روش ارائه‌شده در این مقاله از روش بیزین بهتر است.

جدول (۲): مقایسه شاخص دقت و فراخوانی توسط روش پیشنهادی و بیزین در ۵۰ بازه زمانی

	دقت	فراخوانی
روش بیزین	۳۶/۳۶	۷۲,۷۳
روش پیشنهادی	۶۴/۲۹	۸۱/۸۲



شکل (۷): گره‌های ناهنجار شناسایی شده در هر بازه زمانی با روش پیشنهادی

همان‌طور که مطرح شد تعداد نقاط زمانی یا تعداد گراف‌ها در دنباله گراف در کارایی و مقیاس‌پذیر روش ارائه‌شده مؤثر است. در مقایسه صورت گرفته دو روش در فوق هرروز به پنج بازه تقسیم شده و کلاً بررسی در ۵۰ نقطه زمانی یا تصویر گراف صورت گرفته است. در ۰ برای تعداد بازه‌های زمانی مختلف گره‌های ناهنجار شناسایی شده مقایسه گردیده است. همان‌طور که در ۰مشاهده می‌گردد، روش پیشنهادی عملکرد مناسب‌تری داشته است. روش بیزین برای بازه‌های بیش از ۱۰۰ تعداد تمامی گره‌های شناسایی شده نادرست بوده است. با بررسی بیشتر مشخص می‌گردد که در تعداد بازه‌های ۴۰ تا ۱۰۰ بازه، بازه‌ای که بیشترین گره ناهنجار را داشته است در روز هشتم واقع شده است. اما بیش از ۱۰۰ بازه دیگر بازه دارای بیشترین در روز هشتم نبوده و همه گره‌های شناسایی شده نادرست بوده‌اند. این در حالی است که در تعداد بازه ۴۰ تا ۱۰۰ تمامی گره‌های شناسایی شده در بیشترین بازه درست بوده‌اند و در تعداد بازه بیش از ۱۰۰ گرچه تعداد گره‌های شناسایی شده در بازه‌های روز هشتم بیشترین نبوده‌اند اما گره‌های درستی بوده‌اند.

[1] Recall
[2] Precision
[3] True Positive
[4] True Negative
[5] False Positive
[6] False Negative

جدول (۳): مقایسه تعداد گره‌های ناهنجار مثبت صحیح توسط روش پیشنهادی و بیزین در بازه‌های زمانی مختلف

روش تعداد بازه‌های	روش بیزین	روش پیشنهادی
۴۰	۸	۸
۵۰	۸	۹
۷۰	۸	۹
۸۰	۷	۸
۱۰۰	۷	۸
۱۲۰	۰	۷
۱۶۰	۰	۶
۲۰۰	۰	۷
۲۴۰	۰	۵

۶- نتیجه‌گیری

نظر به اهمیت شناسایی ناهنجاری‌ها در شبکه‌های اجتماعی پویا، در این مقاله یک روش مبتنی بر گراف ارائه گردید. در این روش با استخراج شاخص‌های مرکزیت گراف و محاسبه میانگین نرمال شده این شاخص‌ها، معیار فعالیت برای هر فرد در شبکه محاسبه گردید. با اندازه‌گیری معیار فعالیت در دنباله‌ای از گراف‌ها رفتار هر گره موردسنجش قرار گرفت. انتخاب تعداد نقاط رمانی یا تعداد گراف‌ها در دنباله گراف‌ها چالش مهم تحلیل گراف‌های پویا هست. روش ارائه‌شده بر روی مجموعه داده VAST 2008 مورد ارزیابی قرار گرفت. نتایج نشان داد که روش ارائه‌شده تعداد ناهنجاری بیشتری را نسبت به روش بیزین با دقت ۶۴/۲۹ و فراخوان ۸۱/۸۲ یافته است. علاوه بر این روش ارائه‌شده وابستگی کمتری نسبت به تعداد نقاط زمانی داشت. ازاین‌رو روش ارائه‌شده در این مقاله نتایج مناسب‌تری را ارائه داد.

گرچه کارایی یک روش شناسایی ناهنجاری‌ها در شبکه‌های اجتماعی وابسته به نوع داده آن است. بااین‌حال در ادامه پژوهش سعی خواهد که با مجموعه داده‌های دیگر نیز روش ارزیابی گردد. علاوه بر این در این مقاله از میانگین شاخص‌های مرکزیت استفاده گردید که بررسی اثر تغییرات هر یک از شاخص‌ها به‌طور جداگانه و همچنین وابسته به سایر شاخص‌ها جز کارهای آینده این پژوهش خواهد بود.

۷- مراجع

[1] Y. Liu and S. Chawla, "Social Media Anomaly Detection: Challenges and Solutions," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2317-2318, 2015.

[2] R. Yu, X. He, and Y. Liu, "Glad: Group Anomaly Detection In Social Media Analysis," ACM Transactions on Knowledge Discovery from Data ACM T Knowl Discov D, vol. 10, no. 2, pp. 1-22, 2015.

[3] G. Shahmohammadi and S. Kamalizadeh, "Providing a Method for Identifying Phishing Website of Internet Payment Service, " Electronic and Cyber Defense, vol. 4, no. 3, pp. 11-26, 2016. [Online]. Available: https://ecdj.ihu.ac.ir/article_200116_a099f2e1cdcda998bb5178d572a36436.pdf.

[4] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," ACM computing surveys ACM Comput. Surv., vol. 41, no. 3, pp. 1-58, 2009.

[5] M. E. Newman, D. J. Watts, and S. H. Strogatz, "Random Graph Models of Social Networks," Proceedings of the national academy of sciences, vol. 99, no. 1, pp. 2566-2572, 2002.

[6] L. Akoglu, H. Tong, and D. Koutra, "Graph Based Anomaly Detection and Description: A Survey," Data mining and knowledge discovery Data Min Knowl Disc, vol. 29, no. 3, pp. 626-688, 2015.

[7] M. Zamini and S. M. H. Hasheminejad, "A Comprehensive Survey of Anomaly Detection in Banking, Wireless Sensor Networks, Social Networks, and Healthcare, " Intelligent Decision Technologies FRONT ARTIF INTEL AP, vol. 13, no. 2, pp. 229-270, 2019.

[8] S. Ranshous, S. Shen, D. Koutra, S. Harenberg, C. Faloutsos, and N. F. Samatova, "Anomaly Detection in Dynamic Networks: A Survey, " Wiley Interdisciplinary Reviews: Computational Statistics, vol. 7, no. 3, pp. 223-247, 2015.

[9] R. Hassanzadeh, R. Nayak, and D. Stebila, "Analyzing the Effectiveness of Graph Metrics for Anomaly Detection in Online Social Networks," in International Conference on Web Information Systems Engineering, Springer, pp. 624-630, 2012.

[10] K. Das, S. Samanta, and M. Pal, "Study on Centrality Measures in Social Networks: A Survey," Social Network Analysis and Mining, vol. 8, no. 1, p. 13, 2018.

[11] F. Riquelme, P. Gonzalez-Cantergiani, X. Molinero, and M. Serna, "Centrality Measure in Social Networks Based on Linear Threshold Model," Knowledge-Based Systems Knowl-Based Syst., vol. 140, pp. 92-102, 2018.

[12] M. Z. Al-Taie and S. Kadry, Python for Graph and Network Analysis, Springer, 2017.

[13] A. Hagberg, D. Schult, and P. Swart, "NetworkX Reference, Release 2.3," NetworkX Developers, Last updated on Apr. 11, 2019.

[14] M. Mirzaee and A. Mahabadi, "Anomaly Detection Method in Attributed Graph, Based on Community Detection and Feature Selection," Electronic and Cyber Defense, 2019.

[15] N. A. Heard, D. J. Weston, K. Platanioti, and D. J. Hand, "Bayesian Anomaly Detection Methods for Social Networks," The Annals of Applied Statistics, vol. 4, no. 2, pp. 645-662, 2010.

[16] M. G. Vigliotti and C. Hankin, "Discovery of anomalous behaviour in temporal networks," Social Networks, vol. 41, pp. 18-25, 2015.

[17] V. K. Giri and S. Sachdeva, "Anomaly Detection in Social Networks," in 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, pp. 698-703, 2019.

[18] R. Kaur and S. Singh, "A Comparative Analysis of Structural Graph Metrics to Identify Anomalies in Online Social Networks," Computers & Electrical Engineering Comput. Electr. Eng., vol. 57, pp. 294-310, 2017.

Journal of Electronical & Cyber Defence
Vol. 9, No. 1, 2021, Serial No. 33

Anomaly Detection in Dynamic Social Networks Based on Behavioral Measurements

S. Ghaderi Piraghum, M. Sakhaei-nia*, M. Mansoorizadeh

*Computer Engineering Department, Bu-Ali Sina University

(Received: 28/04/2020, Accepted: 26/10/2020)

ABSTRACT

Since the detection of anomalies in dynamic social networks takes place in a sequence of graphs over time, in addition to the storage management challenge, the detection process is difficult due to the slow evolution of graphs. A number of graphs are selected in the specified time frame, and by examining the changes of these graphs, the possible anomalies are detected. Therefore, choosing the number of time points (graphs) in the sequence of graphs is an important challenge in the detection of anomalies. In this paper, a novel method is proposed to detect anomalies based on structural data extracted from dynamic social network graphs. By extracting the centrality indicators from the network graph and their normalized mean, the activity criterion for each individual has been defined. Over time, changes in the activity criterion for each individual are measured and marked as the possibility of normal or abnormal behavior. If the individual's behavior measure exceeds a certain threshold, it is reported as an anomaly. The results show that the proposed method detects more anomalies with the accuracy and recall of 64.29 and 81.82 respectively, for the VAST 2008 data set. It also, detects more anomalies by selecting different number of time points in the graph sequence.

Keywords: Social Networks, Anomaly Detection, Dynamic Graph, Behavioral Measurement

* Corresponding Author Email: sakhaei@basu.ac.ir