دوفصلنامه مدیریت مهندسی و رایانش نرم/دوره ۸، شماره ۱، بهار و تابستان ۱۴۰۱
(DOI): 10.22091/jemsc صفحات ۲۶۹- ۲۴۵
JEMSC.QOM
مدیریت مهندسی و رایانش نرم ۱۳۹۴

شناسایی افراد ناهنجار در شبکه‌های اجتماعی موقعیت‌محور با استفاده از معیارهای گراف*

فاطمه عدالتی[۱]
مرتضی رموزی[۲]

چکیده

بررسی شبکه‌های اجتماعی برای شناسایی کاربران ناهنجار با توجه به محبوبیت این شبکه‌ها امری ضروری است. در این مقاله هدف، شناسایی کاربران ناهنجار در شبکه‌های اجتماعی موقعیت‌محور است. برای این منظور گراف ego هر کاربر ساخته شده و پنج متغیر درجه رأس، تعداد یال، جمع وزن یال، مرکزیت بینابینی و مرکزیت بردار ویژه با توجه به وزن‌دهی به یال‌های این گراف محاسبه می‌شود. سپس شش ارتباط بین دو متغیر از این متغیرها تشکیل شده و برای هرکدام از این ارتباط‌ها، معادله خط در دستگاه مختصات بین دو متغیر به دست آمده است. از این معادله برای پیش‌بینی مقدار متغیرها استفاده شده است. متناسب با این پیش‌بینی، امتیاز کاربر مشخص می‌شود و کاربران ناهنجار شناسایی می‌شوند. روش پیشنهادی، ناهنجاری در گراف دوستی، محل سکونت و علایق کاربران را بررسی کرده و نتایج حاکی از این است که روش پیشنهادی توانسته با بررسی امتیاز ساختار ستاره و کلیک در گراف، کاربران ناهنجار را شناسایی کند.

واژه‌های کلیدی: *تحلیل شبکه اجتماعی، تشخیص ناهنجاری، شبکه اجتماعی موقعیت‌محور*

---

* تاریخ دریافت: ۱۳۹۸/۰۴/۰۲؛ تاریخ پذیرش: ۱۳۹۸/۰۵/۱۷.

۱. دانشجوی کارشناسی ارشد، مهندسی نرم افزار، واحد کاشان، دانشگاه آزاد اسلامی، کاشان، ایران **(نویسنده مسئول)**
edalati.fatemeh@yahoo.com
۲. استادیار گروه مهندسی برق کامپیوتر، واحد کاشان، دانشگاه آزاد اسلامی، کاشان، ایران. mromoozi@gmail.com

مقدمه

شـــبکه‌های اجتماعی موقعیت‌محور[1] با حضـــور افراد در مکان‌های مختلف و امتیاز کاربران به مکان‌هایی که بازدید کرده‌اند، ســـعی در جمع‌آوری داده‌های مربوط به بازدید کاربران دارد. در این شـبکه اجتماعی، رابطه بین افراد بر اسـاس برچسـب موقعیت مکانی به رسانه‌هایی از قبیل ویدئو، تصویر، عکس و متن در دنیای واقعی ایجادشده است (حسن‌زاده و نیاک، ۲۰۱۳). یک شبکه اجتماعی موقعیت‌محور به معنی اضافه کردن و اشتراک‌گذاری مکان در شبکه‌های اجتماعی نیست بلکه به معنی ساختار اجتماعی است که بر اساس حضور افراد در مکان‌های دنیای واقعی شـــکل‌گرفته اســـت (حســـن‌زاده و نیاک، ۲۰۱۳). تجزیه و تحلیل شـبکه‌های اجتماعی به ما جهت کشـف روابط و الگوهای بین افراد کمک می‌کنند. برخی از مهم‌ترین مسائل در تجزیه و تحلیل شبکه‌های اجتماعی، تجزیه و تحلیل مرکزیت، تشخیص جامعه، انتشار اطلاعات، حداکثر نفوذ، پیش‌بینی لینک، سیستم‌های پیشنهاد دهنده و تشخیص ناهنجاری هستند.

در شـــبکه‌های اجتماعی منظور از ناهنجاری، هرگونه رفتار غیرمنتظره از طرف یک کاربر یا گروهی از کاربران اســـت که نســـبت به رفتار طبیعی کاربران غیرمعمول باشـــد. ناهنجاری‌ها در شبکه‌های اجتماعی زمانی رخ می‌دهد که رفتار کاربران، الگوهای شبکه‌ای غیرمعمول را شکل می‌دهد. برای مثال، در یک شبکه اجتماعی اینترنتی، اکثریت کاربران به دنبال الگوی دوسـتان یک کاربر که اغلب دوسـتان هم هسـتند بوده و اقلیت به دنبال الگوی دسته یا نزدیک به دسته یا ستاره یا نزدیک به ستاره هستند. این دو الگوی اقلیت که در آن همه همســـایگان بیشـــتر با یکدیگر ارتباط داشـــته یا عمدتاً قطع ارتباط دارند می‌توانند کاندیداهای احتمالی برای ناهنجاری باشد.

متناســـب با تعریف ناهنجاری در کائور و ســـینگ (۲۰۱۵) روشـــی برای شـــناســـایی ناهنجاری ارائه شـــده اســـت. این روش بر پایه متغیرهایی برای هر کاربر اســـت که برای

[1] Location Based Social Network

اندازه‌گیری این متغیرها گراف ego برای هر کاربر تشکیل می‌شود. در این گراف متغیرهایی ازجمله درجه، تعداد یال‌ها و مرکزیت بینابینی هر کاربر محاسبه شده است.

در این پژوهش تمرکز روی شناسایی ناهنجاری در شبکه‌های اجتماعی موقعیت‌محور می‌باشد. در تشخیص ناهنجاری در این نوع از شبکه‌های اجتماعی، مسئله مکان باید در نظر گرفته شود. در این تحقیق از داده‌های شبکه اجتماعی موقعیت‌محور foursquare استفاده شده است. بدین ترتیب با الهام از کائور و سینگ (۲۰۱۵) ابتدا گراف ego هر کاربر ساخته می‌شود و متناسب با محل سکونت کاربران و بازدید کاربران از مکان‌های مختلف، این گراف وزن‌دهی می‌شود. متغیرهای مورد بررسی که شامل درجه گره، تعداد یال‌ها در گراف ego، جمع وزن همسایه بودن نودهای گراف ego به لحاظ محل سکونت، مرکزیت بینابینی گره اصلی و مرکزیت بردار مقداردهی می‌شود. سپس ارتباط بین دو به دوی این متغیرها با استفاده از معادلات خطی و نمایی مدلسازی می‌شود و با استفاده از این مدل و مقایسه مقدار واقعی و پیش‌بینی شده با این معادله خط امتیاز ناهنجاری کاربران تعیین می‌شود. کاربران با بیشترین و کمترین مقدار امتیاز به عنوان کاربران ناهنجار با استفاده از ساختارهای ستاره و کلیک در گراف ego شناسایی می‌شوند. نتایج به کارگیری روش پیشنهادی روی مجموعه داده حاکی از آن است که استفاده از متغیرهای محل سکونت کاربران و امتیاز کاربران منجر به شناسایی بهتر رفتارهای ناهنجار کاربران شده است.

در این مقاله در بخش دوم مقاله مروری بر کارهای مرتبط با تشخیص ناهنجاری داریم سپس در بخش سوم روش پیشنهادی معرفی می‌شود و در بخش چهارم نتایج روش پیشنهادی ارائه و بررسی می‌شود.

پیشینه پژوهش

شبکه‌های اجتماعی آنلاین روشی مؤثر برای سازمان‌دهی و به اشتراک‌گذاری مطالب افراد فراهم می‌کند. تجزیه و تحلیل شبکه‌های اجتماعی جهت کشف روابط بین افراد به ما کمک می‌کند و درک الگوهای اصلی را که در این گراف اجتماعی تعبیه شده‌اند، فراهم می‌کند (اسکات ۲۰۱۱). یکی از مهم‌ترین وظایف در تجزیه و تحلیل شبکه اجتماعی

تشخیص ناهنجاری است. در عرصه شبکه‌های اجتماعی، ناهنجاری رفتار غیرمنتظره از یک کاربر یا گروهی از کاربران است که نسبت به رفتار طبیعی کاربران در گراف غیرمعمول است. اکثراً و به طور خاص، ناهنجاری‌ها در شبکه‌های اجتماعی زمانی رخ می‌دهد که رفتار کاربران، الگوهای شبکه‌ای غیرمعمول را شکل می‌دهد. برای مثال، در یک شبکه اجتماعی اینترنتی، اکثریت کاربران به دنبال الگوی دوستان یک کاربر هستند که اغلب دوستانش هستند و اقلیت به دنبال الگوی دسته یا نزدیک به دسته یا ستاره یا نزدیک به ستاره هستند (حسن‌زاده و نیاک، ۲۰۱۳؛ آکوگلا و فالوتسوس، ۲۰۱۰) تشخیص ناهنجاری مبتنی بر گراف، مسئله پیدا کردن ناهنجاری داده‌ها با استفاده از تکنیک استخراج گراف است. گراف‌کاوی که حوزه تحقیقاتی محبوب در سال‌های اخیر بوده است، روش جدیدی برای استخراج دانش مفید از داده‌های گراف با استفاده از تکنیک‌هایی چون یادگیری ماشین، داده‌کاوی، آمار، تشخیص الگو و نظریه گراف است. تشخیص ناهنجاری یک میدان گسترده است که به زمینه خوبی برای تحقیق در طیف گسترده‌ای از حوزه‌های کاربردی تبدیل شده است.

روش‌های مبتنی بر تشخیص ناهنجاری برای پیدا کردن یک واحد رفتار ناهنجار مانند گره، یال یا زیرگراف تلاش می‌کنند. تشخیص ناهنجاری از داده‌های گراف برای پرداختن به مشکلاتی از قبیل نظارت بر فعالیت در شبکه‌های اجتماعی، تشخیص نفوذ به شبکه، نظارت شبکه، فیلتر هرزنامه، تصویربرداری پزشکی، تجزیه و تحلیل شبکه ژن، تشخیص شیوع بیماری، نظارت بر شبکه‌های حسگر، نظارت بر محیط‌زیست و تشخیص نرم‌افزارهای مخرب ضروری است. اگرچه مشکل تشخیص ناهنجاری‌های سنتی به طور فعال در آمار و داده‌کاوی در طول چند دهه گذشته مورد مطالعه قرارگرفته است، تمرکز کمتری روی تشخیص ناهنجاری در داده‌های مبتنی بر گراف شده است. تشخیص ناهنجاری در شبکه‌های اجتماعی دارای ویژگی‌های مختلفی چون ماهیت شبکه‌های ورودی و نوع ناهنجاری است که در این پژوهش گراف‌های ایستا و با ویژگی بررسی شده است. دو روش عمده برای تشخیص ناهنجاری در این نوع گراف‌ها، روش‌های مبتنی بر تشخیص

جامعه و مبتنی بر ساختار گراف است. روش مبتنی بر جامعه برای تشخیص ناهنجاری، گره‌ها یا یال‌هایی را که محل اتصال جوامع مختلف هستند به عنوان خوشه یا جامعه پیدا می‌کند و ناهنجاری را به عنوان گره یا یال‌هایی که به هیچ جامعه‌ای اتصال ندارند تعریف می‌کند. رویکردهای مبتنی بر ساختار گراف نیز مسئله تشخیص ناهنجاری را به مسئله تشخیص داده پرت تبدیل می‌کنند (آکوگلا، چندی و فالوتسوس، ۲۰۱۳؛ هندرسون و همکاران، ۲۰۱۱). یکی از اولین آثار در تشخیص ناهنجاری، رویکرد مبتنی بر جامعه است (آکوگلا، مک گلاهان و فالوتسوس، ۲۰۱۰). در (چاکرابارتی، ژان و فالوتسوس، ۲۰۰۴) یک رویکرد مقیاس‌پذیر، برای تشخیص یال غیرعادی و گره ارائه شده است. در این روش به طور خودکار گراف را با سازمان‌دهی تکرارشونده در سطر و ستون از ماتریس مجاورت به خوشه تقسیم می‌کند. لبه‌هایی که به هیچ خوشه‌ای تعلق ندارند ناهنجاری هستند. در (تانگ و لین، ۲۰۱۱) برای سازمان‌دهی مجدد ماتریس مجاورت، گراف را به بلوک‌های همگن تقسیم‌بندی کرده‌اند. این بلوک‌ها شامل گره‌هایی هستند که در مقایسه با گره‌های باقیمانده در شبکه، به‌طور قوی به هم متصل می‌شوند. در (تانگ و لین، ۲۰۱۱) نویسندگان یک چارچوب تجزیه نامنفی ماتریس[1]، برای مشخص کردن گره‌ها و یال‌های غیرطبیعی پیشنهاد می‌دهند. در (یانگ، شن، وانگ، گونگ و دانگ، ۲۰۱۵) چارچوبی بر اساس گراف دوبخشی و خوشه‌بندی برای کشف کاربران غیرعادی و پیام‌ها در میکروبلاگ‌ها توسعه داده شده است. (گوتا، آگو و آگاروال، ۲۰۱۴) مکانیسمی به نام کشف الگوریتم داده پرت در زیرگراف را توسعه داد. در مورد پژوهش‌های داخلی صورت گرفته در مورد موضوع کمتر مقالات جدید پیدا شد لیکن در (نادی، ۲۰۱۶) ناهنجاری در شبکه‌های اجتماعی موقعیت‌محور بررسی شده است.

[1] Nonnegative matrix factorization

روش‌شناسی پژوهش

کائور و سینگ (۲۰۱۵) روشی برای شناسایی کاربران ناهنجار معرفی کرده‌اند. در این مقاله با الهام از این روش، راهکاری جهت شناسایی ناهنجاری در شبکه‌های اجتماعی موقعیت‌محور ارائه شده است. مراحل سیستم پیشنهادی عبارت است از:

۱. نگهداری داده‌های مورد استفاده
۲. آماده‌سازی داده‌ها
۳. آماده‌سازی گراف شبکه اجتماعی و شبکه ego کاربران
٤. مقداردهی به متغیرهای مورد استفاده
۵. شناسایی ارتباط بین متغیرهای مورد استفاده
٦. شناسایی کاربران ناهنجار
۷. بررسی و تحلیل نتایج

در ابتدا لازم است داده‌های موجود برای اجرای مراحل آماده شوند. سپس این داده‌ها در ساختمان داده مشخص نگهداری شود. پس از جمع‌آوری داده‌ها، نگهداری و آماده‌سازی داده‌ها به‌صورت گراف، شبکه و ساختار ego برای هر کاربر تشکیل می‌شود. سپس در ساختار ego، دوستان یک کاربر و ارتباط بین این دوستان در نظر گرفته می‌شود. در این گراف برای هر کاربر متغیرهای درجه رأس، تعداد یال‌ها، جمع وزن یال‌ها، مرکزیت بینابینی و مرکزیت بردار ویژه در گراف ego محاسبه می‌شود. سپس ارتباط بین این دو متغیر که در محور مختصات x و y هستند به صورت یک معادله خطی و نمایی به دست می‌آید. با استفاده از این فرمول و تعیین امتیاز برای هر کاربر، کاربران ناهنجار مشخص می‌شوند.

مرحله نگه‌داری و آماده‌سازی داده‌ها

داده‌های مورد استفاده در شبکه‌های اجتماعی موقعیت‌محور شامل ارتباط اجتماعی بین کاربران، محل سکونت کاربران، مختصات مکان‌ها، امتیاز و ساعت و تاریخ بازدید از هر مکان می‌باشد. نمونه داده‌های مورد استفاده مطابق جدول ۱ می‌باشد.

جدول ۱. نمونه داده‌های موجود در مجموعه داده

نوع داده	نمونه داده

شبکه اجتماعی کاربران

کاربر دوم	کاربر اول
۱۰	۱
۱	۱۰
۱۱	۱
۱	۱۱
۱۲	۱

محل سکونت کاربران

مختصات y کاربر	مختصات x کاربر	شناسه کاربر
-۹۲/۱۰۰۵	۴۶/۷۸۶۶	۱۹
-۹۵/۳۸۳۱	۲۹/۷۶۲۸	۷۱۶۳۴۴
-۸۲/۴۶۵۱	۲۷/۹۴۹۴۴	۵
۳۵/۵۲۴۷	۳۲/۷۸۹۷	۱۲۰۷۶۷۰
-۷۷/۲۹۷۵	۱۸/۱۰۹۵	۱۰۷۶۴۵۷

مختصات مکان‌ها

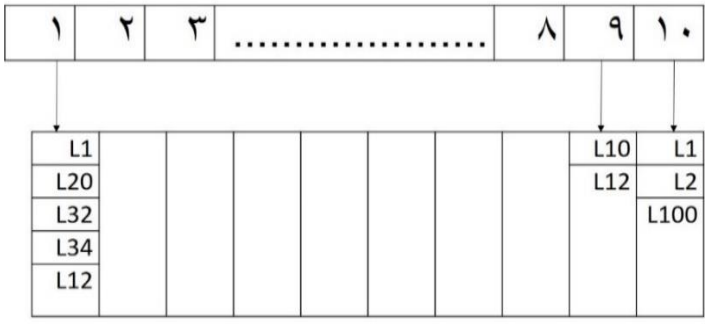
مختصات y مکان	مختصات x مکان	شناسه مکان
-۷۳/۵۶۴۳	۴۵/۵۱۵۴۷	۱۰۱۷۶۲
-۹۴/۶۰۴۴	۳۸/۹۶۲۱۷	۴۴۹۰۶۰
-۸۷/۶۲۵۸	۴۱/۸۸۹۶۲	۲۴۸۶۹

امتیاز به مکان‌ها

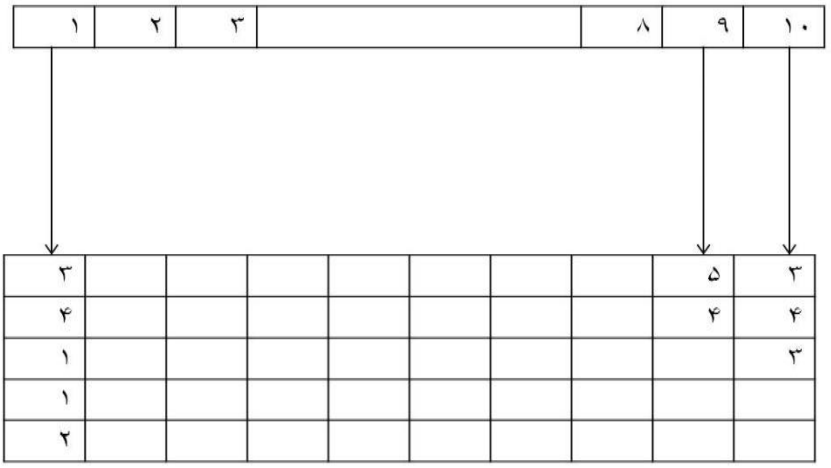
امتیاز	شناسه مکان	شناسه کاربر
۵	۱	۱
۴	۵۱	۱
۲	۵۱	۱
۵	۵۱	۱

	ثانیه	دقیقه	ساعت	روز	ماه	سال	شناسه مکان	شناسه کاربر
ساعت و تاریخ بازدید	۱۸	۳۷	۱۷	۲۱	۴	۲۰۱۲	۵۲۲۲	۱۷۶۴۳۹۱
	۴۳	۴۳	۱۷	۲۱	۴	۲۰۱۲	۵۲۲۲	۴۴۶۵۲
	۵۸	۴۲	۱۷	۲۱	۴	۲۰۱۲	۵۲۲۲	۲۱۴۶۸۴۰
	۳۵	۴۲	۱۷	۲۱	۴	۲۰۱۲	۵۲۲۲	۲۱۴۶۸۴۳
	۴۰	۳۹	۱۷	۲۱	۴	۲۰۱۲	۵۲۲۲	۲۱۴۶۸۴۶

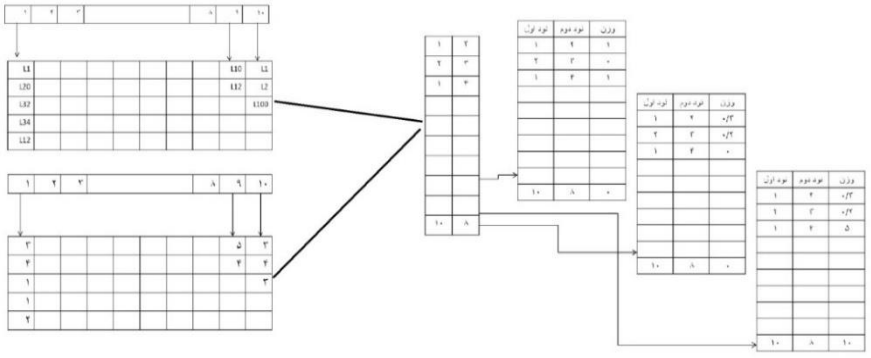
داده‌های مربوط به بازدید کاربران از مکان‌های مختلف مطابق شـــکل (۱) و داده‌های مربوط به امتیاز کاربران به این مکان‌ها مطابق شکل (۲) نگهداری می‌شود و همچنین گراف شبکه اجتماعی کاربران و وزن یال‌های این گراف مطابق شکل (۳) نگهداری می‌شود.



شکل ۱. نحوه نگهداری مکان‌های مشاهده شده توسط هر کاربر



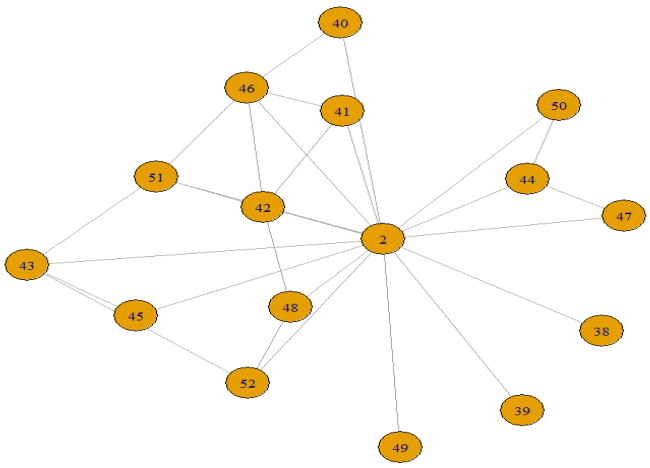
شکل ۲. نحوه نگهداری امتیاز هر کاربر به مکان‌های مشاهده شده توسط هر کاربر



شکل ۳. نگهداری وزن‌های گراف

مرحله آماده‌سازی گراف و شبکه ego کاربران مقداردهی به متغیرها

هدف از این مرحله، ساخت گراف ego از گراف شبکه اجتماعی کاربران است. برای هر کاربر u، دوستان آن کاربر و ارتباط بین دوستان آن کاربر به عنوان یک زیرگراف نگهداری شده است و به آن گراف ego کاربر u گفته می‌شود. نمونه‌ای از گراف ego در شکل ۴ ارائه شده است.



شکل ٤. گراف ego یک کاربر دلخواه

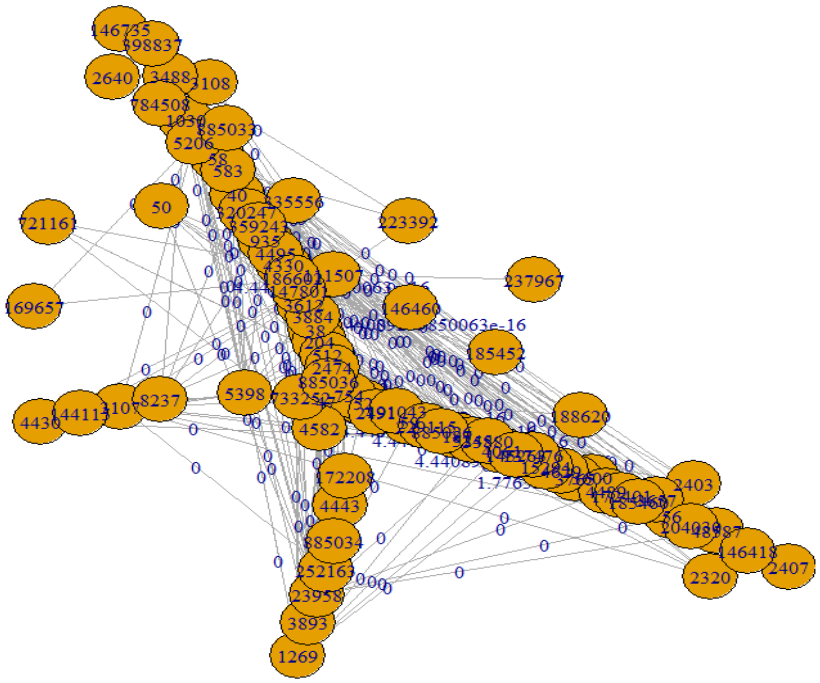
با توجه به گراف ego برای هر کاربر پنج متغیر در نظر گرفته می‌شـــود. این متغیرها عبارتند از:

۱- درجه: برای هر رأس u در گراف ego تعداد یال‌های برخوردی به این رأس درجه آن رأس می‌باشد.

۲- تعداد یال: تعداد یال‌های موجود در گراف ego برای هر کاربر می‌باشد.

۳- جمع وزن یال: برای محاسبه جمع وزن یال در گراف ego ابتدا یال‌های این گراف وزن‌دهی می‌شوند. برای این منظور به ۲ صورت وزن‌دهی انجام گرفته است. در حالت اول وزن یال‌ها در صورتی که دو کاربر از لحاظ محل سکونت همسایه همدیگر باشد یک و در غیر این صـــورت صـــفر در نظر گرفته می‌شـــود. در حالت دوم در صـــورتی که دو کاربر مکان‌های مشترکی را بازدید کرده باشند وزن یال شباهت کسینوسی امتیاز به این مکان‌ها است؛ اما درصورتی‌که دو کاربر مکان مشترکی را ندیده باشند و به لحاظ سکونت همسایه همدیگر باشند وزن یال ۵ و در حالتی که دو کاربر نه مکان مشترکی را دیده‌اند و نه همسایه همدیگر باشـــند وزن یال ۱۰ در نظر گرفته می‌شـــود. نمونه‌ای از این وزن‌دهی در شـــکل ۵

نشـان داده شـده اسـت. این مقادیر برای وزن یال‌ها صـرفاً برای تأثیر دادن همسـایه بودن و نبودن کاربران ازلحاظ محل سکونت با همدیگر در نظر گرفته شده است.



شکل ۵. گراف ego با وزن شباهت کسینوسی برای یک کاربر دلخواه

۴- مرکزیت بینابینی: در گراف ego برای هر کاربر این متغیر مقداردهی می‌شود.

۵- مرکزیت بردار ویژه: در گراف ego برای هر کاربر این متغیر مقداردهی می‌شود.

مرکزیت بینابینی و مرکزیت بردار ویژه به ترتیب طبق رابطه‌های ۱ و ۲ تعریف شـده است که در ابزار R محاسبه شده است. برای محاسبه این فرمول‌ها از کتابخانه Igraph در R استفاده شده است.

رابطه ۱)
$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

در رابطه ۱، σ_{st} وزن کوتاه‌ترین مسیرها از رأس s به رأس t و $\sigma_{st}(v)$ وزن کوتاه‌ترین مسیرها از رأس s به رأس t است که از رأس v عبور می‌کند.

در رابطه ۲ که برای محاسبه مرکزیت بردار ویژه استفاده می‌شود A_{ij} درصورتی که رأس i به رأس j وصل باشد مقدار یک و در غیر این صورت مقدار صفر دارد که برای گراف وزن‌دار این مقادیر به جای صفر و یک، وزن آن یال است و امتیاز مرکزیت گره i ام که با x_i نشان داده می‌شود.

رابطه ۲) $x_i = \frac{1}{\gamma} \sum_{j \in M(i)} x_j = \frac{1}{\gamma} \sum_{j=1}^{N} A_{ij} x_j$

درنهایت برای هر کاربر u این پنج متغیر در محیط R و با استفاده از کتابخانه igraph محاسبه می‌شود.

مرحله شناسایی ارتباط بین متغیرها

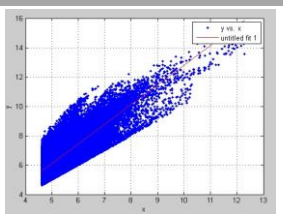
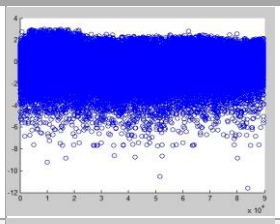
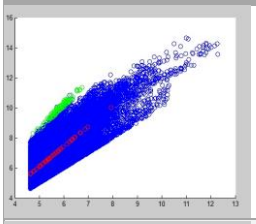
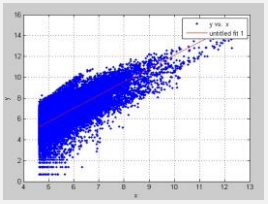
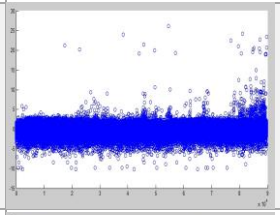
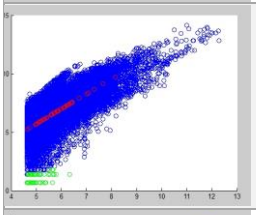
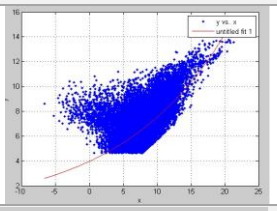
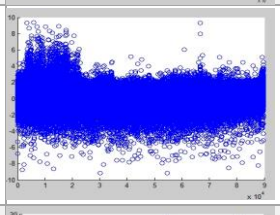
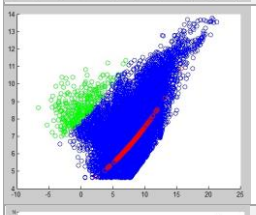
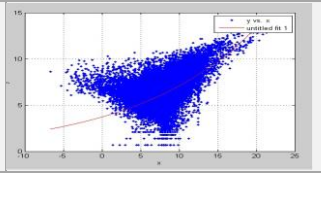
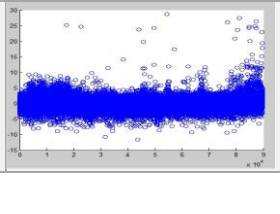
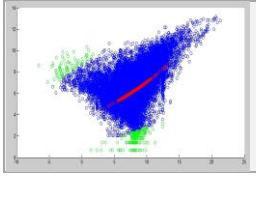
پس از این که در مرحله قبل برای هر کاربر گراف ego آن کاربر را تشکیل دادیم و برای هر کاربر با توجه به گراف ego مقادیر ۵ متغیر به دست‌آمد، در این مرحله لازم است ارتباط بین این متغیرها مدلسازی شود. در اینجا فقط روی ارتباطات دو به دو متغیرها تمرکز شده است. از این رو، ۱۰ ارتباط متمایز می‌تواند وجود داشته باشد؛ اما برای استخراج نوع ارتباط، هر دو متغیر طبق جدول ۲ در دستگاه معادلات دوبعدی در نظر گرفته شده و با استفاده از درون‌یابی، معادله منحنی مربوطه استخراج می‌شود.

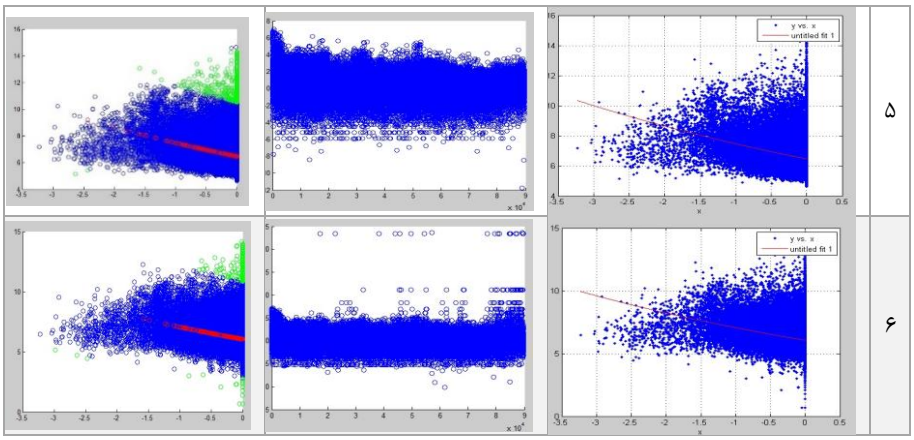
جدول ۲ ارتباط بین متغیرها

شماره ارتباط	محور x	محور y
۱	درجه رأس	تعداد یال
۲	درجه رأس	جمع وزن یال
۳	مرکزیت بینابینی	تعداد یال
٤	مرکزیت بینابینی	جمع وزن یال
٥	مرکزیت بردار ویژه	تعداد یال
٦	مرکزیت بردار ویژه	جمع وزن یال

شکل‌های موجود در ستون اول جدول ۳ نحوه این کار را به خوبی آشکار می‌کند. در این شکل جدول ستون معادله خط مقادیر ۶ متغیر در جدول ۲ را نشان می‌دهد و در همین شکل خط ترسیم شده حاصل نگاشت آن‌ها به دستگاه معادلات دوبعدی است. منحنی منطبق با این نقاط از طریق معادله منحنی برازش بین این دو متغیر موجود در هر سطر جدول بالا به دست‌آمده است. در همین شکل، ستون دوم امتیاز کاربران می‌باشد و ستون سوم کاربرانی که کمترین و بیشترین امتیاز را دارند بارنگ‌های سبز و قرمز مشخص‌شده‌اند. در ادامه نحوه محاسبه امتیاز برای هر کاربر توضیح داده شده است.

جدول ۳ نمودار ارتباط بین متغیرها

	معادله خط	امتیاز کاربران با درجه رأس بزرگ‌تر از ۱۰۰	کاربران با بالاترین امتیاز پایین امتیاز
۱			
۲			
۳			
۴			



شناسایی کاربران ناهنجار

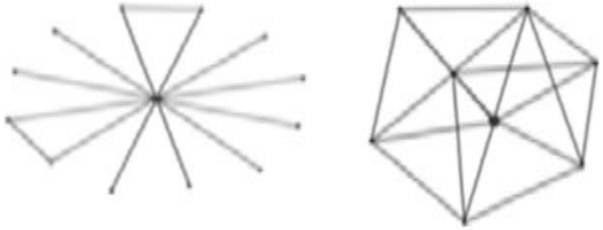
پس از استخراج معادله منحنی به ازای هر دو متغیر متمایز، حال مقدار پیش‌بینی شده برای مقدار ورودی x به وسیله معادله خط، به دست می‌آید. این مقدار را با مقدار واقعی این ورودی x مقایسه می‌کنیم و به هر کاربر امتیاز می‌دهیم. این امتیاز با رابطه ۱ محاسبه می‌شود که در رابطه ۳ برای گراف‌های بدون وزن معرفی و در نظر گرفته شده است.

$$ascore(i) = \frac{\max\left(y_i, Cx_i^{\theta}\right)}{\min\left(y_i, Cx_i^{\theta}\right)} * \log\left(|y_i - Cx_i^{\theta}| + 1\right) \quad \text{رابطه ۳)}$$

در رابطه ۳ مقدار ascore(i) نشان‌دهنده امتیاز کاربر i ام‌است. y_i مقدار واقعی خروجی متغیر کاربر i ام به ازای ورودی x_i می‌باشد و Cx_i^{θ} مقدار خروجی متغیر کاربر i ام به ازای ورودی x_i است که توسط معادله خط پیش‌بینی می‌شود.

در ادامه کاربران بر اساس مقادیر امتیاز مرتب می‌شوند. با توجه به رابطه ۳، افراد با بالاترین امتیاز و همچنین کمترین امتیاز مستعد وجود ناهنجاری هستند. به عنوان مثال ۱۰۰ کاربر ابتدا و ۱۰۰ کاربر انتها در لیست مرتب شده مقادیر امتیازها در نمودار مشخص می‌شوند؛ اما برای شناسایی با دقت بیشتر به ازای این کاربران کاندید، گراف ego این کاربران طبق روش موجود در (کائور و سینگ، ۲۰۱۵) در نظر گرفته می‌شود. بر اساس شکل ۶ در این دو مدل گراف می‌توان ناهنجاری را تشخیص داد. در گراف ego این کاربران کاندید نسبت درجه به تعداد یال‌ها به دست آمده است و تعداد افرادی که مقدار

این نسبت برای آن‌ها در یک محدوده از پیش تعیین شده باشد، به عنوان کاربران ناهنجار شناسایی می‌شوند. برای تعیین مقادیر این محدوده از ساختار کلیک و ستاره استفاده شده است. در بررسی رفتار ناهنجار برای گراف ego دو ساختار کلیک و ستاره در نظر گرفته می‌شود که این ساختارها مطابق شکل ۶ می‌باشد.

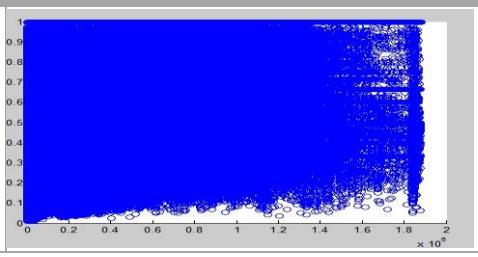
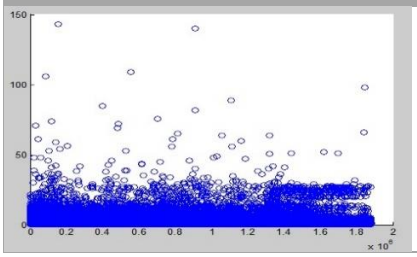


شکل ٦. دو گراف با ساختارهای ستاره و کلیک

در واقع برای هر کاربر نسبت درجه رأس هر کاربر به تعداد یال‌های موجود در گراف ego آن کاربر نشان‌دهنده رفتارش است. برای این منظور نیز پس از این که ۱۰۰ کاربر با امتیاز بالا و ۱۰۰ کاربر با امتیاز پایین شناسایی شدند، در این مرحله نسبت درجه رأس گراف ego هر کاربر به تعداد یال‌های موجود در گراف ego آن کاربر در شکل ۱-۴ نشان داده شده است. در این شکل محور افقی نشان‌دهنده کاربران و محور عمودی نشان‌دهنده نسبت بیان شده است. از آنجا که هم تعداد یال‌های گراف ego و هم وزن این‌ها جداگانه حساب شده است، بنابراین دو مقدار برای هر کاربر به دست می‌آید که یک مقدار در واقع حاصل تقسیم درجه رأس کاربر در گراف ego به تعداد یال‌ها است و یک مقدار هم درجه رأس کاربر در گراف ego به تعداد یال‌ها با وزن یک می‌باشد. در خصوص مقدار در گراف ego، وزن یال‌هایی که کاربران آن به لحاظ محل سکونت همسایه همدیگر هستند یک و مقدار وزن یال‌هایی که کاربران آن به لحاظ محل سکونت همسایه همدیگر نیستند صفر در نظر گرفته می‌شود. در شکل‌های جدول ۴ این مقدار نشان داده شده است. در این نمودارها نسبت درجه به تعداد یال‌ها در گراف ego هر کاربر نشان داده شده است. محور

افقی شـــماره کاربر و محور عمودی نشـــان‌دهنده تقســـیم مقدار درجه رأس کاربر به تعداد یال‌ها در گراف ego این کاربر است.

جدول ٤. بررسی مقادیر درجه و تعداد یال در گراف

نسبت درجه به تعداد یال‌ها در گراف **ego** برای هر کاربر	نسبت درجه به جمع وزن تعداد یال‌ها در گراف **ego** برای هر کاربر
	

یافته‌های پژوهش

اعتبارسنجی روش پیشنهادی

مجموعه داده مورد استفاده و ابزار پیاده‌سازی

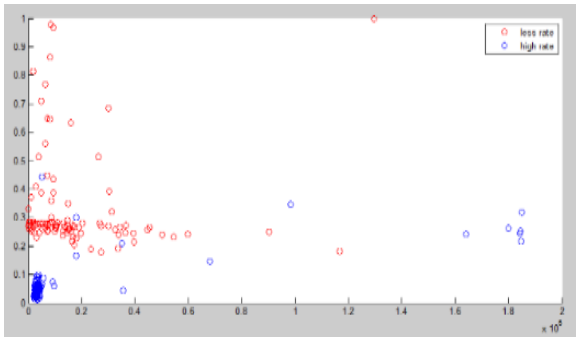
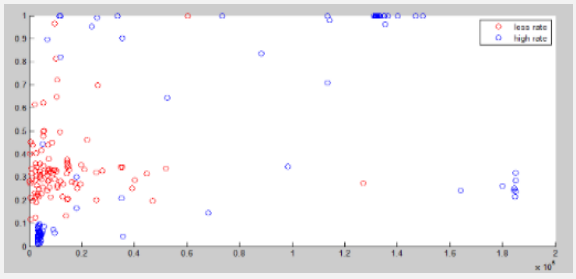
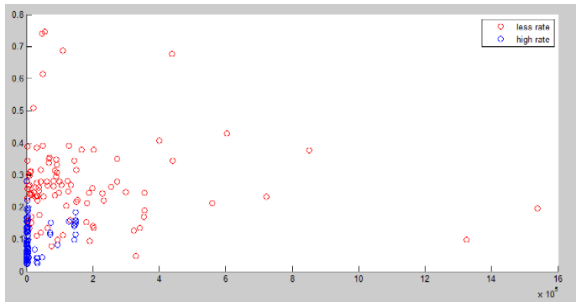
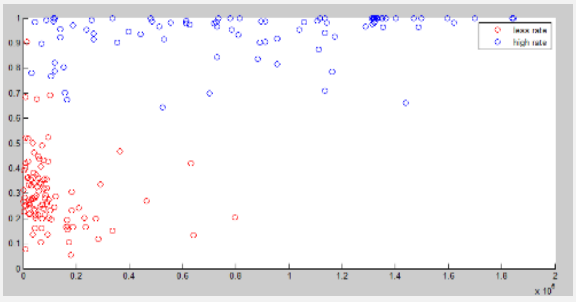
مجموعه داده‌های مورد اســتفاده در این پژوهش از (ســورات و لواندوســکی ۲۰۱۴، لواندوسکی و سوارت ۲۰۱۲) می‌باشد. برای پیاده‌سازی روش پیشنهادی از ابزارهای متلب و R اسـتفاده شـده اسـت. این مجموعه داده، به صـورت یک ماتریس با ابعاد سـه سـتون و ۲۷۰۹۸۴۷۲ سطر به عنوان یال‌ها و وزن بین یال‌های گراف در محیط R وارد شده است. در محیط R و با اسـتفاده از کتابخانه Igraph گراف ارتباط اجتماعی کاربران با وزن‌های مورد استفاده ساخته شده است. تعداد امتیاز کاربران به مکان‌ها به تعداد ۲۸۰۹۵۸۱ امتیاز است که مقدار امتیاز عدد طبیعی از ۱ تا ۵ است.

بررسی ارتباط بین متغیرهای موردمطالعه

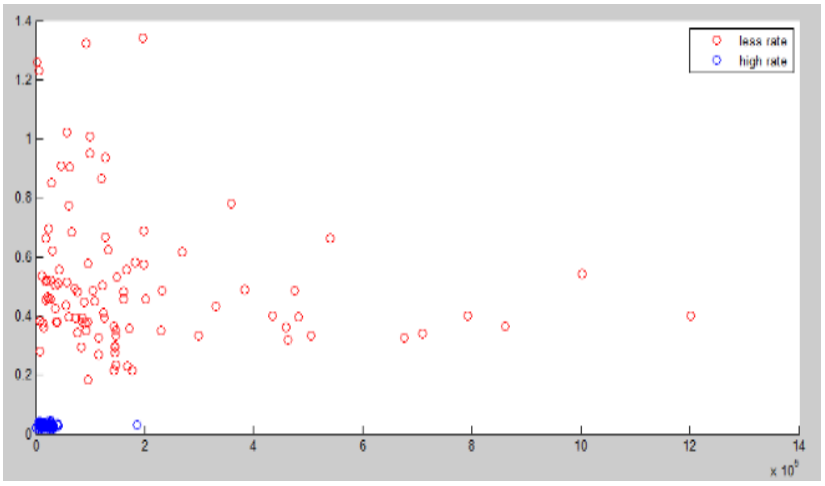
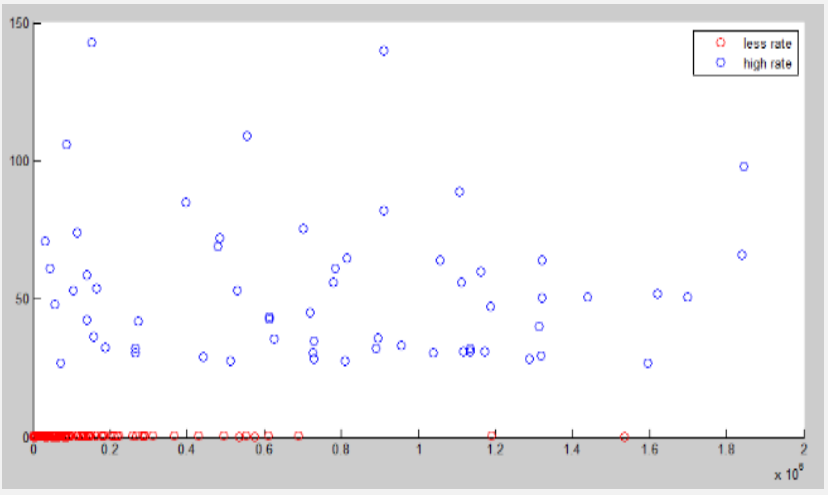
همان‌طور که بیان شـــد برای آزمون‌های مختلف، ۱۰۰ کاربر با امتیاز پایین و ۱۰۰ کاربر با امتیاز بالا انتخاب شد. در بررسی رابطه متغیر اول با دوم، نسبت درجه رأس گراف ego این کاربران به تعداد یال‌های گراف ego در جدول ۵ نشان داده شـده اسـت. همچنین

نسبت درجه رأس گراف ego این کاربران به جمع وزن تعداد یال‌ها در گراف ego در جدول ۶ نشان داده شده است. کاربرانی که امتیاز آن‌ها در محدوده نزدیک یا دور از معادله خط باشد به عنوان کاربران مشکوک در نظر گرفته می‌شوند. برای این منظور ۱۰۰ کاربر با امتیاز بالا و ۱۰۰ کاربر با امتیاز پایین بررسی شده است ۱۰۰ کاربر با امتیاز پایین با رنگ قرمز و ۱۰۰ کاربر با امتیاز بالا بارنگ آبی نشان داده شده است. مقادیر محور عمودی در این شکل نشان‌دهنده تقسیم مقدار درجه رأس به تعداد یال‌ها در گراف ego این ۲۰۰ کاربر می‌باشد. مقادیر نشان می‌دهد که کاربرانی که امتیاز بالا داشتند در واقع کاربرانی بودند که به معادله خط نزدیک بودند. مقادیر تقسیم درجه به تعداد یال نزدیک به ۰/۵ است اما کاربرانی که امتیاز بالایی داشتند در نمودار منحنی برازش با رنگ سبز نشان داده شده بودند، از خط قرمزرنگ منحنی برازش دور بودند. در صورتی‌که به عنوان مثال درجه یک رأس ۲۰۰ باشد و تعداد یال‌های گراف ego این کاربر نیز ۲۰۰ باشد حاصل تقسیم یک می‌شود که درواقع گراف ego این کاربر به شکل ستاره است و نشان‌دهنده رفتار ناهنجار این کاربر است. در صورتی‌که تعداد درجه رأس این کاربر ۱۰۰ و تعداد یال‌ها در گراف ego این کاربر نیز ۲۰۰ باشد مقدار تقسیم نیم می‌شود که نشان‌دهنده رفتار هنجار و متعادل کاربر است.

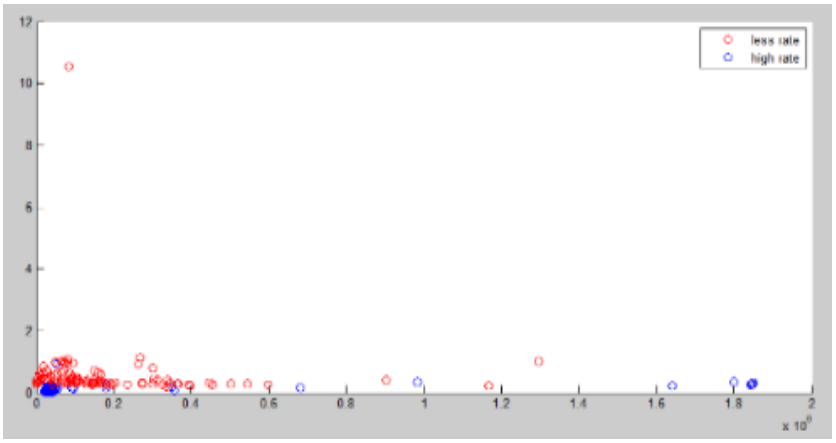
جدول ۵. بررسی متغیرهای درجه و رأس برای ۱۰۰ کاربر با امتیاز بالا

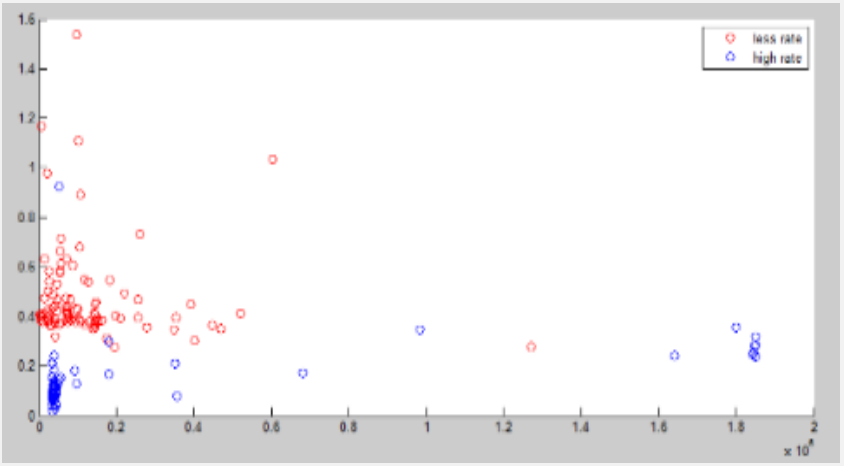
نسبت درجه به تعداد یال‌ها در گراف **ego** برای هر کاربر برای ۱۰۰ کاربر با بالاترین و پایین‌ترین درجه





جدول ٦. بررسی متغیرهای درجه و رأس برای ۱۰۰ کاربر با امتیاز بالا

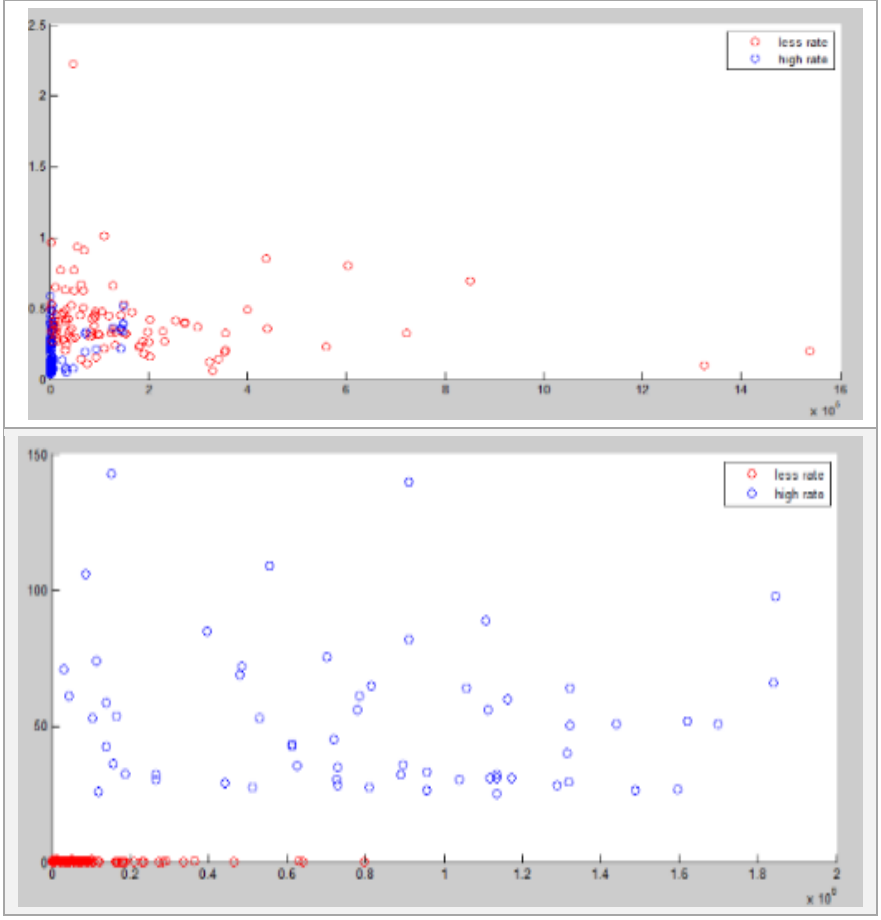
نسبت درجه به جمع وزن تعداد یال‌ها در گراف **ego** برای هر کاربر برای ۱۰۰ کاربر با بالاترین و پایین‌ترین درجه



نسبت درجه به جمع وزن تعداد یال‌ها در گراف **ego** برای هر کاربر برای ۱۰۰ کاربر با بالاترین و پایین‌ترین درجه





نسبت درجه به جمع وزن تعداد یال‌ها در گراف ego برای هر کاربر برای ۱۰۰ کاربر با بالاترین و پایین‌ترین درجه



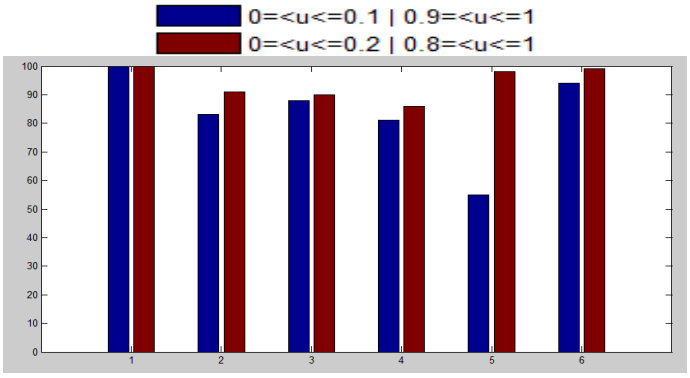
با توجه به این که ۶ ارتباط بین دو متغیر به دست آمده، نتایج حاصل شده از این ارتباط‌ها مشخص می‌کند که این ارتباط‌ها شامل موارد مندرج در جدول ۶ است.

در جدول ۶ شماره ارتباط و این که آن ارتباط بین کدام دو متغیر بوده است نشان داده شده است. در این جدول به عنوان مثال شماره ارتباط ۱، ارتباط بین دو متغیر درجه رأس و تعداد یال می‌باشد.

جدول ٦ شماره ارتباط و مختصات x و y نقاط هر ارتباط

شماره ارتباط	محور x	محور y
۱	درجه رأس	تعداد یال
۲	درجه رأس	جمع وزن یال
۳	مرکزیت بینابینی	تعداد یال
٤	مرکزیت بینابینی	جمع وزن یال
٥	مرکزیت بردار ویژه	تعداد یال
٦	مرکزیت بردار ویژه	جمع وزن یال

با توجه به مقادیر مشخص‌شده در شکل‌های جدول ۵ برای مقادیر رنگ آبی در این شکل‌ها تعداد افرادی که مقادیر محور عمودی آن‌ها در محدوده‌های خاص هست شمارش می‌شود. در شکل ۷ نتایج این شمارش نشان داده شده است در این شکل محور افقی نشاندهنده شماره ارتباط بین متغیرها و محور عمودی درصد افراد در بازه‌های تعیین شده در نمودار را نشان می‌دهد. شماره ارتباط بین متغیرها در جدول ۶ نشان داده شده است و مقدار u در راهنمای شکل ۷ تعداد نقاط آبی است که مقادیر محور عمودی این نقاط در محدوده مشخص شده را نشان می‌دهد. در شکل ۷ دو محدوده در نظر گرفته شده است و به عبارتی افراد ناهنجار را می‌دهد. در هر ارتباط اگر نسبت درجه رأس به تعداد یال در گراف ego این کاربران نزدیک صفر و نزدیک ۱ باشد و مجموعه مقادیر در بازه [۹ و ۰ و۱] و [۸ و ۰ و۱] به‌عنوان مجموعه مقادیر نزدیک به یک و مجموعه مقادیر در بازه [۰ و ۱ و۰] و [۰ و ۲ و ۰]به عنوان مجموعه مقادیر نزدیک به صفر در نظر گرفته شده است و به عنوان کاربر ناهنجار واقعی تشخیص داده می‌شود.



شکل ۷. تعداد افراد با نسبت درجه به تعداد یال در محدوده مشخص شده

نقاط مشخص‌شده در شکل ۷ نشان دهنده این است که ناهنجاری در ارتباط سوم و ششم بهتر از ارتباط دوم بوده است و همچنین وقتی محدوده در نظر گرفته شده بزرگ‌تر شده ناهنجاری در ارتباط سوم و پنجم و ششم بهتر از ارتباط دوم بوده است. به این معنا که در ارتباط دوم ناهنجاری در سطح همسایگی محل سکونت کاربران بوده است ولی در ارتباط سوم و چهارم و پنجم و ششم علاوه بر همسایگی، در سطح علاقه نیز ناهنجاری دیده شده است. نتایج نشان دهنده این است که شناسایی ناهنجاری در ۵ ارتباط مهم‌تر به ترتیب ۸۳، ۸۸، ۸۱،۵۵ و ۹۴ درصد برای محدوده اول و ۹۱، ۹۰، ۸۶،۹۸ و ۹۹ درصد برای محدوده دوم است.

نتیجه‌گیری و پیشنهادها

شبکه‌های اجتماعی در معرض حملات جعلی و آسیب‌پذیر هستند. فعالیت‌های ناخوشایند به دلیل استفاده گسترده حاصل از افزایش محبوبیت آن‌ها رو به افزایش است؛ بنابراین، تشخیص فعالیت‌های غیرعادی، به ویژه در شبکه‌های اجتماعی، اساساً مورد نیاز است زیرا به ما کمک می‌کند تا رفتار کاربران نامتعارف را شناسایی کرده و اطلاعات مهم و قابل توجه در مورد رفتار آن‌ها به دست آوریم. به منظور تشخیص فعالیت‌های غیرمعمول در شبکه‌های اجتماعی، یکی از روش‌های متداول، مشاهده الگوهای تعامل بین کاربران مختلف است. شبکه‌های اجتماعی به خوبی به شکل گراف قابل نمایش هستند. چنین الگوهای تعاملی می‌توانند به طور مؤثر با تحلیل

خواص ساختاری گراف‌ها مورد بررسی قرار گیرند. در این مقاله روش پیشنهادی روی مجموعه داده شبکه اجتماعی موقعیت‌محور foursquare مورد بررسی قرار گرفت و ۵ متغیر در گراف ego برای هر کاربر تعریف شد. سپس ارتباط بین ۶ مورد از دو به دوی این متغیرها در دستگاه مختصات دو بعدی با به دست آوردن معادله خط مدلسازی شد. با مقایسه به دست آمده برای مقادیر واقعی و پیش‌بینی شده، معادلات خط در هر کدام از ۶ ارتباط تعریف شده کاربران با امتیازهای بالا و پایین مشخص شدند. کاربران با امتیازهای بالا از معادله خط دور و کاربران یا امتیاز پایین به معادله خط نزدیک‌تر بودند. سپس برای کاربران با امتیاز بالا نسبت درجه به تعداد یال در گراف ego این کاربران محاسبه شد و این نسبت جداگانه برای کاربران تشخیص داده شده و برای هرکدام از ۶ معادله خط جداگانه به دست آمده است و با شمارش این که چه تعداد از مقادیر این نسبت در محدوده مشخصی قرار دارد ناهنجاری کاربران در گراف دوستی و همسایگی و علایق کاربران بررسی شده است. ازجمله نوآوری این مقاله نسبت به می‌توان به موارد زیر اشاره نمود:

- استفاده از مجموعه داده شبکه اجتماعی موقعیت‌محور.
- وزن‌دهی به یال‌های گراف با استفاده از متغیرهای فاصله محل سکونت کاربران و امتیاز کاربران به مکان‌های مورد بازدید.
- محاسبه معیارهای مرکزیت وزن‌دار در گراف ego و استفاده از این معیارها برای تشخیص ناهنجاری.

متناسب با روش پیشنهادی این مقاله برای ادامه تحقیقات در این حوزه می‌توان به موارد زیر اشاره کرد:

• استفاده از روش‌های موازی برای اندازه‌گیری متغیرهای مورد استفاده در تحقیق ازجمله ارتباط بین یال‌های گراف ego کاربران.

• استفاده از روش پیشنهادی، در چند خط بالاتر نوآوری‌های روش پیشنهادی بیان شده است که برای تشخیص و شناسایی سوء رفتار کاربران به کار می‌رود، به این طریق که کاربرانی که عمداً رأی کم یا زیاد به مکان‌های خاصی می‌دهند تشخیص داده می‌شوند.

منابع

- Akoglu, L. Chandy, R. Faloutsos, C. Opinion fraud detection in online reviews by network effects. ICWSM 2013; 13:2–11
- Akoglu, L. Faloutsos, C. Event detection in time series of mobile communication graphs. In: Army Science Conference. 2010. p. 77–79.
- Akoglu, L. McGlohon, M. Faloutsos, C. Oddball: Spotting anomalies in weighted graphs. In: Advances in Knowledge Discovery and Data Mining. Springer; 2010. p. 410–421.
- Chakrabarti, D. Zhan, Y. Faloutsos, C. R-MAT: A Recursive Model for Graph Mining. In: SDM. SIAM; volume 4; 2004. p. 442–446.
- Gupta, M. Gao, J. Aggarwal, C. Han, J. Outlier detection for temporal data. Synthesis Lectures on Data Mining and Knowledge Discovery 2014a;5(1):1–129
- Hassanzadeh, R. & Nayak, R. (2013). A semi-supervised graph-based algorithm for detecting outliers in online-social-networks. In Proceedings of the 28th Annual ACM Symposium on Applied Computing (pp. 577-582). ACM.
- Hassanzadeh, R., Nayak, R. A rule-based hybrid method for anomaly detection in online-social-network graphs. In: Tools with Artificial Intelligence (ICTAI), 2013 IEEE 25th International Conference on. IEEE; 2013a. p. 351–357.
- Henderson, K., Gallagher, B., Li, L., Akoglu, L., Eliassi-Rad, T., Tong, H., Faloutsos, C. It's who you know: graph mining using recursive structural features. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM; 2011. p. 663–671.
- Kaur, R., & Singh, S. (2015, December). Detecting anomalies in online social networks using graph metrics. In 2015 annual IEEE India conference (INDICON) (pp. 1-6). IEEE.
- Levandoski, J. J., Sarwat, M., Eldawy, A., & Mokbel, M. F. (2012, April). Lars: A location-aware recommender system. In Data Engineering (ICDE), 2012 IEEE 28th International Conference on (pp. 450-461). IEEE.
- Nadi,n,(2016) anomaly detection in social networks using multi-dimensional array decomposition(in Persian).
- Sarwat, M., Levandoski, J. J., Eldawy, A., & Mokbel, M. F. (2014). LARS*: An efficient and scalable location-aware recommender system. IEEE Transactions on Knowledge and Data Engineering, 26(6), 1384-1399.
- Scott, J. Social network analysis: developments, advances, and prospects. Social network analysis and mining 2011;1(1):21–26.
- Tong, H., Lin, C.Y. Non-Negative Residual Matrix Factorization with Application to Graph Anomaly Detection. In: SDM. SIAM; 2011. p. 143–153.
- Yang, W., Shen, G.W., Wang, W., Gong, L.Y., Yu, M., Dong, G.Z. Anomaly detection in microblogging via co-clustering. Journal of Computer Science and Technology 2015;30; 30(5):1097–1108.

استناد به این مقاله: شناسه دیجیتال (DOI): 10.22091/jemsc.2019.1373

عدالتی، فاطمه؛ رموزی، مرتضی. (۱۳۹۷). «شناسایی افراد ناهنجار در شبکه‌های اجتماعی موقعیت‌محور با استفاده از معیارهای گراف». *مدیریت مهندسی و رایانش نرم*، ۸ (۱)، ۲۶۹-۲۴۵.