فصل سوم: رویکرد پیشنهادی

1-3 مقدمه

تشخیص جوامع در شبکه‌های اجتماعی به معنای شناسایی گروه‌هایی از کاربران با ویژگی‌ها و سلایق مشابه است که با هم در ارتباط هستند و به صورت مشترک محتوا و اطلاعات را مبادله می‌کنند. این جوامع می‌توانند براساس موضوعات مختلفی شکل گیری کنند، از جمله علایق، حرفه‌ها، محل زندگی، و یا هر معیار دیگری که مشابهت‌های مشخصی در میان افراد ایجاد کند. تشخیص و تحلیل این جوامع درک بهتری از ساختار و نقش کاربران در شبکه‌های اجتماعی را به ما می‌دهد.

شبکه‌های اجتماعی به عنوان یکی از جوان‌ترین و موثرترین ابزارهای ارتباطی در دنیای امروز شناخته می‌شوند. با رشد روز‌افزون کاربران در این شبکه‌ها، جوامع و گروه‌های مختلفی درون این پلتفرم‌ها شکل گرفته‌اند. تشخیص جوامع در این شبکه‌ها از اهمیت ویژه‌ای برخوردار است. به‌عنوان مثال، توانایی تشخیص این جوامع می‌تواند به کاربران و تاجران امکان ارتقاء ارتباطات خود با افراد مشابه را بدهد و همچنین به مدیران شبکه‌های اجتماعی کمک کند تا بهترین خدمات را به کاربران خود ارائه دهند.

تشخیص جوامع در شبکه‌های اجتماعی از اهمیت بسیاری برخوردار است زیرا این فرآیند به ما امکان می‌دهد ساختار و رفتار کاربران را بهتر درک کنیم. اطلاعات به‌دست‌آمده از تحلیل جوامع می‌تواند در بسیاری از زمینه‌ها به کار گرفته شود، از جمله تبلیغات هدفمند، بهینه‌سازی سیستم‌های پیشنهادگر، کاهش جعل حساب‌ها و ایجاد امنیت بیشتر برای کاربران. همچنین، تشخیص جوامع به ما کمک می‌کند تا به عنوان محققان و دانشمندان، الگوها و روندهای جدید را در ارتباطات اجتماعی شناسایی کرده و تحلیل کنیم.

یکی از مثال‌های بارز تشخیص جوامع در شبکه‌های اجتماعی، تحلیل گروه‌های کاربری مشترک در توییتر است. در این شبکه، افراد با توجه به علاقه‌مندی‌ها، تمایلات سیاسی، یا موضوعات خاصی که به آن‌ها علاقه دارند، در گروه‌های مختلفی عضو می‌شوند. با استفاده از الگوریتم‌های تحلیل شباهت و ارتباطات کاربران، می‌توان این گروه‌ها را شناسایی کرد و با درک بهتر ساختار و نقش‌های این گروه‌ها، محتوای مناسب‌تری را به کاربران ارائه داد. همچنین، این تحلیل به تاجران و تبلیغ‌دهندگان امکان می‌دهد تا با دقت بیشتری مخاطبان هدف خود را شناسایی کنند و به نتیجه‌ای مؤثرتر برسند.

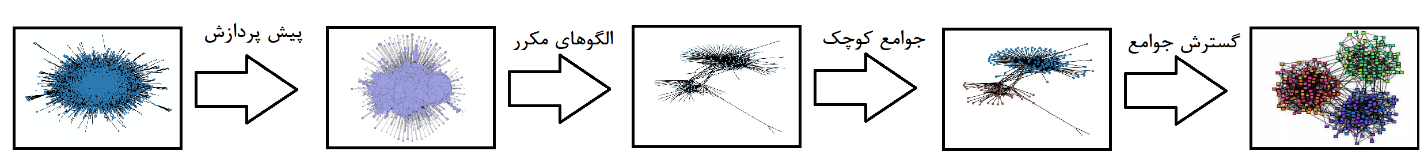
2-3 رویکرد پیشنهادی

رویکرد پیشنهادی ارائه شده برای کشف جوامع در شبکه‌های اجتماعی به ساختار گراف روابط کاربران محدود نمی‌شود، بلکه اقدام به در نظر گرفتن عملیات کاربران و استخراج الگوهای مکرر مرتبط با این عملیات می‌نماید. این رویکرد با تمرکز بر روی الگوهای مکرر در عملیات کاربران، به شناسایی رهبرانی پرداخته و پس از تایید الگوهای مکرر، جوامع کوچک را مشخص می‌کند.

در این رویکرد، ابتدا با تحلیل و استخراج الگوهای مکرر در عملیات کاربران، رهبران جوامع مشخص می‌شوند. این رهبران ممکن است به دلیل ارتباطات نزدیک و فعالیت‌های مشابه با یکدیگر، به عنوان مرکزهای جوامع کوچک شناخته شوند. پس از این مرحله، با انتساب پیروان به رهبران جوامع کوچک بر اساس شباهت، گسترش جوامع انجام ‌می­شود و افراد درون هر جامعه مشخص می‌گردند.

نهایتاً، با استفاده از الگوریتمی ابداعی بر اساس همسایگان هر شخص در گراف روابط و نیز با در نظر گرفتن عملیات کاربران، افراد باقی‌مانده به جوامع مربوطه نسبت داده می‌شوند. این رویکرد ترکیبی از معیارهای جوامع‌یابی مبتنی بر الگوهای مکرر و شباهت گسترش با اطلاعات ساختار گراف و عملیات کاربران، نتایج دقیق‌تری را در تشخیص جوامع در شبکه‌های اجتماعی ارائه می‌دهد. این رویکرد از تمهیدات قوی‌تری برای تشخیص جوامع برخوردار بوده و می‌تواند در انواع شبکه‌های اجتماعی با تعداد بزرگی از کاربران به کار گرفته شود. مراحل اصلی این رویکرد (شکل 1-3) عبارتند از:

1. پیش پردازش و آماده سازی داده­ها
2. استخراج الگوهای پر تکرار
3. تشکیل جوامع کوچک
4. گسترش جوامع



شکل 1-3: مراحل رویکرد پیشنهادی

در رویکرد ارائه شده ابتدا کاربران شبکه اجتماعی از نظر عملیات بررسی می­شوند و الگوهای مکرر بین فعالیت کاربران شناسایی می­شوند. در مرحله بعد این الگوهای مکرر برای استخراج جوامع همگن کوچک پردازش می­شوند. این جوامع کوچک در حکم رهبران شبکه­ی اجتماعی می­باشند. در اخر جوامع کوچک گسترش یافته و پیروان جوامع کوچک، با استفاده از شباهت کاربران شناسایی می­شوند. این مراحل به طور مختصر در شکل 3-1 نمایش داده شده­ اند. در بخش پیش رو جزئیات رویکرد پیشنهادی شرح داده خواهد شد.

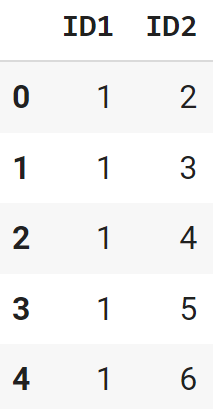
3-3 جزئیات و مراحل پیاده­سازی رویکرد پیشنهادی

همانطور که پیش از این نیز اشاره شد، رویکرد پیشنهادی شامل چهار مرحله اصلی می­باشد. در مرحله اول، در صورت نیاز روی داده­ها پیش­پردازش انجام می­شود و اطلاعات مورد نیاز استخراج شده و دو جدولِ عملیات و روابط کاربران استخراج می­شوند. در مرحله دوم الگوریتم کاوش الگوهای مکرر روی جدول عملیات اعمال شده و الگوهای مکرر استخراج می­شوند. هر الگوی استخراج شده در این مرحله دنباله­ای از کاربران است که یک الگوی پرتکرار در فعالیت کاربران را نشان می­دهد. این الگوها، نمایندگانی برای رهبران جوامع می­باشند. در مرحله سوم این الگوها همگن شده و الگوهای مکرر از نظر شباهت کاربران بررسی می­شوند و در صورت نیاز هرس شده یا به همدیگر می­پیوندند و جوامع کوچکی را تشکیل می­دهند. در نهایت در مرحله چهارم جوامع کوچک با پیوند کاربران به عنوان پیرو، گسترش می­یابند. در ادامه این مراحل به تفصیل شرح داده خواهند شد.

1-3-3 پیش­پردازش و آماده سازی داده­ها

به عنوان ورودی مرحله دوم، برای شناسایی الگوهای مکرر، نیاز به یک جدول خواهیم داشت که عملیات کاربران را نمایش دهد. همچنین ارتباطات بین کاربران توسط یک جدول دیگر که آن را جدول همسایگی می­نامیم مشخص می­شود. در مرحله­ی پیش پردازش هدف ایجاد این دو جدول و آماده سازی داده­ها برای استفاده در مراحل بعد می­باشد.

آنچه که با خزش از سطح توییتر به دست می­آید، معمولا داده­ی خامی است که حاوی دو دسته اطلاعات است؛ یک دسته اطلاعاتی است که از ماهیت روابط کاربران به دست می­آیند، نظیر دنبال کنندگان و دنبال شوندگانِ هر کاربر خاص که در تشکیل گراف جهت دار روابط استفاده می­شوند و دسته­ی دیگر اطلاعاتی است که فعالیت کاربران را نمایش می­دهد. این اطلاعات با بررسی و کاوش فعالیت کاربران حاصل می­شوند. به طور مثال فعالیت­هایی نظیر بازنشر کردن، پاسخ دادن و... از جمله اطلاعاتی است که در این مرحله استخراج می­شوند. بنابراین دو نوع جدول برای نمایش این داده­ها مورد استفاده قرار می­گیرد.

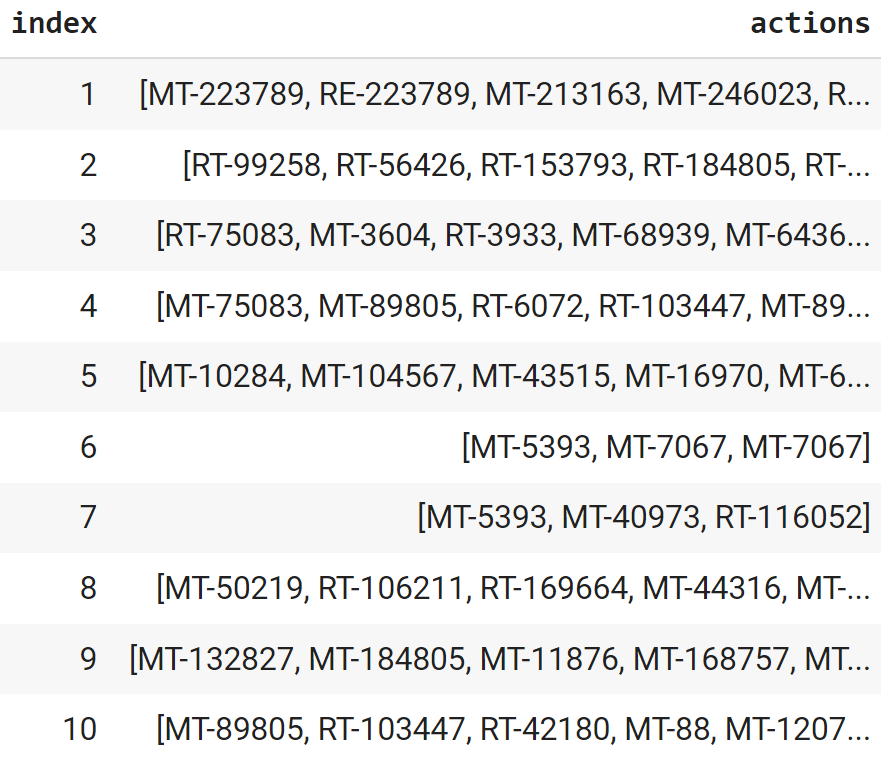


شکل 3-2: جدول روابط

* جدول روابط کاربران

جدول روابط کاربران به عنوان یکی از ورودی‌های مراحل آتی به طور مثال برای تشکیل گراف شبکه و گسترش جوامع بر اساس توپولوژی گراف، مورد استفاده قرار می­گیرد. این جدول اطلاعات مربوط به روابط میان کاربران را نمایش می‌دهد. این جدول حاوی دو ستون اصلی به نام‌های "کاربر مبدا" و "کاربر مقصد" می‌باشد. در این جدول، هر ردیف نمایانگر یک رابطه بین دو کاربر می‌باشد که از ماهیت ارتباطات اجتماعی بین کاربران استخراج شده‌اند. به عبارت دیگر، اگر کاربر A دنبال‌کننده کاربر B باشد، ردیفی در جدول روابط کاربران وجود خواهد داشت که مشخص می‌کند کاربر A و کاربر B به عنوان یک زوج مرتبط با هم هستند. همچنین، این جدول می‌تواند شامل اطلاعات اضافی مانند تاریخ ارتباطات و نوع ارتباط (مثلاً دنبال کردن یا دوستی) نیز باشد. این اطلاعات به ساختار گراف روابط کاربران کمک می‌کند. شکل 3-2 یک نمونه از جدول روابط کاربران را نمایش می­دهد. در این جدول کاربر مبدأ با “ID1” و کاربر مقصد با “ID2” نمایش داده شده است و ارتباط جهت دار دنبال کردن توسط کاربران شبکه اجتماعی توییتر در دیتاست مورد استفاده در این پژوهش را نمایش می­دهد.

* جدول فعالیت کاربران



شکل 3-3: جدول فعالیت کاربران

جدول فعالیت کاربران نیز به عنوان یکی دیگر از ورودی‌های مراحل آتی به خصوص مرحله دوم استفاده می‌شود تا الگوهای مکرر در عملیات کاربران استخراج و تشخیص داده شوند. هر ردیف از این جدول نمایانگر یک گروه از فعالیت­های کاربری می‌باشد که می‌تواند شامل بازنشر کردن، پاسخ دادن، نظر دادن، لایک کردن، ارسال پیام و سایر فعالیت‌ها در شبکه‌های اجتماعی باشد. در شکل 3-3 یک نمونه از جدول فعالیت کاربران نمایش داده شده است. در این جدول هر کاربر با یک شناسه چند رقمی و فعالیت انجام شده به طور اختصار با حروف انگلیسی مشخص شده است. اطلاعات موجود در این جدول با بررسی و کاوش فعالیت‌های کاربران از داده‌های خام توییتر (یا هر شبکه‌اجتماعی دیگر) استخراج می‌شود. تحلیل این اطلاعات به ما اجازه می‌دهد تا الگوهای مکرر و مهم در عملیات کاربران شبکه‌های اجتماعی را شناسایی و بررسی کنیم.

جدول فعالیت کاربران در این پژوهش به عنوان یکی از اجزای مهم و اساسی چارچوب ارائه شده ایفای نقش دارد. در این چارچوب، برای تحلیل شبکه‌های اجتماعی، نقش و اهمیت عملیات کاربران به‌عنوان عناصری اساسی در ساختار و عملکرد شبکه‌ها مورد بررسی قرار می‌گیرد. قابل ذکر است که عملیات کاربران باید بر اساس اختیار و علاقه­ کاربر انجام شده باشد. در واقع عملیاتی به کاربران منتسب می­شود که در رخ دادن آن مختار بوده اند.

در شبکه اجتماعی توییتر نیز عملیات مجاز بسیار زیادی وجود دارد که کاربران می‌توانند به انتخاب و سلیقه خود آن‌ها را انجام دهند. به عنوان مثال، در توییتر، کاربران می‌توانند توییت‌های مختلفی را ارسال کنند که در این شبکه به عنوان عملیات مجاز شناخته می‌شوند. همچنین، کاربران می‌توانند توییت‌های دیگران را لایک (پسندیدن) کنند یا روی آن‌ها کامنت بگذارند که نیز به‌عنوان عملیات‌های مجاز در این شبکه به‌شمار می‌آیند. در توییتر، ریتوییت نیز یکی از عملیات‌های مجاز است که به کاربران اجازه می‌دهد توییت‌های دیگران را به دنبال‌کنندگان خود نمایش دهند. این عملیات نشان‌دهنده تأیید و ارزش‌دانی کاربران از توییت‌ها و محتوای ارسالی دیگران است و به‌عنوان یکی از عملیات مهم در شبکه توییتر در جدول فعالیت کاربران قرار می‌گیرد. از دیدگاه دیگر، توییت‌هایی که با استفاده از هشتگ‌ها (#) مشخص می‌شوند نیز به عنوان عملیات مجاز در نظر گرفته می‌شوند. با ارسال توییت‌هایی با هشتگ‌های خاص، کاربران می‌توانند محتوا و توییت‌های خود را به یک موضوع یا تاپیک خاص مرتبط کنند و توسط سایر کاربران با همان هشتگ دنبال شوند.

با توجه به مثال‌های ذکر شده، جدول فعالیت کاربران در شبکه توییتر به‌عنوان یکی از اجزای مهم چارچوب ارائه شده در این پژوهش مورد توجه قرار گرفته است. این جدول شامل اطلاعات مربوط به فعالیت کاربران می‌شود و نقش حیاتی در تحلیل و شناخت جوامع و عملکرد شبکه توییتر دارد. بنابراین در تحلیل شبکه‌های اجتماعی، توجه به عملیات‌ مجاز و تاثیر آن‌ها بر روابط کاربران از اهمیت ویژه‌ای برخوردار است. انتخاب عملیاتی که کاربران در انجام آن‌ها مختار باشند، از جمله مهمترین ملاک‌ها در جدول فعالیت کاربران است. به همین دلیل، جدول فعالیت کاربران در این پژوهش به‌صورت مختلفی ارائه می‌شود که همگی نقش حیاتی در تحلیل و شناخت جوامع شبکه‌های اجتماعی ایفا می‌کنند.

با استفاده از دو جدول فوق، یک تحلیل جامع از روابط کاربران و فعالیت‌های آن‌ها در شبکه‌های اجتماعی امکان‌پذیر می‌شود. این تحلیل به ما اجازه می‌دهد تا الگوهای مکرر و شناخته شده در رفتار کاربران و روابط اجتماعی آن‌ها را در مراحل بعدی استفاده کنیم تا جوامع را به صورت دقیق‌تر و جامع‌تر شناسایی کنیم. این اطلاعات می‌توانند مبنایی قوی برای اجرای الگوریتم‌ها و معیارهای جوامع‌یابی در مرحله بعدی تحلیل و شناسایی جوامع در شبکه‌های اجتماعی ارائه دهند.

2-3-3 استخراج الگوهای پر تکرار

الگوهای پرتکرار یا الگوهای مکرر در شبکه‌های اجتماعی به مجموعه‌ای از عملیات یا رفتارهای مشابه توسط کاربران اشاره دارد که به صورت تکراری و مکرر انجام می‌شوند. این الگوها اغلب نشان‌دهنده ویژگی‌ها و رفتارهای مشابه در کاربران و نحوه تعاملات آن‌ها در شبکه‌های اجتماعی هستند. با شناسایی این الگوها می‌توان جوامع و زیرگروه‌های مشابه درون شبکه را شناسایی کرد و نقشه‌ی رفتاری و اجتماعی شبکه را بهتر درک کرد. در این پژوهش برای شناسایی الگوهای پرتکرار از جدول عملیات کاربران که در مرحله قبل آماده شد، استفاده می­شود. با اعمال الگوریتم کاوش الگوی مکرر بر روی جدول عملیات کاربران، دنباله­هایی از کاربران به دست می­آیند که بیانگر الگوهای پر تکرار در عملیات کاربران می­باشند. در این مرحله برای کاوش الگوهای پرتکرار از الگوریتم اپریوری[[1]](#footnote-1) استفاده شد. در الگوریتم اپریوری یک حد آستانه در نظر گرفته می­شود که به آن حد پشتیبان[[2]](#footnote-2) می­گویند. الگوهایی که میزان تکرارشان به حد کافی باشد که شرط حد پشتیبان حداقل را ارضا کنند، خروجی این الگوریتم می­باشند. به طور مثال دنباله­ی فرضی }1، 2، 3{ یک دنباله از کاربران است که در صورتی که یک الگوی پشتیبان باشد، یعنی میزان تکرار این دنباله به حد کافی برسد، به‌عنوان یک الگوی پرتکرار استخراج می‌شود. این الگوریتم می‌تواند دنباله‌هایی با طول مختلف را بر اساس حد پشتیبان مشخص شده استخراج کند.

با استخراج الگوهای پرتکرار از داده‌ها، می‌توانیم انواع الگوهای مختلف رفتاری کاربران را شناسایی کنیم. این اطلاعات به ما کمک می‌کند تا جوامع و زیرگروه‌های مشابه درون شبکه را شناسایی کرده و نقشه‌ی رفتاری و اجتماعی شبکه را بهتر درک کنیم. بنابراین خروجی این مرحله لیستی از الگوهای پرتکرار در رفتار کاربران است که با توجه به حد آستانه پشتیبان استخراج می­شوند. این الگوهای مکرر در تشخیص جوامع کوچک و رهبران تعیین کننده خواهند بود.

3-3-3 تشکیل جوامع کوچک

در این مرحله با توجه به الگوهای مکرر استخراج شده در مرحله پیشین، جوامع کوچک همگن به عنوان رهبران جوامع ایجاد می­شوند. برای تشکیل جوامع همگن در آغاز مجموعه کاربران هر الگوی مکرر از نظر شباهت بررسی می­شوند. برای ارزیابی شباهت از ویژگی­های توپولوژیک گراف روابط کاربران در کنار فعالیت کاربران استفاده می­شود. در ادامه نحوه ارزیابی شباهت دو کاربر به تفصیل توضیح داده خواهد شد. بنابراین الگوهای مکرر از نظر شباهت اعضا بررسی می­شوند و اعضایی که شباهت آن­ها با سایر اعضای دنباله­ی مکرری که عضوش هستند از یک حد آستانه کمتر باشد، هرس خواهند شد. حال نمایندگان همگنی از اعضای جوامع رهبر در اختیار خواهیم داشت. در مرحله بعد دنباله­های پرتکراری که شباهت اعضایشان از حدی بیشتر باشد به یکدیگر الحاق شده و جوامع همگنی از الگوهای پرتکرار ایجاد می­شود که اعضایشان پرتکرارترین فعالیت و بیشترین مشابهت را با یکدیگر را دارند. این جوامع همگن رهبران جوامع نامیده می­شوند.

1-3-3-3 شباهت کاربران

برای بررسی شباهت کاربران ارتباطات و فعالیت­های آن­ها بررسی می­شود. شباهت کاربران در شبکه‌های اجتماعی می‌تواند با استفاده از معیارهای مختلفی محاسبه شود. این معیارها بر اساس ویژگی‌ها و رفتارهای کاربران مختلف تعریف می‌شوند. در ادامه، به توضیح و معرفی معیارهای شباهت کاربران خواهیم پرداخت.

1. شباهت بر اساس دوستی: در شبکه­های اجتماعی نظیر توییتر، ارتباط دو نفر ممکن است جهت دار یا بی جهت باشد. در صورتی که این ارتباط جهت دار باشد، سه حالت ممکن است رخ دهد. نخست اینکه هر دو کاربر همدیگر را دنبال کرده باشند. در این صورت امتیاز شباهت 1 در نظر گرفته می­شود. حالت دیگر این است که تنها یکی از آن­ها دیگری را دنبال کرده باشد که در این صورت امتیاز شباهتشان 0.5 خواهد بود و در صورتی که هیچ یک از دو کاربر یکدیگر را دنبال نکنند، امتیاز شباهتشان از نظر این معیار 0 می­باشد.
2. شباهت بر اساس دنبال کنندگان[[3]](#footnote-3) مشترک: این معیار بر اساس ارتباط دنبال کنندگان کاربران تعریف می‌شود. اگر دو کاربر اشتراک بسیاری از کاربران را که دنبال‌کننده‌هایشان هستند داشته باشند، این نشان‌دهنده‌ی شباهت زیادی بین آن دو کاربر است. این معیار می‌تواند با محاسبه تعداد فالوورهای مشترک دو کاربر و نرمال‌سازی آن‌ها بر اساس مجموعه فالوورهای هر کاربر محاسبه شود. نحوه­ی محاسبه­ی این شباهت در معادله­ی 3-1 نمایش داده شده است. در صورت کسر تعداد دنبال کنندگان مشترک دو کاربر و در مخرج این عبارت ماکزیمم تعداد دنبال کنندگان برای نرمال سازی عبارت تقسیم شده است.

*معادله­ی 3-1*

1. شباهت بر اساس دنبال شوندگان مشترک[[4]](#footnote-4): این معیار مشابه معیار پیشین است با این تفاوت که به جای شمارش دنبال­کنندگان، دنبال شوندگان محاسبه می­شوند.

*معادله­ی 3-2*

1. شباهت هشتگ­ها: هشتگ‌ها به عنوان برچسب‌های مشترکی که توسط کاربران برای محتواها و پست‌ها استفاده می‌شوند، به کار می‌روند. اگر دو کاربر هشتگ‌های مشابه را استفاده کنند، نشان‌دهنده شباهت در علاقه‌ها و موضوعات مشابه است. این معیار می‌تواند با محاسبه تعداد هشتگ‌های مشترک دو کاربر و نرمال‌سازی آن‌ها بر اساس مجموعه هشتگ‌های هر کاربر محاسبه شود.

*معادله­ی 3-3*

1. شباهت منشن­[[5]](#footnote-5)ها: در شبکه­های اجتماعی کلمه منشن به معنای اشاره، اعلام یا ذکر کردن یک شخص یا موضوع خاص است. در شبکه‌های اجتماعی مانند توییتر، وقوع منشن به معنای ارجاع به یک کاربر خاص در توییت یا پست دیگری است. با استفاده از نماد @ در ابتدای نام کاربری، می‌توان به صورت مستقیم به یک کاربر ارجاع کرد و او را در توییت ذکر کرد. شباهت منشن­ها میان دو کاربر از حاصل دو دیدگاه محاسبه می­شود. نخست اینکه منشن­های مشترک دو کاربر محاسبه شده و توسط تعداد کل منشن­های دوکاربر نرمال سازی می­شود. در واقع برای کاربر i مجموعه­ی کاربرانی که این کاربر منشن کرده با Mi نمایش داده می­شود. به همین ترتیب Mj مجموعه­ی کاربرانی است که کاربر j منشن کرده است. اشتراک این دو مجموعه نرمال شده توسط مجذور حاصلضرب اندازه­ی این دو مجموعه شباهت دو کاربر از نظر کاربرانی که منشن کرده اند را نشان می­دهد. از طرف دیگر تعداد دفعاتی که این دو کاربر همدیگر را منشن کرده اند نیز ملاک شباهتشان است. بدین معنی که تعداد دفعاتی که این دو کاربر همدیگر را منشن کرده اند محاسبه می­شوند و بر تعداد کل منشن­های این دو نفر نرمال سازی می­شود. از مجموع مقدار حاصل از این دو دیدگاه مشابهت دو کاربر بر اساس منشن محاسبه می­شود. معادله­ی 3-4 نحوه­ی محاسبه­ی این شباهت را نشان می­دهد.

*معادله­ی 3-4*

1. شباهت ریپلای­[[6]](#footnote-6)ها: ریپلای به معنای پاسخ‌دادن به یک توییت خاص است و به کاربران اجازه می‌دهد تا به توییت‌ها و پست‌های دیگری که در توییتر منتشر شده‌اند، پاسخ دهند. با استفاده از ریپلای، کاربران می‌توانند در جریان موضوعات و بحث‌های مختلف در شبکه‌های اجتماعی با هم ارتباط برقرار کنند و نظرات یا پاسخ‌های خود را به اشتراک بگذارند. به طور مشابه با شباهت منشن­ها، شباهت ریپلای­ها نیز از دو بخش تشکیل می­شود که شامل شباهت دو کاربر در ارتباط با سایر کاربران و شباهت دو کاربر در ارتباط با یکدیگر است. معادله­ی 3-5 نحوه­ی محاسبه­ی این شباهت را نمایش می­دهد.

*معادله­ی 3-5*

1. شباهت ریتوییت[[7]](#footnote-7)­ها: ریتوییت یک اصطلاح معمول در توییتر است و به اشتراک‌گذاری دوباره یک توییت از یک کاربر توسط کاربر دیگری اشاره دارد. با ریتوییت کردن توییت یک کاربر، توییت او به صورت عمومی برای دنیای دیگر کاربران قابل مشاهده می‌شود. به همان شکل که برای منشن و ریپلای توضیح داده شد، شباهت ریتوییت نیز محاسبه می­شود. نحوه­ی محاسبه این شباهت در معادله­ی 3-6 نمایش داده شده است.

*معادله­ی 3-6*

با استفاده از این معیارهای شباهت، می‌توان نقاط مشترک و تفاوت‌های کاربران در شبکه‌های اجتماعی را شناسایی کرد. اطلاعات حاصل از این معیارها می‌تواند در ارتقاء سیستم‌های توصیه‌گر و محتوای مشابه مناسب برای کاربران مورد استفاده قرار گیرد. همچنین، این اطلاعات کمک می‌کند تا جوامع مختلف درون شبکه را شناسایی کرده و با بهبود برنامه‌ها و الگوریتم‌ها در جهت بهبود تجربه کاربری شبکه‌های اجتماعی راهکار ارائه داد. در کل، استفاده از معیارهای شباهت کاربران به عنوان ابزاری قدرتمند در تحلیل و بهبود ساختار اجتماعی و رفتاری شبکه‌های اجتماعی می‌تواند نقش مهمی در بهبود کیفیت و عملکرد این شبکه‌ها ایفا کند.

2-3-3-3 همگن سازی

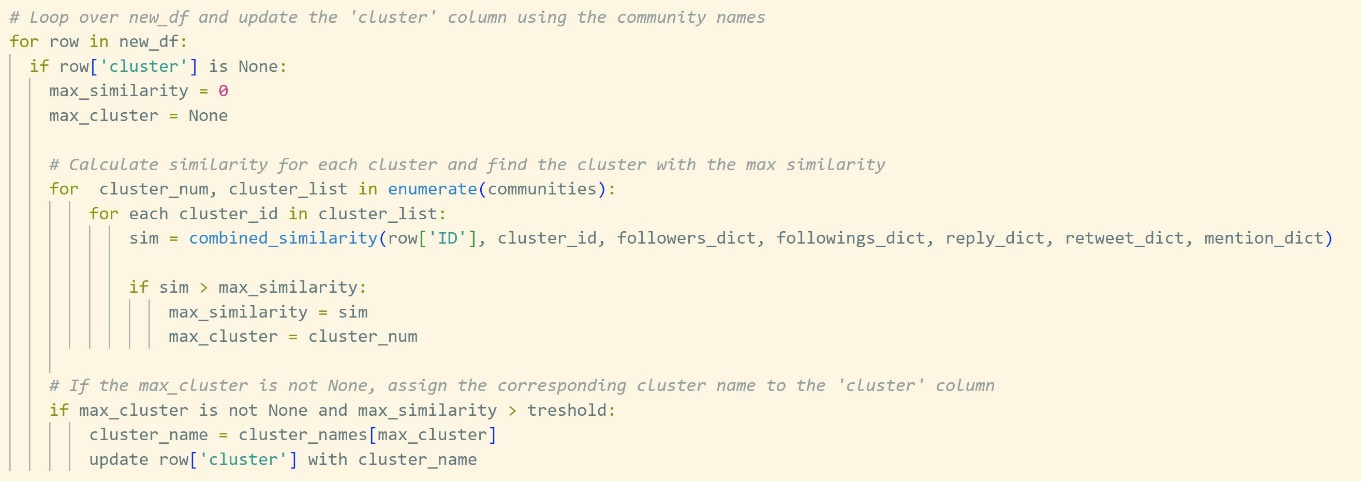
در این مرحله الگوهای مکرر استخراج شده از مرحله پیشین که دنباله­هایی از کاربران را نمایش می­دهند، همگن سازی می­شوند. منظور از همگن سازی این است که اعضای هر دنباله دو به دو از نظر میزان شباهت بررسی می­شوند. در صورتی که میزان شباهت اعضا از حد آستانه­ی در نظر گرفته شده برای همگن سازی کمتر باشد. عضوی که دارای کمترین شباهت است از دنباله حذف می­شود. هدف این مرحله این است که دنباله­های حاصل از استخراج الگوهای مکرر، دارای کاربرانی باشد که از نظر فعالیت و روابط درون شبکه اجتماعی نیز مشابه باشند. در شکل 3-4 شبه کد مربوط به مرحله­ی همگن سازی الگوهای مکرر نشان داده شده است.

3-3-3-3 ایجاد جوامع کوچک



شکل 3-4: شبه کد مرحله­ی همگن سازی

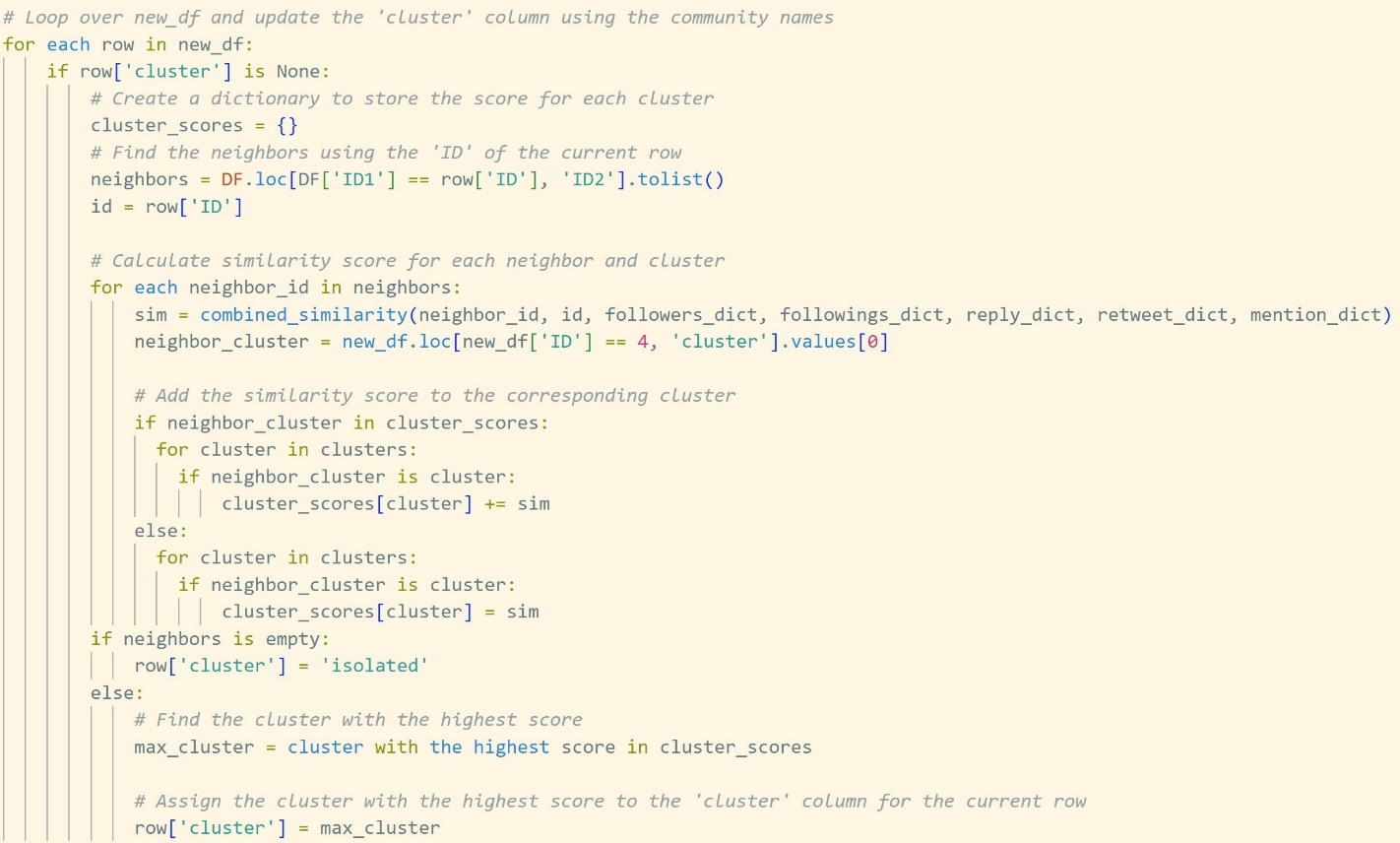
در نهایت پس از همگن سازی دنباله­های مکرر یافت شده، این دنباله­ها در صورتی که اعضایشان با یکدیگر مشابهت کافی داشته باشند، با یکدیگر ادغام شده و جوامع کوچکی را تشکیل می­دهند که اعضایشان الگوهای مکرر یافت شده از فعالیت کاربران هستند و همچنین با یکدیگر مشابهت کافی برای تشکیل جوامع رهبر کوچک را دارند. برای ادغام الگوهای مکرری که اعضای بسیار مشابه دارند می­توان از روش­های مختلفی استفاده کرد. رویکرد ساده استفاده از حلقه بر روی دنباله­های همگن و ادغام آن­ها در صورتی که شباهت اعضا از حد آستانه ای بیشتر شود است. یک رویکرد دیگر که به سادگی قابل اتخاذ است و نتیجه خوبی دارد این است که شباهت اعضا دو به دو محاسبه شود و یک ماتریس مشابهت ایجاد شود. هر درایه در این ماتریس نشان دهنده شباهت دو عنصر مرتبط می­باشد. حال می­توان مقادیر شباهت را از یک مقدار ثابت (حد بالا و ماکزیمم مقدار شباهت انتخاب خوبی است) کم کرد به طوری که هر درایه بیانگر فاصله­ی بین دو عنصر مرتبط باشد. در این صورت شبیه ترین عنصرها کمترین مقدار را خواهند داشت و کم شباهت ترین عناصر بیشترین مقدار را دارند. در این صورت با خوشه بندی به یک روش سلسله مراتبی می­توان الگوهای مکرری که شباهت زیادی به هم دارند را با یکدیگر ادغام کرد. نتیجه­ی این ادغام خروجی این بخش است که جوامع کوچکی هستند که به عنوان رهبر در شبکه اجتماعی در نظر گرفته می­شوند.



شکل 5-3: شبه کد گسترش جوامع بر اساس شباهت

4-3-3 گسترش جوامع

در این مرحله جوامع کوچک استخراج شده در مرحله پیشین، به عنوان هسته­های اصلی جوامع در نظر گرفته شده و کاربران کمی را پوشش می­دهند. سایر کاربران که هنوز عضو جامعه­ای نمی­باشند، قرار است در این مرحله به عنوان پیرو به جوامع رهبر متصل شوند. این عمل گسترش جوامع نامیده می­شود. گسترش جوامع در دو مرحله صورت می­گیرد. در مرحله نخست، افراد بر اساس شباهت به جوامع هسته­ی اصلی منتسب می­شوند. همانطور که شبه کد آن در شکل 3-5 نمایش داده شده است؛ روش کار در این مرحله به این صورت است که هر فرد به جامعه­ای که بیشترین شباهت را با آن دارد منتسب می­شود. با این عمل تعداد زیادی از کاربران به عنوان پیرو به جوامع مرکزی منتسب خواهند شد. می­­توان در این مرحله یک حد آستانه در نظر گرفت که برای انتساب یک عضو به یک جامعه­ میزان شباهت با هسته­ی آن جامعه، حداقل به اندازه حد بیان شده باشد. در مرحله دوم گره­هایی که تا کنون به جامعه­ای انتساب نشده اند بررسی می­شوند و به جامعه­ای منتسب می­شود که همسایگانی از آن گره که بیشترین شباهت را با آن دارند عضو آن جامعه هستند. شبه کد مربوط به این بخش در شکل 3-6 نمایش داده شده است. به طور مثال اگر کاربر A دارای سه همسایه­ی B و C و D باشد و کاربر B عضو جامعه­ی 1 و کاربر C عضو جامعه­ی 2 و کاربر D عضو جامعه­ی 1 باشد و شباهت کاربر A با B برابر یک باشد و شباهت کاربر A با کاربر C برابر سه باشد و شباهت کاربر A با کاربر D نیز یک باشد. برای هر جامعه­ای که در همسایگی کاربر A قرار دارد امتیازی محاسبه می­شود که مجموع حاصل جمع شباهت همسایگان عضو آن جامعه است. پس امتیاز جامعه­ی 1 برابر دو و امتیاز جامعه­ی 2 برابر سه خواهد بود. بنابراین کاربر A عضو جامعه­ی 2 تشخیص داده می­شود.



شکل 3-6:شبه کد گسترش جوامع بر اساس همسایگان

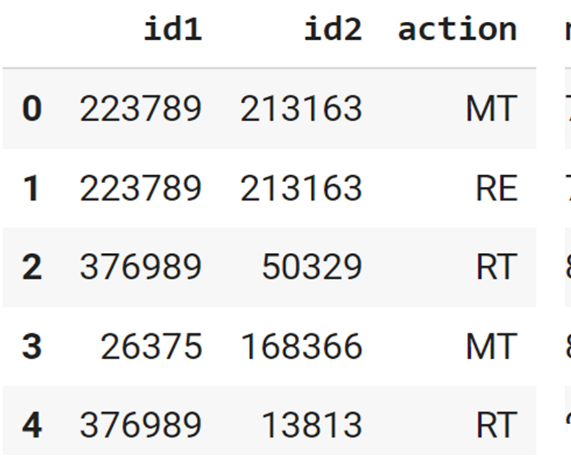
فصل چهارم: پیاده سازی

برای پیاده سازی رویکرد پیشنهادی در ابتدا زبان برنامه نویسی R در نظر گرفته شده بود ولی به منظور بهره برداری از سرویس­های گوگل کولب[[8]](#footnote-8) برای اجرا و همچنین سهولت برقراری ارتباط، دریافت و ذخیره سازی داده­ها، تصمیم به استفاده از محیط پایتون گوگل کولب شد. با توجه به شرایط و توضیحات مذکور، تصمیم گیری برای استفاده از زبان برنامه نویسی پایتون به جای R در پایان نامه دلایل مهمی را به همراه دارد. یکی از دلایل اصلی انتخاب پایتون این است که کدهای خود را بر روی سرورهای گوگل کولب اجرا کرده و این کدها به طور دائمی در گوگل درایو[[9]](#footnote-9) ذخیره می‌شوند، که این امکان دسترسی سریع و دائمی به کدها را فراهم می­کند. همچنین، محیط گوگل کولب با منابع قدرتمندی شامل گرافیک، پردازنده و حافظه پر سرعت، این امکان را ایجاد می‌کند تا به بهترین شکل از قابلیت‌های موجود برای پردازش و تحلیل داده‌ها استفاده کرد. به علاوه، امکان دسترسی مستقیم به دیتاست‌ها از سایت‌های معتبر مانند کگل[[10]](#footnote-10) (که تنها در محیط پایتون آن امکان پذیر است) نقطه قوت دیگری است که امکان ایجاد ارتباط آسانتر و سریعتر با منابع داده را فراهم می‌کند. در نتیجه، با توجه به محدودیت‌های حجم دیتاست و کندی عملکرد در محیط R در سیستم شخصی، انتخاب محیط پایتون گوگل کولب به عنوان ابزار اصلی برای تحقیقات و تحلیل در این پژوهش، تصمیم مناسبی به نظر می‌رسد.

1-4 پیش پردازش و آماده سازی داده­ها

همانطور که پیش از این مطرح شده بود، شبکه اجتماعی هدف در این پژوهش شبکه­ی اجتماعی توییتر می­باشد. در نخستین گام از پیاده سازی رویکرد پیشنهادی، نیاز به در اختیار داشتن داده­های لازم از روابط و فعالیت کاربران در شبکه اجتماعی توییتر می­باشد. ایده آل این بود که با استفاده از خزشگر و رابط کاربری توییتر اطلاعات مورد نیاز از بستر اینترنت خزش شوند. اما با توجه به تغییرات صورت گرفته در رابط کاربری توییتر پس از انتقال مالکیت آن به ایلان ماسک، ناگریز باید از دیتاست­های منتشر شده استفاده کرد. مناسب­ترین دیتاست یافت شده برای هدف این پژوهش، مجموعه داده­ی هیگز ]2[ می­باشد. این مجموعه­ی داده شامل اطلاعات حدود 500 هزار کاربر توییتر یعنی اشخاصی که دنبال می­کنند، دنبال شوندگانشان، ریتوییت­ها، ریپلای­ها و منشن­ها می­باشد. گراف روابط اجتماعی این شبکه 456626 گره و 14855842 یال دارد. همچنین گراف ریتوییت­های آن شامل 256491 گره و 328132 یال، گراف ریپلای­ها شامل 38918 گره و 32523 یال و گراف منشن­ها شامل 116408 گره و 150818 یال می­باشد. لازم به ذکر است که در گراف­های ذکر شده هر گره بیانگر یک کاربر منحصر به فردا در شبکه­ی توییتر می­باشد.

یکی از چالش­های روبرو شده در طی استفاده از این مجموعه داده حجم بزرگ آن می­باشد. به طوری که در مراحل مختلف پیاده­سازی رویکرد پیشنهادی (به طور مثال یافتن الگوهای مکرر) منجر به تکمیل ظرفیت حافظه­ی دستگاه می­شد و برنامه کرش می­کرد. بنابراین رویکردی برای کاهش ابعاد مجموعه داده و در عین حال حفظ کیفیت آن باید در نظر گرفته می­شد. مسلماً نمونه گیری تصادفی انتخاب خوبی برای کاهش ابعاد مجموعه داده نمی­باشد، در این راستا طبق توصیه­ی لسکوک و همکاران ]3[ روش آتش جنگل[[11]](#footnote-11) برای کاهش سایز گراف­های بزرگ مناسب می­باشد. مدل پیاده سازی شده برای کاهش سایز توسط روش آتش جنگل توسط لسوک و همکاران ]4[ ارائه شده است نهایتاً تصمیم بر این شد که به کمک این روش تعداد گره­های گراف به 10 درصد سایز اصلی کاهش پیدا کند و یال­های زیرگراف شامل گره­های باقی مانده نگهداری شوند. نتیجه زیر گرافی با 45756 گره و 2931409 یال بود. در ادامه­ی پژوهش این گراف و گره­های موجود در آن برای پیاده سازی رویکرد پیشنهاد در نظر گرفته می­شود.

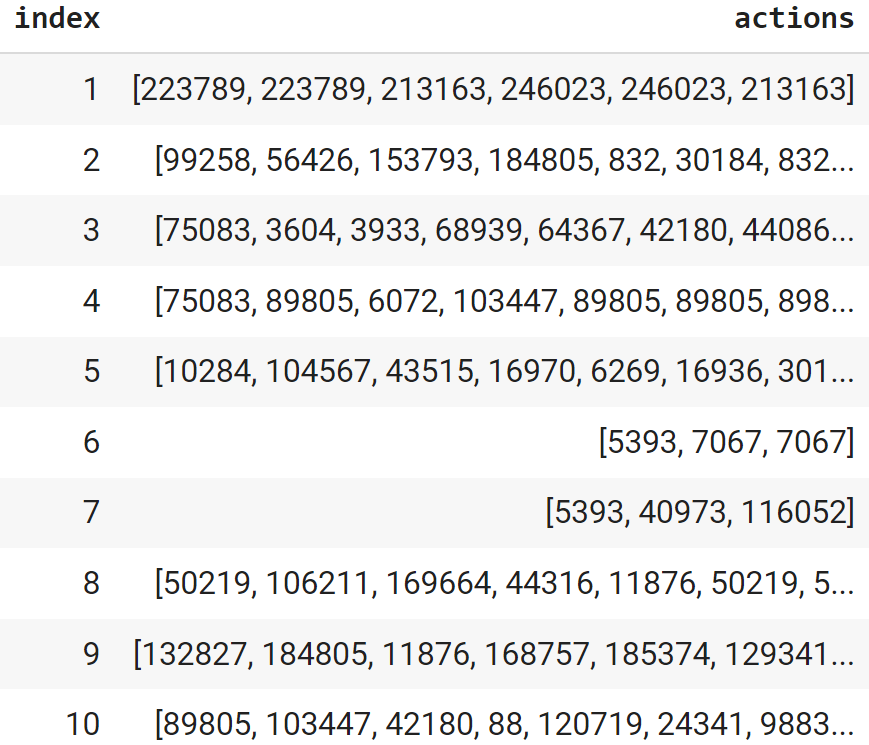


شکل 4-1: فعالیت کاربران در مجموعه داده مورد استفاده

حال پس از کاهش ابعاد مجموعه­ی داده و ایجاد مجموعه داده با حجم مناسب، نوبت به تشکیل جداول روابط کاربران و فعالیت کاربران می­رسد. در مجموعه داده­ی مورد استفاده جدول روابط کاربران به صورت شکل 3-2 موجود می­باشد و حاصل آن گراف جهت داری است که ارتباطات دنبال کردن در شبکه اجتماعی توییتر در مجموعه داده­ی مورد استفاده را نشان می­دهد. اما این مجموعه داده اطلاعات فعالیت کاربران را به صورت شکل 4-1 نمایش می­دهد. همانطور که دیده می­شود، در این جدول هر ردیف بیانگر یک نوع فعالیت است که با حروف اختصاری در ستون مربوطه مشخص شده است. به طور مثال ردیف نخست از این جدول مشخص می­کند که کاربر با شناسه­ی 223789 کاربر دیگری با شناسه­ی 213163 را منشن کرده است. برای مشخص کردن نوع عملیات در این جدول از حروف اختصاری MT، RE و RT به ترتیب برای منشن، ریپلای و ریتوییت استفاده شده است. متاسفانه در این مجموعه داده اطلاعاتی درباره­ی هشتگ­های مورد استفاده و متن توییت­ها وجود ندارد.

به منظور آماده سازی جدول فعالیت برای استخراج الگوهای مکرر، فعالیت خود مختار کاربران به عنوان عملیات در نظر گرفته می­شود. بنابراین منشن شدن و ریتوییت شدن و ریپلای شدن جزء فعالیت­های مد نظر نخواهند بود و تنها فعالیت­های منشن کردن، ریتوییت کردن و ریپلای کردن که با اختیار کاربر انجام می­شوند در نظر گرفته می­شوند. هر ردیف نیز بیانگر عملیات با اختیار انجام شده توسط کاربران دیگر بر روی یک کاربر خاص می­باشد. در نهایت جدول فعالیت کاربران با در نظر گرفتن نوع عملیات به صورت شکل 3-3 خواهد بود. اما برای سادگی بیشتر نوع عملیات را صرف نظر کرده و نهایتاً به صورت شکل 4-2 پیاده سازی می­شود.

2-4 استخراج الگوهای پر تکرار



شکل 2-4: جدول فعالیت­ کاربران

پس از تشکیل جدول فعالیت کاربران مشابه با شکل 4-2، الگوریتم اپریوری روی این جدول پیاده سازی شد. در این مرحله انتخاب حد پشتیبان[[12]](#footnote-12) مناسب، در الگوهای مکرر خروجی تاثیر مستقیم دارد. این حد پشتیبان ارتباط مستقیم با نوع شبکه و مجموعه داده­ی مورد استفاده دارد و باید توسط متخصص شبکه اجتماعی تعیین شود. در مدل پیاده سازی شده با توجه به مقیاس بزرگ شبکه، حد پشتیبان 0.003 انتخاب شد که منجر به استخراج 42 مجموعه­ی پرتکرار یک عضوی و دو عضوی شد. در صورتی که این حد پشتیبان 0.00094 در نظر گرفته می­شد، 3291 الگوی مکرر با 1 عضو تا 5 عضو در هر مجموعه به دست می­آمد. در شکل 4-3 برخی از این الگوهای مکرر نمایش داده شده اند.

شکل 3-4: برخی الگوهای مکرر به دست آمده با حد پشتیبان 0.003



3-4 تشکیل جوامع کوچک

در این مرحله­ معیارهای شباهت اشاره شده در فصل پیشین پیاده سازی شده و برای همگن­سازی دنباله­های کاربران حاصل از الگوهای مکرر استفاده می­شود. لازم به ذکر است با توجه به اینکه مجموعه داده­ی مورد استفاده، اطلاعاتی راجع به هشتگ مورد استفاده توسط کاربران در اختیار نمی­گذارد، در محاسبه شباهت کاربران این معیار لحاظ نشده است. همچنین شباهت یک کاربر با خودش مقدار 1 در نظر گرفته شده است. پس از اجرای الگوریتم همگن سازی، اگر دنباله­ای از کاربران هیچ یک شباهت کافی (بیش از حد شباهت در نظر گرفته شده که در اینجا عدد یک می­باشد) را نداشته باشند. کل دنباله حذف می­شود. اما اگر عضوی از یک دنباله شباهت آن با سایر اعضای دنباله کمتر از حد مشخصی باشد، به عنوان عضو نا­مرتبط تشخیص داده می­شود و از الگوی مکرر مد نظر حذف می­شود. بنابراین خروجی این بخش گروه­هایی از اعضای مشابه که الگوهای مکرر نیز هستند، می­باشد.

پس از همگن سازی الگوهای مکرر، نوبت به تشکیل جوامع کوچک یا رهبر می­رسد. برای تشکیل جوامع کوچک، الگوهای همگن استخراج شده در مرحله قبل بر اساس مشابهت اعضا، با هم ادغام می­شوند. در این بخش برای ادغام الگوهای حاوی اعضای مشابه از الگوریتم خوشه­بندی سلسله مراتبی تجمعی[[13]](#footnote-13) استفاده شد اما هر روش دیگر که ادغام الگوهای مکرر را بر اساس شباهتشان انجام دهد و حاصل آن جوامعی از الگوهای مکرر و به حد کافی مشابه باشد، مورد قبول است. در روش مورد استفاده فاصله­ی بین اعضا به صورت معادله­ی 4-1 در نظر گرفته شد. علت کسر شدن شباهت از عدد 6 این است که فاصله­ی کاربران مشابه کمتر و فاصله­ی کاربران نامشابه بیشتر به دست آید. و عدد 6 یک حد بالای در نظر گرفته شده برای شباهت است.

*معادله­ی 4-1*

در نهایت تعداد هفت جامعه از الگوهای مکرر همگن به دست آمد که اعضای آن­ها حداقل شباهت 1 را با یکدیگر خواهند داشت.

4-4 گسترش جوامع

برای گسترش جوامع همانطور که در فصل سوم توضیح داده شد، در دو مرحله جوامع گسترش خواهند یافت. نخست براساس شباهت افراد با جوامع رهبر استخراج شده در بخش 4-3، افراد به جامعه­ای که بیشترین شباهت را با آن دارند منتسب خواهند شد. برای این انتساب می­توان یک حد شباهت در نظر گرفت و افراد به جامعه­ای منتسب می­شوند که بیشترین شباهت را به آن داشته باشند و شباهتشان با آن جامع نیز از حد در نظر گرفته شده بیشتر باشد. در این پیاده سازی این بخش حد صفر در نظر گرفته شد بدین معنی که تنها افراد به گروهی منتسب می­شوند که بیشترین شباهت را با آن خواهند داشت. جدول 4-1 تعداد اعضای منتسب شده به هر جامعه را پس از این مرحله نشان می­دهد. همانطور که دیده می­شود تعداد 837 کاربر پس از این مرحله بدون جامعه باقی مانده اند.

جدول 4-1 جوامع و تعداد اعضایشان پس از مرحله­ی نخست از گسترش جوامع

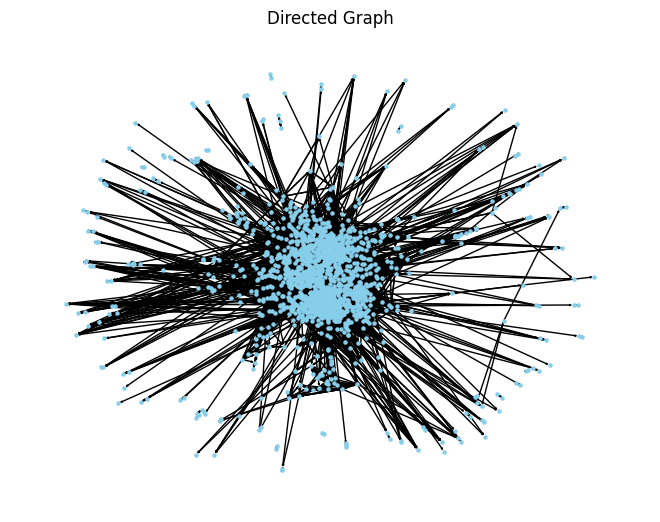
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| جامعه | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Nan |
| اعضا | 2453 | 17233 | 18834 | 2144 | 1426 | 862 | 1967 | 837 |

در مرحله بعد آن دسته از کاربران که همچنان جامعه­ای ندارند، بررسی خواهند شد و به جامعه­ای منتسب می­شوند که همسایگانی از آن­ها که بیشترین تشابه را با آن­ها دارند عضو آن جامعه باشند. پس اجرای این الگوریتم تنها کاربرانی که هیچ همسایه­ای ندارند و شباهتی نیز با هیچ یک از رهبران شناسایی شده ندارند، بدون جامعه باقی می­مانند که این کاربران در واقع داده­های پرت[[14]](#footnote-14) مجموعه داده مورد استفاده هستند و می­توان آن­ها را حذف کرد. در این پژوهش به جای حذف با برچسب ایزوله شده[[15]](#footnote-15) شناسایی شدند. در جدول 4-2 تعداد اعضای منتسب به هر جامعه در پایان گسترش جوامع مشخص شده است.

جدول 4-2 جوامع و تعداد اعضایشان پس از اتمام گسترش جوامع

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| جامعه | 1 | 2 | 3 | 4 | 5 | 6 | 7 | پرت |
| اعضا | 2498 | 17532 | 19285 | 2159 | 1429 | 864 | 1967 | 22 |

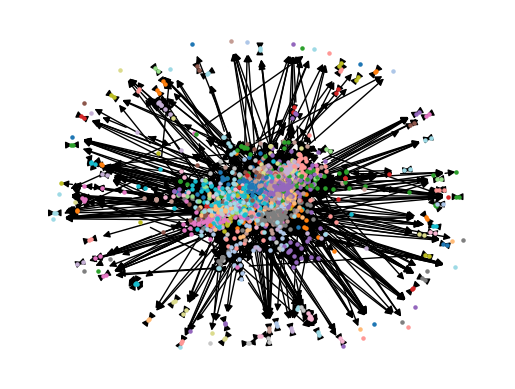
فصل پنجم: ارزیابی



شکل 5-1: نمایش گرافی مجموعه­ی داده مورد استفاده در ارزیابی

با توجه به این مسئله که رویکرد اعلام شده یک روش بدون نظارت است و در گام­های آن از برچسب­های از پیش مشخص شده برای یادگیری تشخیص جوامع استفاده نمی­شود و اساساً برچسبی بر روی مجموعه داده وجود نداشته که جوامع را مشخص کند، معیار مطلقی برای ارزیابی کیفیت جوامع تشخیص داده شده وجود ندارد؛ اما به منظور ارزیابی رویکرد پیشنهاد می­توان نتیجه­ی جوامع استخراج شده را با برخی دیگر از روش­های موجود مقایسه کرد. بدین منظور در این بخش تصمیم گرفته شد که در مرحله نخست برای هر کاربر در مجموعه­ی داده یک برچسب در نظر گرفته شود و سپس برچسب تولید شده توسط رویکرد پیشنهادی نسبت به برچسب در نظر گرفته شده سنجیده شود. به منظور تولید برچسب برای مجموعه­ی داده از روش لووین[[16]](#footnote-16) در نظر گرفته شد و برای مقابله با مرتبه زمانی بالای محاسبه­ی این الگوریتم، مجموعه­ی داده یک مرحله دیگر نیز کاهش داده شد و تنها 1864 گره منتخب از گراف ارتباطات باقی ماندند. سپس یک زیر گراف وزن دار جهت دار از گره­های متخب حاصل شد که وزن هر یال برابر با شباهت دو گره­ی دو طرف آن می­باشد. سپس الگوریتم لووین روی زیرگراف حاصل اعمال شد و جوامع تشخیص داده شده به عنوان برچسب­های کاربران در نظر گرفته شد. شکل 5-1 زیرگراف حاصل برای ارزیابی را نمایش می­دهد. شکل 5-2 نیز نتیجه­ی اعمال تشخیص جوامع به روش لووین را نشان می­دهد. لازم به ذکر است که علت اضافه کردن وزن به گراف روابط این است که فعالیت­های کاربران نیز که توسط معیار شباهت سنجیده می­شود، در روابط ادغام شود.

شکل 5-2: تشخیص جوامع با الگوریتم لووین



حال رویکرد پیشنهادی روی مجموعه­ی داده­ی برچسب دار (که برچسب آن توسط لووین ایجاد شده است) اعمال می­شود و جوامع توسط رویکرد پیشنهادی نیز تشخیص داده می­شوند. حال مطلوب است که از نظر معیارهای نظارت شده نظیر معیار اف و بازیابی و دقت نتایج به دست آمده بررسی شوند. با توجه به اینکه تعداد جوامع تشخیص داده شده در روش­های مختلف یکسان نیستند، یک چالش دیگر نیز در ارزیابی جوامع تشخیص داده شده نیز مطرح می­شود. برای مقابله با این چالش نیز در هر جامعه از برچسب، بررسی شد که چه میزان از کاربران در پیش بینی به کمک رویکرد پیشنهادی نیز عضو یک جامعه در نظر گرفته شد­ه­اند. بدین ترتیب نتایج جدول 5-1 برای ارزیابی تشخیص جوامع توسط رویکرد پیشنهادی نسبت به روش لووین در گراف روابط جهت دار و وزن دار (با وزن شباهت فعالیت کاربران) به دست آمدند.

جدول 5-1: ارزیابی رویکرد پیشنهادی

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| جامعه | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| دقت | 0.17 | 0.93 | 0.83 | 0.63 | 0.36 | 0.47 | 0.22 |
| بازیابی | 0.72 | 0.81 | 0.77 | 0.66 | 0.86 | 0.53 | 0.7 |
| معیار اف | 0.27 | 0.87 | 0.80 | 0.65 | 0.51 | 0.50 | 0.34 |

همچنین به منظور مقایسه نتایج به دست آمده با یک رویکرد دیگر، یکبار نیز تشخیص جوامع به روش حریصانه صورت گرفت و نتیجه­ی ارزیابی آن نسبت به روش لووین در جدول 5-2 ارائه شده است. شکل 5-3 نتیجه ارزیابی توسط معیار اف را برای رویکرد پیشنهادی و روش حریصانه نمایش می­دهد.

شکل 5-3: مقایسه نتیجه­ی ارزیابی توسط معیار اف بین دو روش حریصانه و رویکرد پیشنهادی

جدول 5-2: ارزیابی تشخیص جوامع با روش حریصانه

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| جامعه | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| دقت | 0.013 | 0.81 | 0.59 | 0 | 0 | 0 | 0.02 |
| بازیابی | 0.5 | 0.59 | 0.55 | 0 | 0 | 0 | 1 |
| معیار اف | 0.02 | 0.68 | 0.57 | 0 | 0 | 0 | 0.04 |

1. Apriori [↑](#footnote-ref-1)
2. Support [↑](#footnote-ref-2)
3. Followers [↑](#footnote-ref-3)
4. Followings [↑](#footnote-ref-4)
5. Mention [↑](#footnote-ref-5)
6. Reply [↑](#footnote-ref-6)
7. Retweet [↑](#footnote-ref-7)
8. Google colab [↑](#footnote-ref-8)
9. Google drive [↑](#footnote-ref-9)
10. Kaggle [↑](#footnote-ref-10)
11. Forest Fire (FF) [↑](#footnote-ref-11)
12. Support [↑](#footnote-ref-12)
13. Agglomerative hierarchical clustering [↑](#footnote-ref-13)
14. Outlier [↑](#footnote-ref-14)
15. Isolated [↑](#footnote-ref-15)
16. Louvain [↑](#footnote-ref-16)