

# توصیه کاربر در پیام‌رسان تلگرام با تحلیل گراف و مدل‌سازی ریاضی رفتار کاربران

داود کریم‌پور\* محمدعلی زارع چاهوکی\*\* علی هاشمی\*\*\*

\* دانشجوی کارشناسی ارشد مهندسی کامپیوتر، دانشکده مهندسی کامپیوتر، دانشگاه یزد

\*\* استادیار دانشکده مهندسی کامپیوتر، دانشگاه یزد

\*\*\* دانشجوی دکتری مهندسی کامپیوتر، دانشکده مهندسی کامپیوتر، دانشگاه یزد

تاریخ پذیرش: ۱۴۰۰/۰۵/۰۵

تاریخ دریافت: ۱۳۹۹/۱۱/۲۹

نوع مقاله: پژوهشی

## چکیده

سامانه‌های توصیه‌گر به‌منظور کاهش تولید و پردازش پرس‌وجو به‌وجود آمده‌اند. توصیه کاربران در شبکه‌های اجتماعی و پیام‌رسان‌ها برای کاربران عادی در یافتن دوست و برای بازاریابان جهت یافتن مشتریانی جدید، بسیار مفید است. در شبکه‌های اجتماعی مانند فیس‌بوک، یافتن کاربران هدف برای بازاریابی پیش‌بینی شده است؛ اما در پیام‌رسان‌هایی همچون تلگرام امکانی جهت یافتن جامعه هدف وجود ندارد. در این مقاله با استفاده از گراف و مدل‌سازی رفتار کاربران و همچنین تعریف ویژگی‌هایی مرتبط با گروه‌ها، روشی جهت توصیه کاربران تلگرام، ارائه شده است. روش پیشنهادی دربردارنده هشت گام است و هر یک از گام‌ها، می‌توانند روشی جهت توصیه کاربر در نظر گرفته شوند. مهاجرت، روشی جدید جهت مدل‌سازی علایق کاربران، براساس سوابق عضویت آنان در گروه‌ها است. داده‌های این پژوهش، مجموعه داده‌ای واقعی شامل بیش از ۹۰۰,۰۰۰ سوپرگروه و ۱۲۰ میلیون کاربر تلگرامی است. نتایج ارزیابی روش پیشنهادی بر روی ۱۰۰ گروه باکیفیت، حاکی از مؤثر بودن توصیه‌هایی برگرفته از سوابق کاربران و مدل‌سازی رفتار آنان نسبت به‌صرف استفاده از این اطلاعات است. رویکرد پیشنهادی با ارائه گام‌هایی در ادامه روش گروه‌های مشابه که جهت توصیه گروه در تلگرام ارائه شده بود، توانسته میانگین خطای  $RMSE$  را از ۰,۸۷ به ۰,۷۹ و میانگین خطای  $MAE$  را از ۰,۷۷ به ۰,۶۴ کاهش دهد. **واژگان کلیدی:** سامانه‌های توصیه‌گر، پیام‌رسان تلگرام، تحلیل گراف، رفتار کاربران.

## ۱. مقدمه

به‌عنوان مثال در شبکه اجتماعی فیس‌بوک<sup>۱</sup> جامعه ارتباطی کاربران گسترده و در پیام‌رسان‌هایی همچون تلگرام<sup>۲</sup> و واتساپ<sup>۳</sup> جامعه ارتباطی آنان محدود است. نرم‌افزارهای پیام‌رسان به‌طور رسمی با پیام‌رسان فوری<sup>۴</sup> شناخته می‌شوند. پیام‌رسان فوری به گفتگوی برخط دو یا چند کاربر به‌صورت متنی گفته می‌شود. متن‌های کاربران در این محیط‌ها با استفاده از اینترنت و یا سایر شبکه‌ها به‌صورت

امروزه شبکه‌های اجتماعی و پیام‌رسان‌ها علاوه بر استفاده عمومی کاربران، مورد توجه بسیاری از بازاریابان و شرکت‌ها قرار گرفته‌اند. تفاوت شبکه اجتماعی و پیام‌رسان را می‌توان در نحوه و میزان ارائه خدمات آنان بیان نمود. کاربران در پیام‌رسان‌ها با تعداد محدودی از مخاطبین در ارتباط هستند؛ اما در شبکه‌های اجتماعی، مخاطب کاربران عام است و می‌توانند با بسیاری از کاربران در ارتباط باشند.

<sup>۱</sup> Facebook

<sup>۲</sup> Telegram

<sup>۳</sup> WhatsApp

<sup>۴</sup> Instant messenger (IM)

نویسنده مسئول: محمدعلی زارع چاهوکی chahooki@yazd.ac.ir

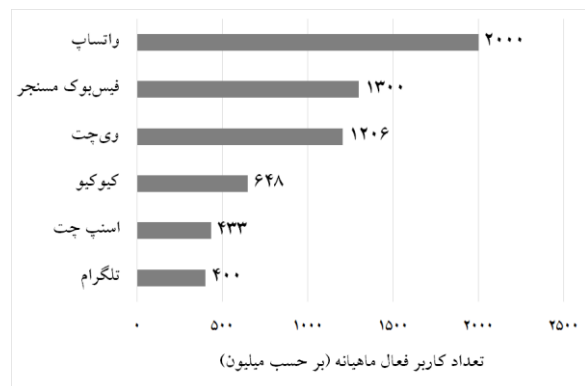
ایده‌کاوه<sup>۵</sup> به‌عنوان یک برنامه‌شخص‌ثالثی ایجاد شده است. در این سامانه، با جستجوی کلمات موردنظر، می‌توان گروه‌ها و کانال‌های بسیاری را در تلگرام یافت. معماری سامانه ایده‌کاوه در بخش ۱،۴ مورد بررسی قرار گرفته است.

ویژگی بسیار پرکاربرد در اکثر شبکه‌های اجتماعی، یافتن کاربران است. افزودن این ویژگی در تلگرام بسیار ارزشمند برای کاربران جهت یافتن دوست و برای بازاریابان جهت یافتن مشتریانی، مشابه با مشتریان قدیمی‌شان است. بازاریابی در شبکه‌های اجتماعی و پیام‌رسان‌ها به شیوه‌های متفاوتی انجام می‌شود. میلانی و زوزی [۴] تأثیر استفاده از انواع روش‌های بازاریابی در شبکه‌های اجتماعی را مورد بررسی قرار داده‌اند. یافتن مجموعه کاربرانی هدفمند جهت ارسال تبلیغ، از گذشته موردتوجه بازاریابان بوده است. چراکه در صورت ارسال تبلیغ به افرادی که هیچ علاقه‌ای به آن ندارند، موجب نارضایتی آنان می‌شود. این نارضایتی در پیام‌رسان‌ها موجب می‌شود تا کاربران پس از دریافت پیامی نامرتب، کاربر ارسال‌کننده را متخلف و پیام او را به‌عنوان هرزنامه گزارش کنند. فتحیان و حضرتقلی‌زاده [۵] روشی جهت دسته‌بندی پیام‌های تبلیغاتی، متناسب با پروفایل کاربران ارائه دادند. بهترین روش جهت مدل‌سازی و یافتن کاربران، استفاده از سامانه‌های توصیه‌گر است. سامانه‌های توصیه‌گر در سایت‌های تجاری، جهت توصیه اقلام و در شبکه‌های اجتماعی جهت توصیه دوست، کارایی دارند. مهم‌ترین هدف در این سامانه‌ها، جذب مشتریان و جلب اعتماد آنان به‌وسیله ارائه بهترین توصیه‌ها، مطابق با علایقشان است [۶].

در سامانه‌های توصیه‌گر، مرتبط بودن کاربران پیشنهادی با علاقه کاربر هدف بسیار با اهمیت است؛ چراکه در صورت توصیه کاربرانی نامرتب به یک کاربر، باعث نارضایتی او و نادیده گرفتن دیگر موارد توصیه‌شده می‌شود. هرچقدر عدم قطعیت در توصیه‌ها کاهش یابد توصیه‌گر دارای خطای کمتری در پیش‌بینی است [۷]. در این پژوهش با افزودن گام‌هایی در ادامه روش مقاله [۸] و با در نظر گرفتن دو موضوع مهم مدل‌سازی رفتار کاربران و کیفیت توصیه‌ها، خطای توصیه‌های ارائه شده تا حد امکان کاهش یافته است. علاوه بر دقت بالا و خطای کم پیش‌بینی، کیفیت توصیه‌ها در سامانه‌های توصیه‌گر نیز بسیار مهم هستند. اینکه توصیه‌گر قادر باشد کیفیت کاربران توصیه شده را تضمین نماید از مسائل بسیار مهمی است که برای بازاریابان جهت یافتن مشتریانی با کیفیت اهمیت دارد. در این پژوهش، برای اولین بار کیفیت کاربران بر اساس مقادیر عددی استخراج‌شده از گروه‌هایشان تجزیه و تحلیل شده است. روش پیشنهادی با ارائه سه پارامتر متفاوت توانسته است کیفیت کاربران پیشنهادی توسط توصیه‌گر را تضمین نماید.

بلادرنگ منتقل می‌شود. پیام‌رسان‌های فوری سرویس‌های متنوعی از جمله ارسال متن در زمان واقعی<sup>۱</sup>، تماس صوتی و تصویری، چت مخفی<sup>۲</sup> و ارسال فایل را در اختیار کاربران قرار می‌دهند.

تلگرام یک برنامه پیام‌رسان فوری مبتنی بر رایانش ابری است که بیش از ۴۰۰ میلیون کاربر فعال ماهیانه دارد [۱]. این پیام‌رسان یکی از محبوب‌ترین برنامه‌های پیام‌رسانی در ایران است. چراکه که بیش از ۵۰ میلیون کاربر تلگرام، ایرانی هستند [۲]. با توجه به شکل ۱، تلگرام در اکتبر ۲۰۲۰ در رتبه ششم محبوب‌ترین پیام‌رسان‌های جهانی قرار گرفته است. این پیام‌رسان علاوه بر امکانات پیام‌رسانی ساده، امکانات فراتری از جمله ایجاد گروه، کانال، ربات و ارسال پیام صوتی را ارائه می‌دهد.



شکل ۱. محبوب‌ترین برنامه‌های پیام‌رسان تلفن همراه جهانی در اکتبر ۲۰۲۰. بر اساس تعداد کاربران فعال ماهیانه [۳].

گروه‌ها، قابلیت‌های جهت بحث و تبادل نظر کاربران حول موضوعی واحد هستند. کانال‌ها، کاربردی یک‌طرفه دارند و اعضای کانال تنها بیننده پیام‌ها هستند و نمی‌توانند همانند گروه پیام ارسال نمایند. بسیاری از شرکت‌ها یک گروه جهت پشتیبانی و یک کانال برای معرفی محصولاتشان دارند. همچنین بسیاری از خبرگزاری‌ها برای انتشار اخبار از کانال استفاده می‌نمایند.

موتور جستجوی تلگرام، به‌صورت جستجوی سراسری<sup>۳</sup> است؛ به‌طوری‌که یافتن کاربر، گروه و کانال را در یک بخش ارائه می‌دهد. یافتن کاربر، تنها با داشتن آی‌دی<sup>۴</sup> دقیق او امکان‌پذیر است و تلگرام قادر به یافتن کاربران بدون آی‌دی نیست. همچنین یافتن گروه و کانالی خاص، تنها با داشتن نام دقیق آنان امکان‌پذیر است. در غیر این صورت، می‌توان کلمات دلخواهی که به دنبال گروه یا کانال مرتبط با آن هستیم در بخش جستجو وارد نماییم. نتایج جستجو بسیار محدود است و این پیام‌رسان قادر به ارائه تمامی نتایج مرتبط نیست. موتور جستجوی تلگرام بر روی گروه‌ها و کانال‌ها هیچ‌گونه تحلیلی انجام نمی‌دهد. به دلیل وجود این نیاز گسترده، سامانه

<sup>۴</sup>@ Id  
<sup>۵</sup>idekav.com/

<sup>۱</sup> Real-time text  
<sup>۲</sup> Secret chat  
<sup>۳</sup> Global Search

## Archive of SID

و آزمون<sup>۳</sup> و به‌وسیله دو خطای جذر میانگین مربع‌ها<sup>۴</sup> و میانگین قدرمطلق‌ها<sup>۵</sup> مورد بررسی قرار گرفته است. نوآوری این پژوهش در ارائه روشی طی گام‌هایی جهت تکمیل و توسعه مقاله [۸] است. خروجی هر گام، فهرستی از گروه‌های رتبه‌بندی شده است که به ورودی گام بعدی وارد می‌شود. درواقع هر گام با افزودن پارامتری جدید، سعی در رفع نواقص گام قبل را دارد. جهت توصیه کاربر در هر گام، با توجه به فهرست گروه‌های رتبه‌بندی شده، اعضای گروه‌ها به ترتیب فهرست تا رسیدن به تعداد کاربر موردنظر با یکدیگر ادغام می‌شوند. انگیزه ما از ارائه هر گام، بهبود رتبه‌بندی گروه‌ها با درنظر گرفتن اطلاعاتی جدید از گروه‌های کاربران ورودی به‌همراه رفع نواقص موجود در دیگر گام‌ها می‌باشد. به‌طور خلاصه گام‌های روش پیشنهادی عبارتند از:

- ایجاد روشی جهت بررسی مهاجرت بین گروهی کاربران و جایگزینی هر گروه با مهاجرتش در گروه‌های به‌دست‌آمده از روش مقاله [۸] (انگیزه: تمایز در وضعیت متفاوت کاربران در هر گروه و مدل‌سازی رفتار بین گروهی آنان)
  - استخراج گروه‌های کاربران ادمین ورودی توصیه‌گر و جایگزینی با مهاجرتشان و همچنین ادغام گروه‌های جدید با گروه‌های گام قبل و اولویت دادن به رتبه‌بندی گروه‌های جدید (انگیزه: مدل‌سازی رفتار کاربران ادمین ورودی)
  - رتبه‌بندی گروه‌ها بر اساس درصد اشتراک با کاربران ورودی (انگیزه: رفع مشکل رتبه‌بندی گروه‌های با تعداد عضو زیاد)
  - بهبود درصد اشتراک هر گروه از طریق ضرب نمودن درصد اشتراک در لگاریتم تعداد اعضای هر گروه (انگیزه: رفع مشکل عدم تناسب در عدد صورت و مخرج کسر حاصل از گام قبل و بهبود رتبه‌بندی گروه‌های با تعداد عضو کم)
  - رتبه‌بندی گروه‌ها بر اساس درصد کاربر جاری زیاد (انگیزه: تضمین کیفیت با رتبه‌بندی حداکثر کاربران جاری هر گروه)
  - رتبه‌بندی گروه‌ها بر اساس درصد کاربر ادمین کم (انگیزه: تضمین کیفیت با رتبه‌بندی حداقل کاربران ادمین هر گروه)
  - رتبه‌بندی گروه‌ها بر اساس درصد کاربر سابق کم (انگیزه: تضمین کیفیت با رتبه‌بندی حداقل کاربران سابق هر گروه)
- ادامه مقاله به شرح زیر بیان می‌شود. در بخش ۲ مطالعات پیشین بررسی می‌شوند. در بخش ۳ روش پیشنهادی مطرح می‌شود. بخش ۴ به انجام آزمایش‌ها و بیان نتایج ارزیابی می‌پردازد. درنهایت بخش ۵ نتیجه‌گیری و کارهای آینده را بیان می‌نماید.

توصیه‌گر ارائه‌شده در این مقاله متفاوت از بسیاری سامانه‌های توصیه‌گر معمول است. در شبکه‌های اجتماعی بسیاری از سامانه‌های توصیه‌گر با در نظر گرفتن اطلاعات و علایق یک کاربر، مجموعه‌ای کاربر به او توصیه می‌نمایند. درحالی‌که این پژوهش قادر است بدون هیچ‌گونه محدودیتی در دریافت کاربر، اطلاعات چندین کاربر را دریافت و مجموعه‌ای کاربر مشابه با آنان را توصیه نماید. درواقع اینجا این مسئله مطرح می‌شود که چگونه کاربرانی با علایق متفاوت را تحلیل نماییم؟

در این پژوهش ما با به‌دست آوردن متغیرهایی عددی از گروه‌های تلگرام فرمولی ریاضی جهت توصیه کاربر ارائه داده‌ایم. این فرمول با بررسی گروه‌های کاربران و تحلیل مهاجرت گروه‌هایشان به‌وسیله گراف، اشتراک علایق کاربران ورودی را به‌صورت یکجا تحلیل می‌نماید. گراف استفاده‌شده در این مقاله گرافی ناهمگون است که کاربران را از طریق عضویت در گروه‌ها به یکدیگر متصل می‌نماید. در این مقاله سعی شده تا نتیجه به‌دست‌آمده در مقاله [۸] بهبود داده شود. آنچه موجب متمایز شدن روش پیشنهادی در این مقاله شده است، در نظر گرفتن وضعیت متفاوت کاربران در هر گروه است. این پژوهش کاربران ورودی را همانند مقاله [۸] به یک دید نمی‌نگرد و کاربران را از سه جهت (ادمین، جاری و سابق) بررسی می‌نماید. در نظر گرفتن وضعیت‌های متفاوت برای هر کاربر، موجب مدل‌سازی بهتر کاربران و به‌موجب آن کاهش خطای پیش‌بینی گردیده است. رفتار کاربران در این پژوهش به‌وسیله عضویتشان در گروه‌ها و مهاجرت به گروه‌های دیگر مدل‌سازی شده است. مهاجرت کاربران در بین گروه‌ها با حداکثر کاربران مهاجرت کرده تحلیل شده است؛ چراکه در نظر گرفتن تغییر علایق اشتراکی کاربران نسبت به ترک یک گروه و حضور در گروهی دیگر بسیار مهم است. در این مقاله با توجه به مدل‌سازی مهاجرت بین گروهی کاربران، رفتار توصیه‌گر بر اساس اطلاعات استخراج‌شده از رفتار کاربران تغییر می‌کند.

به‌طورکلی هدف از ارائه این سامانه توصیه‌گر ایجاد روشی جامع و مبتنی بر گراف، جهت توصیه کاربرانی مشابه با کاربران ورودی توصیه‌گر است. از مزایای این پژوهش می‌توان به عمومیت آن اشاره نمود. روش پیشنهادی، محدود به پیام‌رسان تلگرام نمی‌شود و می‌توان آن را بر روی پیام‌رسان‌هایی که قابلیت ایجاد گروه توسط کاربران را در بر دارند و همچنین رابط برنامه‌نویسی<sup>۱</sup> آنان در دسترس است، پیاده‌سازی و بررسی نمود. داده‌های مورداستفاده شامل بیش از ۹۰۰ هزار سوپرگروه و ۱۲۰ میلیون کاربر در پیام‌رسان تلگرام است. ارزیابی روش پیشنهادی بر روی ۱۰۰ گروه باکیفیت انجام شده است. خطای پیش‌بینی، با تقسیم‌بندی داده‌ها به دو قسمت آموزش<sup>۲</sup>

<sup>۴</sup> Root Mean Square Error (RMSE)

<sup>۵</sup> Mean Absolute Error(MAE)

<sup>۱</sup> Application Programming Interface (API)

<sup>۲</sup> Train

<sup>۳</sup> Test

## ۲. مطالعات پیشین

طی سال‌های اخیر، توصیه اقلام مختلف به کاربران مورد توجه بسیاری از محققان و بازاریابان قرار گرفته است. سامانه‌های توصیه‌گر، بر اساس میزان اطلاعات استخراج شده از کاربران دسته‌بندی می‌شوند و همچنین توصیه‌ها با افزایش این اطلاعات، دقیق‌تر می‌شوند. توصیه به کاربر شامل موارد مختلفی از جمله کتاب [۹]، موسیقی [۱۰]، سفر [۱۱]، فیلم [۱۲] و همچنین کاربر است. به‌طور کلی ادبیات تحقیق در این مقاله به دو دسته توصیه کاربر به یک کاربر و توصیه کاربر به گروهی از کاربران تقسیم می‌شود.

### ۱.۲. توصیه کاربر به یک کاربر

این دسته از سامانه‌های توصیه‌گر، تنها کاربران هدف را به یک کاربر توصیه می‌نمایند. ما در این دسته، پیشینه پژوهش را با سه الگوی عنوان توصیه، نوع پالایش و کاربرد توصیه‌گر در محیط‌های مختلف دسته‌بندی نموده‌ایم.

**عنوان توصیه:** در پژوهش‌های مختلف، توصیه کاربر در عناوین توصیه افراد<sup>۱</sup> [۱۳] تا [۱۵]، توصیه دوست<sup>۲</sup> [۱۶] تا [۱۸] و توصیه کاربر [۱۹] تا [۲۱] بیان شده‌اند. می‌توان گفت توصیه دوست و توصیه افراد زیرمجموعه‌ای از توصیه کاربر هستند. به‌عنوان مثال، توصیه افراد شامل پیشنهاد فروشنده به مشتری و پیشنهاد افراد با تخصص‌های متفاوت در لینکدین<sup>۳</sup> است. توصیه دوست اغلب در شبکه‌های اجتماعی بکار می‌رود که با تجزیه و تحلیل علاقه کاربر، دوستانی با علایق مشابه به او پیشنهاد می‌شود. عنوان توصیه در این پژوهش، به دلیل اینکه ویژگی‌های دوستی، تخصص و یا مهارت‌های کاربران را در نظر ندارد، به‌طور کلی با توصیه کاربر بیان شده است.

**نوع پالایش:** سامانه‌های توصیه‌گر، با توجه به در نظر گرفتن میزان اطلاعات کاربران به روش‌های پالایشی متفاوتی تقسیم‌بندی می‌شوند. در ادامه سه نمونه از پرکاربردترین پالایش‌ها به‌طور خلاصه شرح داده شده‌اند.

- **پالایش مبتنی بر محتوا<sup>۴</sup>:** این نوع پالایش، تنها از اطلاعات خود کاربر جهت توصیه کاربرانی مشابه استفاده می‌نماید و اطلاعات دیگر کاربران را در نظر ندارد. از ویژگی‌های این پالایش می‌توان به در نظر گرفتن پیام‌های کاربر، جنسیت کاربر، رنگ مورد علاقه کاربر و غیره اشاره نمود.
- **پالایش اشتراکی<sup>۵</sup>:** این پالایش یکی از محبوب‌ترین روش‌های پالایشی در سامانه‌های توصیه‌گر است که در سایت‌های آمازون<sup>۶</sup> و نتفلیکس<sup>۷</sup> نیز استفاده فراوانی دارد. این پالایش با جستجو و یافتن کاربرانی که علایق مشابهی با کاربر هدف دارند، سعی در

توصیه‌هایی دقیق‌تر می‌نماید. در این پالایش فرض بر این است که کاربرانی که در گذشته علایق مشابه داشته‌اند، در آینده نیز علایقی مشابه خواهند داشت. پالایش اشتراکی به دو دسته مبتنی بر حافظه<sup>۸</sup> و مبتنی بر نمودار<sup>۹</sup> تقسیم می‌شود. روش مبتنی بر حافظه بر اساس بازخورد گرفتن از کاربر است و روش مبتنی بر نمودار از یک گراف که فعالیت و رفتار کاربران را مدل‌سازی می‌کند، جهت توصیه استفاده می‌نماید [۱۵]. این مقاله با استفاده از یک گراف که عضویت کاربران را در گروه‌ها با وضعیت‌های متفاوت در نظر دارد، توانسته است رفتار کاربران را جهت توصیه، مدل‌سازی نماید. رفتار کاربران بر اساس سابقه عضویت، جاری بودن عضویت و ادمین بودن آنان در گروه‌ها در نظر گرفته شده است. روش پیشنهاد داده شده در این مقاله، روشی مبتنی بر نمودار در پالایش اشتراکی است.

### • پالایش ترکیبی<sup>۱۰</sup>: این پالایش، با ترکیب نمودن دیگر

پالایش‌ها سعی در کاهش محدودیت آنان می‌نماید.

پژوهش‌های بسیاری در ارتباط با سامانه‌های توصیه‌گر مبتنی بر پالایش‌ها و عناوین توصیه‌ای مختلف انجام شده است که در این بخش تعدادی از آن‌ها بر اساس عنوان توصیه، نوع پالایش و کاربرد توصیه‌گر در جدول ۱ ارائه شده است.

### ۲.۲. توصیه کاربر به گروهی از کاربران

برخی از پژوهش‌ها به جای توصیه کاربر به یک کاربر به توصیه کاربر به گروهی از کاربران وابسته هستند. در توصیه گروهی با در نظر گرفتن ترجیحات مجموعه‌ای از کاربران، به صورت گروهی به آنان کاربرانی مشابه توصیه می‌شود. در این روش‌ها اطمینان از رضایت همه اعضای گروه با ترجیحات مختلف بسیار دشوار است [۲۲]. جیو و همکاران [۲۳] الگوریتمی جهت توصیه کاربر به گروهی از کاربران، به وسیله الگوریتم  $K - means$  و شبکه عصبی ارائه داده‌اند. آنان توصیه را بر اساس چند معیار و ترجیحات گروهی از کاربران بیان کرده‌اند. اورتگا و همکاران [۲۴] معیاری مبتنی بر ترکیب متغیرهای عددی و غیر عددی، جهت توصیه اقلام مورد علاقه گروهی را در بین کاربران ارائه داده‌اند. کریم‌پور و همکاران [۲۵] روشی جهت توصیه کاربرانی به گروهی کاربر بر اساس محتوای نمایه‌ای استخراج شده از گروه‌ها ارائه دادند، آنان از روش‌های پردازش زبان طبیعی جهت ساخت کیسه کلماتی برای هر گروه بهره برده‌اند و سپس به وسیله جستجو در بیشترین گروه‌های به دست آمده، کاربران را توصیه می‌نمایند. یالکین و همکاران [۲۶] روشی جهت توصیه گروه مبتنی

<sup>۱</sup> Amazon.com

<sup>۲</sup> Netflix.com

<sup>۳</sup> Memory-based

<sup>۴</sup> Model-based

<sup>۵</sup> Hybrid-based filtering

<sup>۱</sup> People Recommendation

<sup>۲</sup> Friend Recommendation

<sup>۳</sup> LinkedIn

<sup>۴</sup> Content-based filtering

<sup>۵</sup> Collaborative filtering

اطلاعات زمینه را با استفاده از یادگیری نیمه نظارتی بر روی مدل توصیه اعمال کردند. برخی پژوهش‌ها همانند [۲۷] روشی برای غلبه بر پراکندگی داده‌ها<sup>۴</sup> در توصیه کاربر به گروهی از کاربران ارائه داده‌اند.

بر آنتروپی<sup>۱</sup> بر اساس نقطه نظرات مشترکی که بین کاربران یک گروه وجود دارد ارائه داده‌اند. جیوگ و همکاران [۲۲] روشی جهت توصیه کاربر به گروهی از کاربران به‌وسیله مدل‌سازی رفتار گروهی کاربران و اطلاعات مبتنی بر زمینه<sup>۲</sup> ارائه داده‌اند. آنان اطلاعات زمینه را از یک شبکه مبتنی بر اطلاعات ناهمگن<sup>۳</sup> استخراج کرده‌اند. سپس

جدول ۱. مقایسه پژوهش‌های پیشین بر اساس نوع پالایش، عنوان توصیه، کاربرد و رویکرد

مقاله	نوع پالایش	عنوان توصیه	کاربرد	رویکرد مقاله
[۱۶]	ترکیبی	دوست	یلپ <sup>۵</sup>	روش این مقاله ترکیبی از توصیه مبتنی بر پالایش اشتراکی و توصیه معنایی و اجتماعی است. بخش معنایی، به توصیه دوست بر اساس شباهت بین هر کاربر و دوستانش می‌پردازد و بخش اجتماعی، رفتار کاربر را از جمله درجه دوستی بررسی می‌نماید.
[۱۳]	ترکیبی	افراد	لینکدین	این مقاله از ویژگی متمایز افراد مانند عملکرد، سطح کار و تجربه کاری جهت توصیه استفاده کرده است. روش پیشنهادی با استفاده از داده‌کاوی، شبکه‌های عصبی و تکنیک‌های فازی از یک استراتژی ترکیبی استفاده می‌کند که ترکیبی از پالایش‌های اشتراکی و محتوایی است.
[۱۹]	محتوا	کاربر	توییتر <sup>۶</sup>	این مقاله روشی جهت مدل‌سازی و توصیه کاربر با بهره‌گیری از روابط صریح <sup>۷</sup> برای ساخت مشخصات شخصی کاربر ارائه داده است. این روش با هدف ایجاد پروفایل‌های شخصی از طریق یک روش جایگزین برای ارائه توصیه‌های دقیق‌تر ایجاد شده است.
[۱۴]	محتوا	افراد	توییتر	این مقاله روشی جهت مدل‌سازی و توصیه کاربران متخصص به‌وسیله اطلاعات پروفایل هر کاربر ارائه داده است. متخصص توصیه‌شده، بر اساس محتوای توییت که منتشر می‌کنند و اینکه آیا این محتوا موردعلاقه کاربر است انتخاب می‌شود. در این مقاله از تکنیک‌های استخراج متن برای پردازش کلمات در توییت‌ها استفاده شده و از این طریق شباهت بین کاربران تعیین می‌شود.
[۱۷]	اشتراکی	دوست	توییتر فیس‌بوک	در این مقاله دو الگوریتم مجزا برای توصیه دوست با استفاده از پالایش اشتراکی مبتنی بر مدل ارائه شده است. الگوریتم اول، تعداد دوستان متقابل هر کاربر را در نظر می‌گیرد و الگوریتم دوم برای اولویت‌بندی کاربران و تأثیر کاربران مختلف ایجاد شده است. به‌طوری‌که برای هر کاربر یک رتبه تأثیرگذاری نسبت داده می‌شود. (به‌عنوان مثال اگر یک کاربر ضریب تأثیر ۱ را دارد این ضریب در بین دوستانش تقسیم می‌شود).
[۲۰]	محتوا	کاربر	توییتر	در این مقاله احساسات کاربران موردتوجه قرار گرفته است و یک تابع وزن به نام SVO (احساس-حجم-واقعیت) <sup>۸</sup> پیشنهاد شده است. این روش به‌وسیله پالایش مبتنی بر محتوا و استخراج کلمات کلیدی از پیام‌های کاربران، به‌وسیله شباهت کسینوسی شباهت کلمات کلیدی کاربران را با یکدیگر محاسبه می‌نماید.
[۱۸]	محتوا	دوست	فلیکر	این مقاله از ویژگی‌های جنسیت، رنگ، سن و علاقه‌مندی کاربران استفاده می‌نماید. توصیه دوست در این پژوهش بر اساس یک روش دولایه‌ای انجام شده است. لایه اول برای بررسی گراف دوستی بین کاربران و لایه دوم گراف برچسب خورده از ویژگی‌های هر کاربر را در نظر گرفته است.
[۲۱]	محتوا	کاربر	توییتر سیناویبو <sup>۹</sup> اپنین <sup>۱۰</sup>	این مقاله روشی با استفاده از استخراج روابط اجتماعی و علاقه‌مندی کاربر در دو فاز ارائه کرده است. فاز اول شامل استخراج موضوعات موردعلاقه و موضوعات اجتماعی کاربران بر اساس روابط بین دنبال‌شونده <sup>۱۱</sup> ، دنبال‌کننده <sup>۱۲</sup> و ارتباط متقابل دنبال‌کردن <sup>۱۳</sup> برای بررسی ترجیحات کاربران است. فاز دوم، توصیه کاربر است که یک روش مبتنی بر تشکیل جوامع کاربران بر اساس تخمین موضوعات می‌باشد. درنهایت فاکتورگیری ماتریس در هر جامعه برای تولید کاربران مرتبط با کاربر هدف انجام می‌شود.
[۱۵]	اشتراکی	افراد	شبکه‌اجتماعی آنلاین تجاری	این مقاله یک الگوریتم جدید پالایش اشتراکی مبتنی بر مدل را با استفاده از همسایگان کاربران ایجاد کرده است تا بر اساس شباهت از نظر جذابیت و سلیقه، برای کاربر مشخصی سایر کاربرانی که ممکن است علاقه‌مند ارتباط با او باشند را پیش‌بینی کند.

<sup>۸</sup> Sentiment-volume-objectivity

<sup>۹</sup> Sina Weibo

<sup>۱۰</sup> Epinions

<sup>۱۱</sup> Followee

<sup>۱۲</sup> Follower

<sup>۱۳</sup> Following

<sup>۱</sup> Entropy

<sup>۲</sup> Context

<sup>۳</sup> Heterogeneous information network (HIN)

<sup>۴</sup> Data sparsity

<sup>۵</sup> Yelp

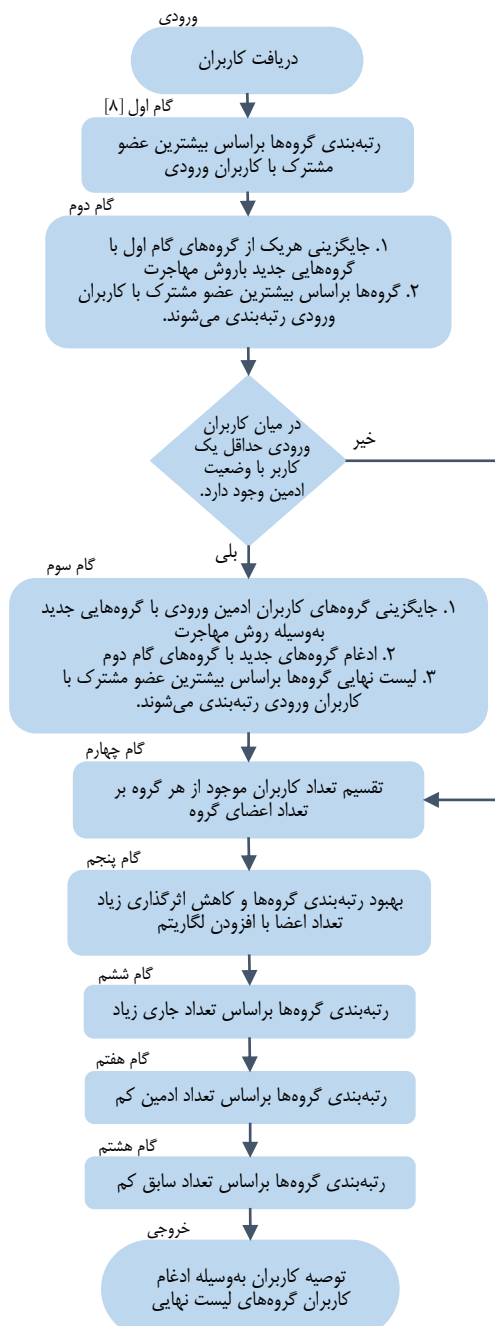
<sup>۶</sup> Twitter

<sup>۷</sup> Explicit

## Archive of SID

در بخش مطالعات پیشین بیان شد، با توجه به اینکه گام‌های روش پیشنهادی در ادامه روش گروه‌های مشابه [۸] هستند، روش مقاله [۸] گام اول نام‌گذاری شده است و روش پیشنهادی از گام دوم شروع شده است. در گام اول ابتدا گروه‌های کاربران ورودی استخراج می‌شوند، سپس این گروه‌ها بر اساس بیشترین اشتراک با کاربران ورودی رتبه‌بندی می‌شوند. خروجی این گام، فهرستی از گروه‌های رتبه‌بندی شده می‌باشد که در ادامه با متغیر  $K$  بیان می‌شود. شکل ۳، شبه کد گام اول را نشان می‌دهد.

جهت درک بهتر کارکرد هر یک از گام‌ها، مثالی جامع و عددی که تمامی گام‌های روش پیشنهادی را پوشش می‌دهد، پس از توضیحات هر گام آورده شده است. در واقع این مثال با اعمال روی هر یک از



شکل ۲. فرآیند اجرای روش پیشنهادی

با توجه به پژوهش‌های مطرح‌شده، این پژوهش با مدل‌سازی رفتار کاربران ورودی، به‌وسیله وضعیت آنان در گروه‌ها، فهرستی از کاربران را که برای توصیه به کاربران ورودی مناسب هستند، توصیه می‌نماید. در این مقاله جهت شروع و رتبه‌بندی اولیه گروه‌های کاربران ورودی از روش مقاله [۸] که از گرافی متناسب با گراف موردنظر در این پژوهش بهره برده است، استفاده شده است. گراف مقاله [۸]، همانند گراف مورد استفاده در این پژوهش، کاربران را بر اساس عضویتشان در گروه‌ها به یکدیگر متصل می‌نماید. ما روش مقاله [۸] را که برای توصیه گروه بیان شده بود، جهت توصیه کاربر در نظر گرفته‌ایم و روش پیشنهادی این پژوهش را با آن مقایسه نموده‌ایم.

در مقاله [۸] ابتدا تمامی گروه‌های کاربران ورودی، از گراف عضویت استخراج می‌شوند. سپس بررسی می‌شود که از هر گروه، چه تعداد کاربر در مجموعه کاربران ورودی قرار دارند. برای تمامی گروه‌های کاربران ورودی، این اشتراک بررسی می‌شود و سپس گروه‌ها به ترتیب بیشترین عضو مشترک، رتبه‌بندی می‌شوند. از معایب این روش می‌توان به دید کلی روش، نسبت به کاربران و رتبه‌بندی همیشگی گروه‌های با تعداد عضو زیاد اشاره نمود. همچنین این روش، گروه‌ها را بدون در نظر گرفتن ترجیحات دیگر کاربران توصیه می‌کند و هیچ‌گونه تحلیلی جهت رتبه‌بندی کیفیت گروه‌ها در نظر نمی‌گیرد. در پژوهش پیش‌رو، کاربران ورودی با سه وضعیت (عضو سابق، جاری و ادمین) در هر گروه در نظر گرفته شده‌اند. همچنین این پژوهش علاوه بر رفع رتبه‌بندی گروه‌های با تعداد عضو زیاد، به مدل‌سازی مهاجرت بین گروهی کاربران و کیفیت گروه‌ها پرداخته است. با توجه به اینکه روش پیشنهادی در این مقاله، شامل گام‌هایی در ادامه روش مقاله [۸] است، روش مقاله [۸] با "گام اول" نام‌گذاری شده است.

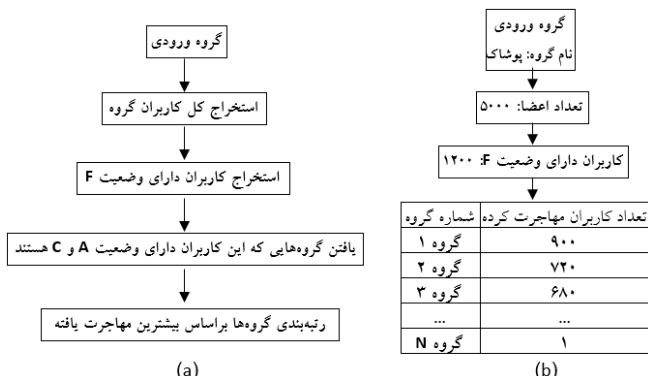
### ۳. روش پیشنهادی

روش پیشنهادی به‌طورکلی در هشت گام دنبال می‌شود. شکل ۲ فرآیند اجرای روش پیشنهادی را نشان می‌دهد. هر یک از گام‌ها، گام قبل خود را تکمیل می‌نماید و هر گام تمامی گام‌های قبلی را دربر دارد. گام‌های ششم تا هشتم جهت افزودن کیفیت به روش پیشنهادی، مطرح شده‌اند. گام هشتم به‌عنوان آخرین گام، دربردارنده تمامی گام‌های روش پیشنهادی است. ورودی هر گام، مجموعه‌ای از کاربران است و خروجی هر یک، فهرستی از گروه‌های رتبه‌بندی شده است که کاربران این گروه‌ها به ترتیب فهرست تا رسیدن به تعداد کاربر موردنظر با یکدیگر ادغام می‌شوند. هدف در هر گام، دستیابی به مجموعه‌ای کاربر (با تعداد موردنظر)، مشابه با کاربران ورودی است. تعداد کاربر موردنظر، در واقع خط پایانی است که تعیین‌کننده تعداد کاربر موجود در فهرست نهایی کاربران پیشنهاد شده توسط توصیه‌گر است. به‌عنوان مثال تعداد کاربر موردنظر را می‌توان ۱۰، ۲۰، ۳۰ برابر ورودی و یا هر مقدار دلخواهی تعیین نمود. همان‌طور که

## Archive of SID

**گام ۲:** این گام به مدل‌سازی ترجیحات و رفتار اعضای گروه‌ها می‌پردازد. ورودی این گام نتایج حاصل از گام اول است. در این گام ابتدا روشی جهت بررسی مهاجرت حداکثری کاربران از گروه‌ها بیان می‌شود و در نهایت این گام تکمیل می‌شود.

**مهاجرت:** روشی جهت تحلیل حرکت عمده کاربران از هر گروه به گروهی دیگر است. این روش، بر اساس تحلیل وضعیت کاربران در گروه‌ها به دست می‌آید. طبق اطلاعات مورداستفاده در این مقاله، هر کاربر در گروهی که وجود دارد، دارای یکی از سه وضعیت  $F$  یا  $C$  یا  $A$  است. شکل ۵- (a) فرآیند اجرای روش مهاجرت را نشان می‌دهد. با توجه به این شکل، کلیت اجرای روش مهاجرت بدین سبک است که برای هر گروه، ابتدا کاربران با وضعیت  $F$  استخراج می‌شوند، سپس بررسی می‌شود که این کاربران در چه گروه‌هایی دارای وضعیت  $A$  و  $C$  هستند. گروه‌های به‌دست‌آمده، بر اساس بیشترین تعداد کاربر مهاجرت یافته، رتبه‌بندی می‌شوند و گروهی که بیشترین تعداد کاربران به آن مهاجرت کرده‌اند، به‌عنوان گروه نهایی حاصل می‌شود. برای بیان بهتر این روش، مثالی عددی در شکل ۵- (b) نشان داده شده است. با توجه به این شکل، گروهی با نام "پوشاک" و تعداد اعضای ۵۰۰۰ در نظر گرفته شده است. ابتدا تمامی کاربران این گروه را که دارای وضعیت  $F$  هستند، استخراج می‌نماییم. در این مثال تعداد کاربران دارای وضعیت  $F$ ، ۱۲۰۰ کاربر می‌باشند. حال برای هر یک از ۱۲۰۰ کاربر، به‌صورت جداگانه بررسی می‌نماییم که در چه گروه‌هایی دارای وضعیت  $A$  و  $C$  هستند. طبق جدول نشان داده شده در شکل ۵- (b) که نتیجه نهایی حاصل از روش مهاجرت را بر روی گروه "پوشاک" نشان می‌دهد، فهرستی از گروه‌های رتبه‌بندی شده به‌دست‌آمده است. بر اساس این جدول تعداد ۹۰۰ کاربر از گروه "پوشاک" به گروه ۱، ۷۲۰ کاربر به گروه ۲، ... و یک کاربر به گروه  $N$  مهاجرت کرده‌اند.



شکل ۵. (a) فرآیند اجرای روش مهاجرت، (b) مثالی از اجرای مهاجرت

در این گام، روش مهاجرت به‌صورت جداگانه بر روی هر یک از گروه‌های حاصل از گام اول (همانند شکل ۵- (b))، اعمال می‌شود و تنها گروهی که بیشترین کاربران به آن مهاجرت کرده‌اند، با گروه موردنظر جایگزین می‌شود. به‌عنوان مثال در شکل ۵- (b) گروه ۱ که

### Pseudo-code of the first step

```

1. Input:  $U$  (list of all input users)
2. Output:  $P$  (final user list)
3. Variables:
4. var  $u_i$ : each user in  $U$ 
5. var  $G$ : [initially empty] set of all groups inferred from  $U$ 
6. var  $g_k$ : each group in  $G$ 
7. var  $S$ : count of  $g_k$  members
8. var  $T$ : a temporary list
9. Pseudo-code:
10. for each user  $u_i$  in  $U$ :
11.      $G.append(u_i \text{ groups});$  // all groups associated with  $u_i$  by any relation is appended to  $G$ 
12. for each group  $g_k$  in  $G$ :
13.     set  $S$  = count of  $g_k$  members;
14. set  $T$  = a list of users of  $g_k$ s with highest  $S$ ; //  $g_k$  groups are ranked with the most  $S$  and users are merged from the list of groups, respectively.
15. while  $P.length() <$  desired number of output users:
16.      $P.append(T.pop());$  // one of users from  $T$  is appended to  $P$  and removed from  $T$ 
17. return  $P$ ;
    
```

شکل ۳. شبه کد گام اول

گام‌ها تا گام هشتم پیش می‌رود. در این مثال، ۱۶۰۰ کاربر به‌عنوان مجموعه کاربران ورودی به گام اول داده می‌شود و در طی آن هر یک از گام‌ها نتایج حاصل از گام قبل خود را بهبود می‌دهد. جهت بررسی مثال عددی جامع، خروجی هر گام (فهرستی از گروه‌های رتبه‌بندی شده) و ورودی گام بعدی می‌باشد. در نهایت، مجموعه‌ای کاربر با تعداد دلخواه توصیه می‌شود. در شکل ۴ نتایج حاصل از اجرای گام اول بر روی این مثال نشان داده شده است. با توجه به شکل ۴ که نتایج حاصل از اجرای گام اول را بر روی مثال عددی نشان می‌دهد، این گام گروه‌های کاربران را بر اساس بیشترین اشتراک با کاربران ورودی رتبه‌بندی نموده است. با توجه به فهرست  $S$  که نتیجه اجرای گام اول می‌باشد، از تعداد ۱۶۰۰ کاربر ورودی، ۱۳۰۰ کاربر در گروه شماره ۱، ۹۶۰ کاربر در گروه شماره ۲، ... و یک کاربر در گروه شماره  $N$  وجود دارد.

(S)

بیشترین اشتراک	شماره گروه
۱۳۰۰	گروه ۱
۹۶۰	گروه ۲
۸۵۰	گروه ۳
...	...
۱	گروه N

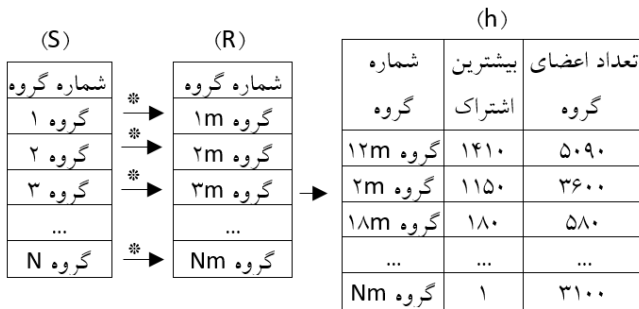
کاربران ورودی: تعداد: ۱۶۰۰

شکل ۴. نتایج حاصل از اجرای گام اول بر روی مثال عددی

در ادامه جهت سهولت در بیان هریک از گام‌های روش پیشنهادی به بیان برخی اختصارات می‌پردازیم.

- $U$ : مجموعه کاربران ورودی
- $M$ : تعداد اعضای گروه
- $A$ : وضعیت کاربر در گروه، ادمین یا مدیر است.
- $F$ : وضعیت کاربر در گروه، سابق یا قدیمی است.
- $C$ : وضعیت کاربر در گروه، جاری است.

مشترک با کاربران ورودی رتبه‌بندی می‌شوند. فهرست نهایی گام دوم با نام  $h$  نشان داده شده است.



\* بررسی روش مهاجرت و یافتن گروهی با بیشترین مهاجرت یافته.

شکل ۷. نتایج حاصل از اجرای گام دوم بر روی مثال عددی

**گام ۳:** این گام، کاربران مجموعه  $U$  را از جهتی دیگر و بر اساس مدل‌سازی مهاجرت گروه‌های کاربران با وضعیت  $A$  بررسی می‌نماید. در این گام، ابتدا کاربران با وضعیت  $A$  از بین مجموعه کاربران ورودی ( $U$ ) استخراج می‌شوند. گروه‌هایی که این کاربران دارای وضعیت  $A$  هستند از گراف عضویت به دست می‌آید. سپس این گروه‌ها بر اساس روش مهاجرت، با گروه‌هایی جدید، جایگزین می‌شوند. در نهایت، فهرست گروه‌های جدید، بر اساس بیشترین عضو مشترک با کاربران ورودی رتبه‌بندی می‌شود. در ادامه جهت سهولت در بیان مطالب، فهرست جدید با نام مهاجرت ادمین‌ها بیان می‌شود. ما پس از بررسی و اشتراک‌گیری گروه‌های فهرست مهاجرت ادمین‌ها با گروه‌های حاصل از گام دوم، به این نتیجه دست‌یافتیم که در اکثر مواقع، بسیاری از گروه‌های موجود در فهرست مهاجرت ادمین‌ها، در فهرست گروه‌های گام دوم وجود دارند. در نتیجه، گروه‌های مشترک، از فهرست گروه‌های گام دوم حذف می‌شوند و فهرست مهاجرت ادمین‌ها در ابتدای فهرست ویرایش شده (پس از حذف گروه‌های مشترک با گروه‌های مهاجرت ادمین‌ها) گام دوم، قرار می‌گیرند. این گام با متغیر  $h'$  و در رابطه (۲) نشان داده شده است. در رابطه (۲)، متغیر  $S$  و  $h$  به ترتیب بیانگر فهرست گروه‌های حاصل از گام اول و دوم هستند.

$$\text{step}^3: S \rightarrow h \rightarrow h' \quad (2)$$

شکل ۸ نتیجه اجرای گام سوم را بر روی فهرست  $h$  حاصل از گام دوم برای مثال عددی نشان می‌دهد. در این شکل با توجه به هدف گام دوم، ابتدا کاربران دارای وضعیت  $A$  از بین کاربران ورودی استخراج می‌شوند. در صورتی که بین کاربران ورودی کاربری با وضعیت  $A$  وجود نداشته باشد، این گام اجرا نخواهد شد و نتایج گام دوم به گام چهارم انتقال می‌یابد. در این مثال با توجه به اینکه تعداد کاربران ورودی با وضعیت  $A$ ، ۶۰ کاربر است گام سوم قابل اجرا می‌باشد.

بیشترین کاربر از گروه پوشاک به آن مهاجرت کرده‌اند با گروه پوشاک جایگزین می‌شود. برای هر یک گروه‌های گام اول به صورت جداگانه این مهاجرت بررسی و هر گروه با گروهی جدید جایگزین می‌شود. در نهایت، فهرست گروه‌های جدید به دست آمده که گروه‌های مهاجرت یافته یکایک گروه‌های گام اول هستند بر اساس بیشترین عضو مشترک با کاربران ورودی رتبه‌بندی می‌شود. این گام در ادامه با متغیر  $h$  بیان می‌شود و در رابطه (۱) نشان داده شده است. در رابطه (۱) متغیر  $S$  بیانگر فهرست گروه‌های نهایی حاصل از گام اول است. شبه کد گام دوم در شکل ۶ نشان داده شده است.

$$\text{step}^2: S \rightarrow h \quad (1)$$

#### Pseudo-code of the second step of the proposed method

```

1. Input:  $U$  (list of all input users)
2. Output:  $P$  (final user list)
3. Variables:
4. var  $u_i$ : each user in  $U$ 
5. var  $G$ : [initially empty] set of all groups inferred from  $U$ 
6. var  $g_k$ : each group in  $G$ 
7. var  $u'_i$ : each user in  $g_k$ 
8. var  $g'_i$ : each group  $u'_i$  is or was a member or admin
9. var  $R$ : [initially empty] migration list
10. var  $g_r$ : each group in  $R$ 
11. var  $h$ : count of  $g_r$  members
12. var  $T$ : a temporary list
13. Pseudo-code:
14. for each user  $u_i$  in  $U$ :
15.      $G.append(u_i \text{ groups});$  // all groups associated with  $u_i$  by any relation is appended to  $G$ 
16. for  $g_k$  in  $G$ :
17.     for  $u'_i$  in  $g_k$ :
18.         if status of  $u'_i$  is  $F$ : // if  $u'_i$  is a former member
19.             for  $g'_i$  in  $u'_i$  groups:
20.                 if status of  $u'_i$  in  $g'_i$  is  $C$  or  $A$ : // if  $u'_i$  is a current member or admin
21.                      $R.append(g'_i);$ 
22. for each group  $g_r$  in  $R$ :
23.     set  $h = \text{count of } g_r \text{ members};$ 
24. set  $T = \text{a list of users of } g_r \text{ s with highest } h;$  //  $g_r$  groups are ranked with the most  $h$  and users are merged from the list of groups, respectively.
25. while  $P.length() < \text{desired number of output users}$ :
26.      $P.append(T.pop());$  // one of users from  $T$  is appended to  $P$  and removed from  $T$ 
27. return  $P$ ;
    
```

شکل ۶. شبه کد گام دوم

شکل ۷ نتیجه اجرای گام دوم را بر روی فهرست  $S$  حاصل از گام اول برای مثال عددی نشان می‌دهد. طبق شکل ۷، برای هر یک از گروه‌های موجود در فهرست  $S$  روش مهاجرت به صورت جداگانه اجرا می‌شود و هر یک از گروه‌ها با گروه مهاجرت یافته‌اش جایگزین می‌شود. فهرست جدید مهاجرت یافته گروه‌ها با  $R$  نشان داده شده است. حرف  $m$  در کنار شماره هر گروه بیانگر مهاجرت است، به‌عنوان مثال گروه  $1m$  مهاجرت یافته گروه ۱ است. در انتهای این گام همانند گام اول، گروه‌های فهرست  $R$ ، بر اساس بیشترین عضو



Archive of SID

$H$  بر تعداد اعضای گروه ( $M$ ) تقسیم شده است تا گروه‌ها بر اساس درصد مشارکت رتبه‌بندی شوند. در انتهای این گام، فهرست گروه‌های به‌دست‌آمده، بر اساس درصد مشارکت ( $\frac{H}{M}$ ) نزولی رتبه‌بندی می‌شوند. برتری این گام در به دست آوردن درصد موجود بودن کاربران هر گروه بر اساس  $M$  است.

$$step\epsilon: S \rightarrow \frac{H}{M}, H: h \text{ or } h' \quad (3)$$

شکل ۹ نتیجه اجرای گام چهارم را بر روی نتایج حاصل از گام دوم و سوم برای مثال عددی نشان می‌دهد. در صورتی که که بین کاربران ورودی کاربری با وضعیت  $A$  وجود نداشته باشد ورودی این گام، فهرست خروجی گام دوم ( $h$ ) و اگر در بین کاربران ورودی حداقل یک کاربر با وضعیت  $A$  وجود داشته باشد، ورودی این گام، فهرست خروجی گام سوم ( $h'$ ) خواهد بود. فهرست  $a$  و  $b$  به ترتیب بر اساس نتایج حاصل از گام سوم و دوم، تعداد اشتراک هر گروه را بر تعداد اعضای آن تقسیم می‌نمایند تا گروه‌ها به‌جای تعداد اشتراک بر اساس درصد اشتراک رتبه‌بندی شوند. در فهرست  $a$  گروه  $35am$  مقدار  $0,319$  و در فهرست  $b$  گروه  $2m$  با مقدار  $0,833$  دارای بیشترین درصد اشتراک می‌باشند.

(a)

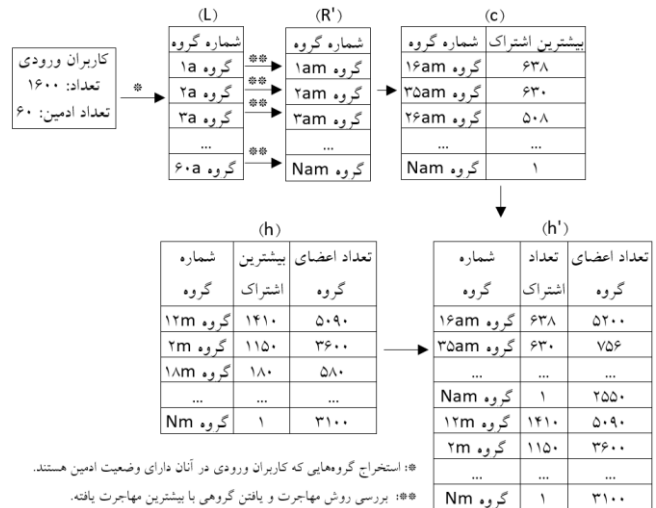
بیشترین درصد اشتراک	شماره گروه
$\frac{630}{756}$	گروه $35am$
$\frac{1150}{3600}$	گروه $2m$
$\frac{180}{580}$	گروه $18m$
$\frac{1410}{5090}$	گروه $12m$
$\frac{638}{5200}$	گروه $16am$
...	...
$\frac{1}{2550}$	گروه $Nam$
...	...
$\frac{1}{3100}$	گروه $Nm$

(b)

بیشترین درصد اشتراک	شماره گروه
$\frac{1150}{3600}$	گروه $2m$
$\frac{180}{580}$	گروه $18m$
$\frac{1410}{5090}$	گروه $12m$
...	...
$\frac{1}{3100}$	گروه $Nm$

شکل ۹. نتایج حاصل از اجرای گام چهارم بر روی مثال عددی

**گام ۵:** این گام جهت بهبود نتایج حاصل از گام چهارم ایجاد شده است. مشکل موجود در گام چهارم این است که با افزایش مقدار مخرج در کسر  $\frac{H}{M}$ ، مقدار صورت نمی‌تواند به همان اندازه افزایش یابد و این باعث رتبه‌بندی گروه‌های با تعداد  $M$  پایین می‌شود. این مشکل در این گام رفع و در رابطه (۴) نشان داده شده است. با توجه به رابطه (۴)، کسر  $\frac{H}{M}$  حاصل از گام سوم در لگاریتم  $M$  ضرب می‌شود. این ضرب باعث کاهش اثرگذاری زیاد  $M$  در مخرج می‌شود. برای کم اثر کردن  $M$  علاوه بر لگاریتم می‌توان کسر  $\frac{H}{M}$  را در ریشه  $M$  نیز ضرب نمود. ما جهت کم اثر کردن  $M$  با انجام آزمایش‌هایی عددی به این نتیجه دست یافتیم که استفاده از لگاریتم در این گام، بهتر از ریشه



شکل ۸. نتایج حاصل از اجرای گام سوم بر روی مثال عددی

در شکل ۸، فهرست  $L$  نشان‌دهنده گروه‌هایی است که ۶۰ کاربر ادمین ورودی در آنان دارای وضعیت ادمین هستند. هر یک از ۶۰ گروه با استفاده از روش مهاجرت با گروهی دیگر جایگزین می‌شود. این جایگزینی در فهرست  $R'$  نشان داده شده است. در فهرست  $R'$  شماره هر گروه با پسوند  $am$  که  $a$  مخفف ادمین و  $m$  مخفف مهاجرت است، نام‌گذاری شده است. به‌عنوان مثال، گروه  $2am$  مهاجرت یافته گروه  $2a$  می‌باشد. فهرست  $C$  گروه‌های فهرست  $R'$  را همانند گام اول، بر اساس بیشترین عضو مشترک با کاربران ورودی رتبه‌بندی نموده است. فهرست  $h$  نتایج حاصل از گام دوم را به همراه تعداد اعضای هر گروه نشان می‌دهد. اکنون جهت ایجاد فهرست نهایی در این گام، اشتراک گروه‌های فهرست  $C$  با گروه‌های فهرست  $h$  بررسی می‌شود و گروه‌های مشترک از فهرست  $h$  حذف می‌شوند. سپس خروجی این گام، فهرستی با نام  $h'$  می‌باشد که فهرست  $C$  در ابتدای آن و فهرست ویرایش شده  $h$  در ادامه گروه‌های  $C$  قرار گرفته است.

**گام ۴:** ورودی این گام، در صورتی که کاربری با وضعیت  $A$  در مجموعه  $U$  نباشد، گام دوم و در صورت وجود کاربر با وضعیت  $A$ ، گام سوم است. بر اساس نتایج به‌دست‌آمده از گام قبل (دوم یا سوم)، به دلیل اینکه گروه‌ها بر اساس تعداد کاربران مشترک با مجموعه کاربران ورودی نمایش داده می‌شوند؛ این امکان وجود دارد که گروهی که در رتبه اول فهرست قرار دارد تعداد  $M$  بسیار زیادی نسبت به گروه رتبه دوم داشته باشد و این یک عیب است که باعث می‌شود گروه‌های با تعداد  $M$  بسیار زیاد به دلیل اشتراک بالا با کاربران ورودی، همیشه در ابتدای فهرست قرار گیرند. این مشکل در گام اول تا سوم وجود دارد و در این گام، رفع شده است. این گام در رابطه (۳) نشان داده شده است. در رابطه (۳)، متغیر  $H$  تعداد کاربران موجود از هر گروه در بین مجموعه  $U$  است، که بر اساس نتایج گام دوم ( $h$ ) یا گام سوم ( $h'$ ) به‌دست‌آمده است. طبق رابطه (۳)، متغیر

## Archive of SID

**گام ۶:** از گام اول تا گام پنجم بر روی کیفیت گروه‌ها هیچ‌گونه تحلیلی انجام نشده است. در این گام، اولین کیفیت به گام پنجم بر اساس  $G_C$  افزوده می‌شود. برای هر گروه، هرچقدر نسبت  $M$  به  $G_C$  بیشتر باشد آن گروه، گروه خوبی است؛ چراکه آن گروه، کاربران جاری بیشتری نسبت به دیگر گروه‌ها دارد و کاربران کمتری آن گروه را ترک کرده‌اند. در واقع این روش فارغ از وضعیت ادمین و سابق بودن، کاربران هر گروه را بر اساس اعضای واقعی‌شان رتبه‌بندی می‌نماید. این گام در رابطه (۵) نشان داده شده است. با توجه به رابطه (۵)، به دلیل اینکه در کسر  $\frac{G_C}{M}$  مقدار  $G_C$  نسبت به  $M$  نمی‌تواند زیاد افزایش یابد و اینکه در برخی گروه‌ها  $G_C$  بسیار کم است، این نسبت در ریشه دوم  $M$  ضرب شده است. در این گام، دلیل انتخاب ریشه به جای لگاریتم جهت کم اثر کردن مخرج، این است که ما در بسیاری از آزمایش‌های عددی، ریشه را صحیح‌تر از لگاریتم به دست آوردیم.

$$step\ 6: S \rightarrow \frac{H}{M} * \log_1 M * \left( \frac{G_C}{M} * \sqrt{M} \right) \quad (5)$$

شکل ۱۱ نتیجه اجرای گام ششم را بر روی نتایج حاصل از گام پنجم برای مثال عددی نشان می‌دهد. فهرست  $I_1$  نشان‌دهنده  $G_C$  یا تعداد کاربران با وضعیت  $C$  در هر گروه است. فهرست‌های  $a$  و  $b$  به ترتیب بیانگر اجرای گام ششم بر روی فهرست‌های  $a$  و  $b$  در شکل ۱۰ می‌باشند. در این گام برای هر گروه، مقدار عددی حاصل از گام پنجم در مقدار  $\frac{G_C}{M} * \sqrt{M}$  ضرب شده است و همچنین گروه‌ها بر اساس بیشترین مقدار عددی حاصل از بهبود درصد اشتراک بر اساس  $G_C$  رتبه‌بندی شده‌اند. به‌عنوان مثال گروه  $18m$  در رتبه پنجم فهرست  $a$  و در رتبه سوم فهرست  $b$  قرار دارد. این گروه در دو فهرست  $a$  و  $b$  مقدار  $7,834$  دارد که از محاسبه  $\frac{220}{580} * \log_1 580 * \frac{220}{580}$  به دست آمده است. برای دیگر گروه‌ها نیز روش محاسبه به همین ترتیب است.

(I <sub>1</sub> )		(a)		(b)	
شماره گروه	G <sub>C</sub>	بهبود درصد اشتراک بر اساس G <sub>C</sub>	شماره گروه	بهبود درصد اشتراک بر اساس G <sub>C</sub>	شماره گروه
۳۵am گروه	۴۴۳	۴۷.۳۳	۲m گروه	۴۷.۳۳	۲m گروه
۲m گروه	۲۵۰۰	۳۸.۶۴	۳۵am گروه	۳۸.۶۴	۱۲m گروه
۱۲m گروه	۲۱۸۸	۳۱.۴۹	۱۲m گروه	۳۱.۴۹	۱۸m گروه
۱۸m گروه	۲۲۰	۱۳.۷۸	۱۶am گروه	۱۳.۷۸	۱۸m گروه
۱۶am گروه	۲۱۸۱	۷.۸۳۴	۱۸m گروه	۷.۸۳۴	...
...	...	...	...	...	...
Nam گروه	۱۲۸۳	۰.۰۳۹	Nm گروه	۰.۰۳۹	Nm گروه
...	...	...	...	...	...
Nm گروه	۱۹۷۵	۰.۰۳۳	Nam گروه	۰.۰۳۳	Nm گروه

شکل ۱۱. نتایج حاصل از اجرای گام ششم بر روی مثال عددی

**گام ۷:** این گام، دومین کیفیت را به گام ششم بر اساس  $G_A$  می‌افزاید. برای هر گروه، هرچقدر نسبت  $G_A$  به  $M$  کمتر باشد، آن گروه، گروه خوبی است. دلیل این امر این است که هرچقدر  $G_A$  بیشتر شود، گروه دچار مدیریتی چندوجهی می‌شود و کیفیت گروه

عمل می‌نماید. در رابطه (۴)، متغیر  $S$  و کسر  $\frac{H}{M}$  به ترتیب بیانگر نتایج حاصل از گام اول و گام چهارم هستند.

$$step\ 5: S \rightarrow \frac{H}{M} * \log_1 M \quad (4)$$

شکل ۱۰ نتیجه اجرای گام پنجم را بر روی نتایج حاصل از گام چهارم برای مثال عددی نشان می‌دهد. فهرست‌های  $a$  و  $b$  به ترتیب بیانگر اجرای گام پنجم بر روی فهرست‌های  $a$  و  $b$  در شکل ۹ می‌باشند. در این گام برای هر گروه، کسر حاصل از گام چهارم در لگاریتم مخرج آن، ضرب شده است. سپس گروه‌ها بر اساس بیشترین مقدار عددی حاصل از بهبود درصد اشتراک رتبه‌بندی می‌شوند. به‌عنوان مثال، در فهرست  $a$  گروه شماره  $35am$  با مقدار  $\frac{630}{756} * \log_1 756$  و در فهرست  $b$  گروه شماره  $2m$  با مقدار  $\frac{1150}{3600} * \log_1 3600$  دارای بیشترین مقدار (بهبود درصد اشتراک) می‌باشند. برای دیگر گروه‌ها نیز روش محاسبه به همین ترتیب است.

(a)		(b)	
شماره گروه	بهبود درصد اشتراک	شماره گروه	بهبود درصد اشتراک
۳۵am گروه	$\frac{630}{756} * \log_1 756$	۲m گروه	$\frac{1150}{3600} * \log_1 3600$
۲m گروه	$\frac{1150}{3600} * \log_1 3600$	۱۲m گروه	$\frac{1410}{5090} * \log_1 5090$
۱۲m گروه	$\frac{1410}{5090} * \log_1 5090$	۱۸m گروه	$\frac{180}{580} * \log_1 580$
۱۸m گروه	$\frac{180}{580} * \log_1 580$	...	...
۱۶am گروه	$\frac{638}{5200} * \log_1 5200$	Nm گروه	$\frac{1}{3100} * \log_1 3100$
...	...	...	...
Nam گروه	$\frac{1}{2550} * \log_1 2550$	Nm گروه	$\frac{1}{3100} * \log_1 3100$
...	...	...	...
Nm گروه	$\frac{1}{3100} * \log_1 3100$		

شکل ۱۰. نتایج حاصل از اجرای گام پنجم بر روی مثال عددی

### کیفیت‌سنجی گروه‌ها:

در این زیر بخش سه گام جدید به دنبال یکدیگر جهت رتبه‌بندی گروه‌ها بر اساس بیشترین کیفیت، بررسی و محاسبه می‌شوند. این سه کیفیت در ادامه گام پنجم، کیفیت گروه‌ها را بر اساس تعداد جاری زیاد، ادمین کم و سابق کم رتبه‌بندی می‌نماید. هر یک از گام‌های ششم، هفتم و هشتم همانند گام‌های اول تا پنجم می‌توانند روشی جداگانه جهت توصیه کاربر در نظر گرفته شوند. همچنین خروجی این سه گام، فهرستی از گروه‌های رتبه‌بندی شده است که کاربران این گروه‌ها، به ترتیب فهرست تا رسیدن به مجموعه کاربران مورد نظر با یکدیگر ادغام می‌شوند. در ادامه جهت سهولت در محاسبه هر یک از کیفیت‌ها به بیان برخی اختصارات می‌پردازیم.

- $G_A$ : تعداد کاربر با وضعیت  $A$  در گروه
- $G_F$ : تعداد کاربر با وضعیت  $F$  در گروه
- $G_C$ : تعداد کاربر با وضعیت  $C$  در گروه

Archive of SID

$$\text{step} \lambda: S \rightarrow \frac{H}{M} * \log_1 M * \left( \frac{G_C}{M} * \sqrt[M]{M} \right) * \left( \frac{M}{G_A} * \frac{1}{\sqrt[M]{M}} \right) \quad (7)$$

شکل ۱۴ نتیجه اجرای گام هشتم را بر روی نتایج حاصل از گام هفتم برای مثال عددی نشان می‌دهد. فهرست  $I_7$  نشان‌دهنده  $G_F$  یا تعداد کاربران با وضعیت  $F$  در هر گروه است. فهرست‌های  $a$  و  $b$  به ترتیب بیانگر اجرای گام هشتم بر روی فهرست‌های  $a$  و  $b$  در شکل ۱۲ می‌باشند. در این گام برای هر گروه، مقدار عددی حاصل از گام هفتم در مقدار  $\frac{M}{G_F} * \frac{1}{\sqrt[M]{M}}$  ضرب شده است و همچنین گروه‌ها بر اساس بیشترین مقدار عددی حاصل از بهبود درصد اشتراک بر اساس  $G_F$  رتبه‌بندی شده‌اند. به‌عنوان مثال گروه  $18m$  در رتبه چهارم فهرست  $a$  و در رتبه سوم فهرست  $b$  قرار دارد. این گروه در دو فهرست  $a$  و  $b$  مقدار  $4,23$  دارد که از محاسبه  $\frac{1}{\sqrt[58]{58}} * \frac{58}{357} * 62,8$  به دست آمده است. برای دیگر گروه‌ها نیز روش محاسبه به همین ترتیب است. این گام، آخرین مرحله از روش پیشنهادی است و کاربران به وسیله فهرست نهایی گروه‌های رتبه‌بندی شده توصیه می‌شوند. در شکل ۱۴، فهرست‌های  $d$  و  $e$  به ترتیب گروه‌های رتبه‌بندی شده در فهرست‌های  $a$  و  $b$  هستند. جهت دستیابی به مجموعه کاربران نهایی با تعداد مورد نظر می‌توان هدف را دست‌یابی به ۱۰ برابر، ۲۰ برابر کاربران ورودی و یا هر مقدار دلخواه دیگری انتخاب نمود. در این مثال، هدف دستیابی به مجموعه‌ای کاربر با تعداد ۱۰ برابر مجموعه کاربر ورودی می‌باشد. برای دستیابی به این هدف، کاربران گروه‌های فهرست  $d$  یا  $e$ ، به ترتیب تا رسیدن به مجموعه کاربرانی کمتر مساوی ۱۶۰۰۰ با یکدیگر ادغام می‌شوند. این کار را جهت بررسی نتایج هر یک از گام‌ها، می‌توان در انتهای هر گام اجرا نمود و اعضای گروه‌های رتبه‌بندی شده را تا رسیدن به حداکثر ۱۶۰۰۰ کاربر با یکدیگر ادغام نمود.

کاهش می‌یابد. این گام در رابطه (۶) نشان داده شده است. با توجه به رابطه (۶)، در صورتی که این گام را بتوان به گام ششم افزود، باید متغیر  $M$  بر  $G_A$  تقسیم شود. این گام همانند گام ششم، نیازمند کم اثر کردن مخرج است که جهت آمدن در کنار دیگر گام‌ها، در معکوس ریشه  $M$  ضرب شده است.

$$\text{step} \nu: S \rightarrow \frac{H}{M} * \log_1 M * \quad (6)$$

$$\left( \frac{G_C}{M} * \sqrt[M]{M} \right) * \left( \frac{M}{G_A} * \frac{1}{\sqrt[M]{M}} \right)$$

شکل ۱۲ نتیجه اجرای گام هفتم را بر روی نتایج حاصل از گام ششم برای مثال عددی نشان می‌دهد. فهرست  $I_7$  نشان‌دهنده  $G_A$  یا تعداد کاربران با وضعیت  $A$  در هر گروه است. فهرست‌های  $a$  و  $b$  به ترتیب بیانگر اجرای گام هفتم بر روی فهرست‌های  $a$  و  $b$  در شکل ۱۱ می‌باشند. در این گام برای هر گروه، مقدار عددی حاصل از گام ششم در مقدار  $\frac{M}{G_A} * \frac{1}{\sqrt[M]{M}}$  ضرب شده است و همچنین گروه‌ها بر اساس بیشترین مقدار عددی حاصل از بهبود درصد اشتراک بر اساس  $G_A$  رتبه‌بندی شده‌اند. به‌عنوان مثال گروه  $18m$  در رتبه پنجم فهرست  $a$  و در رتبه سوم فهرست  $b$  قرار دارد. این گروه در دو فهرست  $a$  و  $b$  مقدار  $62,8$  دارد که از محاسبه  $\frac{1}{\sqrt[58]{58}} * \frac{58}{3} * 7,834$  به دست آمده است. برای دیگر گروه‌ها نیز روش محاسبه به همین ترتیب است.

$(I_7)$		(a)		(b)		
شماره گروه	$G_A$	بهبود درصد اشتراک	شماره گروه	بهبود درصد اشتراک	شماره گروه	
گروه ۳۵am	۳	بر اساس $G_A$	گروه ۲m	۱۴۱۹,۹	گروه ۲m	۱۴۱۹,۹
گروه ۲m	۲	گروه ۲m	۳۷۴,۴	گروه ۱۲m	۳۷۴,۴	
گروه ۱۲m	۶	گروه ۳۵am	۳۵۴,۱	گروه ۱۸m	۶۲,۸	
گروه ۱۸m	۳	گروه ۱۶am	۱۴۱,۹	...	...	
گروه ۱۶am	۷	گروه ۱۸m	۶۲,۸	...	...	
...	...	...	...	...	...	
گروه Nam	۲	گروه Nam	۰,۸۳۳	گروه Nm	۰,۴۳۴	
...	...	...	...	...	...	
گروه Nm	۵	گروه Nm	۰,۴۳۴			

شکل ۱۲. نتایج حاصل از اجرای گام هفتم بر روی مثال عددی

**گام ۸:** این گام، سومین کیفیت را به گام هفتم بر اساس  $G_F$  می‌افزاید. برای هر گروه، هرچقدر نسبت  $G_F$  به  $M$  کمتر باشد آن گروه، گروه خوبی است. دلیل این بهتر بودن این است که هرچقدر تعداد کاربرانی که گروه را ترک کرده‌اند افزایش یابد، می‌تواند دلیلی بر بی‌کیفیت بودن آن گروه باشد. این گام در رابطه (۷) نشان داده شده است. با توجه به رابطه (۷)، در صورتی که این گام را بتوان به گام هفتم افزود، باید متغیر  $M$  را بر  $G_F$  تقسیم نمود. همچنین این گام نیازمند کم اثر کردن مخرج است که جهت آمدن در کنار دیگر گام‌ها، در معکوس ریشه  $M$  ضرب شده است. این گام، تمامی گام‌های روش پیشنهادی را در بر دارد. شبه کد نهایی روش پیشنهادی در شکل ۱۳ نشان داده شده است.

$(I_7)$		(a)		(b)		
شماره گروه	$G_F$	بهبود درصد اشتراک	شماره گروه	بهبود درصد اشتراک	شماره گروه	
گروه ۳۵am	۳۱۰	بر اساس $G_F$	گروه ۲m	۷۷,۵	گروه ۲m	۷۷,۵
گروه ۲m	۱۰۹۸	گروه ۲m	۹,۲	گروه ۱۲m	۹,۲	
گروه ۱۲m	۲۸۹۶	گروه ۳۵am	۳۱,۴	گروه ۱۸m	۴,۲۳	
گروه ۱۸m	۳۵۷	گروه ۱۲m	۹,۲	گروه ۱۶am	۳,۳۹	
گروه ۱۶am	۳۰۱۲	گروه ۱۸m	۴,۲۳	...	...	
...	...	گروه ۱۶am	۳,۳۹	...	...	
گروه Nam	۱۲۶۵	...	...	...	...	
...	...	گروه Nam	۰,۰۳۳	...	...	
گروه Nm	۱۱۲۰	...	...	...	...	
		گروه Nm	۰,۰۲۱			

(d)	(e)
شماره گروه	شماره گروه
گروه ۲m	گروه ۲m
...	...
گروه Nam	گروه ۱۲m
...	...
گروه Nm	گروه Nm

شکل ۱۴. نتایج حاصل از اجرای گام هشتم بر روی مثال عددی

**Pseudo-code of the proposed method**

```

1. Input:  $U$  (list of all input users)
2. Output:  $P$  (final user list)
3. Variables:
4. var  $u_i$ : each user in  $U$ 
5. var  $G$ : [initially empty] set of all groups inferred from  $U$ 
6. var  $g_k$ : each group in  $G$ 
7. var  $u'_i$ : each user in  $g_k$ 
8. var  $g'_i$ : each group  $u'_i$  is or was a member or admin
9. var  $R$ : [initially empty] migration list
10. var  $g_r$ : each group in  $R$ 
11. var  $h$ : count of  $g_r$  members
12. var  $L$ : [initially empty] set of all groups inferred from  $U$ 
13. var  $g_l$ : each group in  $L$ 
14. var  $u''_i$ : each user in  $g_l$ 
15. var  $g''_i$ : each group  $u''_i$  is or was a member or admin
16. var  $R'$ : [initially empty] admin migration list
17. var  $g'_r$ : each group in  $R'$ 
18. var  $E$ : [initially empty] set to store the union of  $R$  and  $R'$ 
19. var  $g_e$ : each group in  $E$ 
20. var  $h'$ : count of  $g_e$  members
21. var  $T$ : a temporary list
22. Pseudo-code:
23. for each user  $u_i$  in  $U$ :
24.      $G.append(u_i \text{ groups})$ ; // all groups associated with  $u_i$  by any relation is appended to  $G$ 
25. for  $g_k$  in  $G$ :
26.     for  $u'_i$  in  $g_k$ :
27.         if status of  $u'_i$  is  $F$ : // if  $u'_i$  is a former member
28.             for  $g'_i$  in  $u'_i$  groups:
29.                 if status of  $u'_i$  in  $g'_i$  is  $C$  or  $A$ : // if  $u'_i$  is a current member or admin
30.                      $R.append(g'_i)$ ;
31. if status of  $u_i$  in  $U$  is  $A$ : // if  $u_i$  is an admin (There is at least one user with admin status Among incoming users)
32.      $L.append(u_i \text{ groups})$ ; // all groups associated with  $u_i$  by any relation is appended to  $L$ 
33.     for  $g_l$  in  $L$ :
34.         for  $u''_i$  in  $g_l$ :
35.             if not status of  $u''_i$  is  $F$ : // if  $u''_i$  is a former member
36.                 for  $g''_i$  in  $u''_i$  groups:
37.                     if status of  $u''_i$  in  $g''_i$  is  $C$  or  $A$ : // if  $u''_i$  is a current member or admin
38.                          $R'.append(g''_i)$ ;
39.     for each  $g_r$  in  $R$ :
40.          $E.append(g_r)$ 
41.     for each  $g'_r$  in  $R'$ :
42.         if not  $E.contains(g'_r)$  then  $E.append(g'_r)$  // remove common groups in two sets  $R$  and  $R'$ 
43.     for each group  $g_e$  in  $E$ :
44.         set  $h'$  = count of  $g_e$  members;
45.     Set  $T$  = a list of users of  $g_e$ s with highest  $\frac{h'}{M} * \log_{10} M * \left(\frac{G_C}{M} * \sqrt[2]{M}\right) * \left(\frac{M}{G_A} * \frac{1}{\sqrt{M}}\right) * \left(\frac{M}{G_F} * \frac{1}{\sqrt[3]{M}}\right)$ ;
        //  $M$  is the number of group members.
        //  $G_C$  is the number of users with current status in the group.
        //  $G_A$  is the number of users with Admin status in the group.
        //  $G_F$  is the number of users with Former status in the group.
46.     while  $P.length() <$  desired number of output users:
47.          $P.append(T.pop())$ ; // one of users from  $T$  is appended to  $P$  and removed from  $T$ 
48.     return  $P$ ;
49. else if status of  $u_i$  in  $U$  is not  $A$ : // if  $u_i$  isn't an admin (There are no users with admin status among incoming users)
50.     for each group  $g_r$  in  $R$ :
51.         set  $h$  = count of  $g_r$  members;
52.     Set  $T$  = a list of users of  $g_r$ s with highest  $\frac{h}{M} * \log_{10} M * \left(\frac{G_C}{M} * \sqrt[2]{M}\right) * \left(\frac{M}{G_A} * \frac{1}{\sqrt{M}}\right) * \left(\frac{M}{G_F} * \frac{1}{\sqrt[3]{M}}\right)$ ;
        //  $M$  is the number of group members.
        //  $G_C$  is the number of users with current status in the group.
        //  $G_A$  is the number of users with Admin status in the group.
        //  $G_F$  is the number of users with Former status in the group.
53.     while  $P.length() <$  desired number of output users:
54.          $P.append(T.pop())$ ; // one of users from  $T$  is appended to  $P$  and removed from  $T$ 
55.     return  $P$ ;

```

شکل ۱۳. شبه‌کد نهایی روش پیشنهادی

**۴. نتایج تجربی**

در این بخش ابتدا به بیان داده‌های جمع‌آوری شده و روش ارزیابی پرداخته شده است و سپس نتایج آزمایش‌ها بررسی و تحلیل می‌شود.

**۴.۱ داده‌های جمع‌آوری شده و محیط پیاده‌سازی**

داده‌های استفاده شده در این پژوهش، داده‌هایی واقعی از پیام‌رسان تلگرام است که توسط سامانه ایده‌کاو به دست آمده‌اند. ایده‌کاو دارای

خوشه‌ای از خزنده‌ها همانند خزش‌گرهای وب است. خزش‌گرهای ایده‌کاو با ورود به لینک‌های جدید، عضو گروه‌ها می‌شوند. هریک از خزش‌گرها به صورت لحظه‌ای پس از عضویت در هر گروه تمامی اطلاعات آن را از جمله کاربر جدید و پیام‌های گروه دریافت می‌کند. این اطلاعات توسط خزش‌گر به بخش مدیریت اطلاعات ارسال می‌شود. در بخش مدیریت، داده‌هایی که توسط خزش‌گر به دست آمده است، بررسی می‌شود و گروه‌های هرزنامه جهت ترک و

## Archive of SID

خبره انتخاب کرده‌ایم. اطلاعات نام و تعداد اعضای ۱۰۰ گروه منتخب در جدول ۳ نشان داده شده است. میانگین تعداد اعضای گروه‌های منتخب ۱۳،۸۹۳ است. دلیل انتخاب گروه‌هایی تخصصی جهت ارزیابی این است که تمامی اعضای این گروه‌ها، کاربرانی هستند که واقعاً به موضوع گروه علاقه‌مندند و در گروه پیام‌هایی نامرتب با موضوع گروه ارسال نمی‌نمایند.

ارزیابی میزان پیش‌بینی و خطای گام‌های روش پیشنهادی، به‌وسیله هر یک از ۱۰۰ گروه تخصصی به‌صورت جداگانه انجام شده است. برای ارزیابی گام‌ها توسط هر گروه تخصصی، اعضای گروه به دو بخش ۸۰ درصد جهت آموزش و ۲۰ درصد جهت بررسی میزان پیش‌بینی و خطا، تقسیم شده‌اند. به‌طور کلی روش ارزیابی بدین‌صورت است که برای هر گروه تخصصی، ابتدا ۸۰ درصد کاربران گروه، به‌ورودی هر یک از گام‌های روش پیشنهادی داده می‌شود، سپس در هر گام به‌صورت جداگانه گروه‌هایی رتبه‌بندی شده به دست می‌آید. کاربران گروه‌ها، به ترتیب فهرست تا رسیدن به مجموعه‌ای کاربر با تعداد موردنظر با یکدیگر ادغام می‌شوند. تعداد کاربر موردنظر، در واقع خط پایانی است که تعیین‌کننده تعداد کاربر موجود در فهرست کاربران نهایی پیشنهاد شده توسط توصیه‌گر است. به‌عنوان مثال تعداد کاربر موردنظر را می‌توان ۱۰، ۲۰، ۳۰ برابر مجموعه کاربر ورودی و یا هر مقدار دلخواهی تعیین نمود. در این مقاله تعداد کاربر موردنظر، ۱۰ برابر مجموعه کاربر ورودی در نظر گرفته شده است. در نهایت، برای هر گام پس از دستیابی به مجموعه کاربران نهایی، میزان پیش‌بینی و خطا با استفاده از ۲۰ درصد کاربر نگهداری شده، مورد ارزیابی قرار می‌گیرد.

لینک‌های جدید جهت عضویت به خزش‌گر ارسال می‌شود. اطلاعات خزش‌گر در خوشه‌ای از شاخص‌های معکوس<sup>۱</sup> ذخیره می‌شود. این کار باعث جستجوی سریع کلمات کلیدی می‌شود. در نهایت، سرویس جستجوی ایده‌کاو که بر اساس جست‌وجوگر لوسین<sup>۲</sup> است، به‌صورت مستقیم از شاخص‌ها پرس‌وجو می‌کند [۲۸]. اطلاعات مورد استفاده در این مقاله شامل بیش از ۹۰۰ هزار سوپرگروه و ۱۲۰ میلیون کاربر است. ما در این مقاله از یک گراف که کاربران را بر اساس گروه مشترک به یکدیگر متصل می‌نماید، بهره برده‌ایم. اطلاعات دقیق داده‌ها در جدول ۲ نشان داده شده است. با توجه به جدول ۲ میانگین تعداد اعضای کل گروه‌ها ۱۱۳۶ است، این نکته بیانگر این است که بسیاری از گروه‌های تلگرام تعداد اعضایشان در بازه هزار تا دوهزار است. پیاده‌سازی و ارزیابی گام‌های روش پیشنهادی در یک سیستم ۶۴ بیتی core-i۷ با ۸ گیگابایت رم<sup>۳</sup> انجام شده است. گراف عضویت کاربران به‌صورت جدولی در پایگاه داده MySQL پیاده‌سازی شده است. همچنین، از زبان برنامه‌نویسی پایتون<sup>۴</sup> برای تجزیه و تحلیل داده‌ها و پیاده‌سازی الگوریتم‌ها استفاده شده است.

جدول ۲. آمار داده‌های تلگرام

تعداد کاربران	تعداد سوپرگروه‌ها	میانگین تعداد اعضای گروه‌ها
۱۲۵۲۶۹۵۲۲	۹۲۰۸۱۰	۱۱۳۶

## ۲.۴ روش ارزیابی

در این مقاله ارزیابی روش پیشنهادی به‌وسیله گروه‌های تخصصی در تلگرام انجام شده است. گروه‌های تخصصی گروه‌هایی با کیفیت در تلگرام هستند که اعضا در آنان، پیام‌هایی حول موضوع گروه ارسال می‌نمایند. ما در این مقاله، ۱۰۰ گروه تخصصی و با کیفیت را با بررسی

جدول ۳. مشخصات ۱۰۰ گروه مورد ارزیابی

شماره گروه	نام گروه	تعداد اعضا	شماره گروه	نام گروه	تعداد اعضا
۱	انجمن تخصصی گردو ایران	۱۶۶۵۸	۵۱	دامپزشکی تخصصی اسب	۱۰۴۴۲
۲	اعلام بار آذربادگان سراسری و بین‌المللی	۱۳۶۹۰	۵۲	ابر گروه سنگ فروشان	۱۰۷۴۱
۳	اعلام بار ستاره بار	۱۲۲۴۰	۵۳	اعلام بار جامع کفی بندرعباس	۱۷۳۰۳
۴	گروه کشاورزی خاورمیانه	۱۰۰۸۲	۵۴	گروه گیاهی و تخصصی "طب الرضا"	۱۰۲۴۷
۵	گروه مهندسان، پیمان‌کاران راه و ساختمان	۱۱۱۵۴	۵۵	گروه کوهنوردی اسطوره ایران زمین	۱۵۷۲۹
۶	گروه گفتمان بورسی	۱۴۳۴۰	۵۶	گروه پرسش و پاسخ گل و گیاه نگار	۱۱۱۶۲
۷	آزمون نظام مهندسی	۱۴۰۹۳	۵۷	گروه انجمن شهرسازی ایران	۱۱۸۶۷
۸	گروه صنعت کیفیت و کفش خاص	۱۵۶۵۶	۵۸	گروه تخصصی آبزیان آکواریومی	۱۴۶۶۵
۹	خریدوفروش خانه در تورنتو	۱۵۴۳۸	۵۹	آکادمی رنگ و میکاپ تخصصی	۱۰۰۴۰
۱۰	سلامت و پزشکی (توس طب)	۱۰۰۱۸	۶۰	گروه تخصصی املاک شرق گیلان	۱۳۹۳۸
۱۱	فروشگاه تخصصی لباس و کفش کودک	۱۰۶۷۳	۶۱	پرسش و پاسخ داروخانه سلامت	۱۴۶۸۴
۱۲	آکادمی فوق تخصصی آرایشی	۱۶۰۳۸	۶۲	پرسش و پاسخ تندرستی با طب سنتی	۱۲۷۰۷

<sup>۲</sup> Ram  
<sup>۴</sup> Python

<sup>۱</sup> Inverted index  
<sup>۲</sup> Lucene

شماره گروه	نام گروه	تعداد اعضا	شماره گروه	نام گروه	تعداد اعضا
۱۳	گروه تخصصی نماز	۱۴۸۷۸	۶۳	گروه حل فوری مشکلات کامپیوتری	۱۰۷۹۸
۱۴	آموزش تخصصی شنا ویژه بانوان	۱۰۴۱۷	۶۴	چالش‌های مهاجرت و پناهندگی آمریکا	۱۰۶۶۶
۱۵	آکادمی تخصصی بدن‌سازی آدرنالین	۱۶۸۶۸	۶۵	خریدوفروش لندکروز و قطعات	۱۰۱۸۳
۱۶	اعلام بار ملی ایران خراسان	۱۱۵۲۳	۶۶	سوپر گروه روان‌شناسان	۱۴۱۸۴
۱۷	بزرگ‌ترین گروه کناف و فن‌آوری ساختمان	۱۷۹۶۲	۶۷	اعلام بارهای صادراتی	۱۱۴۶۵
۱۸	گروه انواع برند و قطعات خودرو GSR	۱۸۲۸۲	۶۸	بررسی مهاجرت یونان	۱۳۹۱۰
۱۹	شبکه بزرگ اعلام بار آریا	۱۸۷۱۷	۶۹	پرسش و پاسخ حقوق مهندسی ایران	۱۵۸۰۸
۲۰	خریدوفروش تجهیزات کوهنوردی، سنگ‌نوردی	۱۵۰۸۴	۷۰	گروه تخصصی همفکری رولت، بت، پوکر	۱۴۰۸۰
۲۱	گروه بهترین رمان‌ها (نقد و بررسی رمان)	۱۸۸۳۰	۷۱	چالش‌ها و روش‌های مهاجرت به آلمان	۱۲۴۴۵
۲۲	مشاوره‌ی تخصصی پوست/امو	۱۹۶۲۱	۷۲	کتابخانه تخصصی روانشناسی	۱۱۸۴۰
۲۳	کلینیک مشاوره ورزشی (فیت استایل)	۱۹۶۰۱	۷۳	برق و صنعت ساختمان	۱۱۳۴۶
۲۴	آموزش رایگان بورس ویژه فرهنگیان	۱۹۲۴۴	۷۴	اعلام بار آنلاین (اورجینال) اصفهان	۱۲۵۰۵
۲۵	تخصصی تأسیسات صنعتی و فوق صنعتی	۱۹۳۳۲	۷۵	بزرگ‌ترین گروه طب سنتی و مدرن	۱۵۸۹۳
۲۶	پیوند مهندسین ایرانی - ترکیه‌ای در استانبول	۱۵۰۱۱	۷۶	گروه صنعت کفش ایران- تبریز	۱۵۸۴۰
۲۷	انجمن نرم‌افزاری اندروید	۱۲۷۲۹	۷۷	انجمن سلامتی و طب سنتی	۱۰۲۲۳
۲۸	گروه تحلیل رایگان طلا، سکه، دلار	۱۰۸۲۶	۷۸	آموزش زبان ترکی استانبولی	۱۳۲۹۳
۲۹	اطلاعات جامع آلمان	۱۰۲۲۳	۷۹	گروه تخصصی کفش در ایران	۱۴۶۱۵
۳۰	اعلام بار ده هزار نفری استان کرمانشاه	۱۰۳۰۱	۸۰	نمایشگاه قطعات خودرو	۱۴۶۰۷
۳۱	انجمن زعفران کاران خراسان بزرگ	۱۴۷۳۵	۸۱	سوپر گروه رانندگان و کامیون‌داران	۱۴۵۹۴
۳۲	گروه تجارت طیور ایران (محبی)	۱۴۶۰۷	۸۲	اعلام بار خاور (مسقف داران) سراسری ایران	۱۰۷۷۹
۳۳	کانون اساتید حق‌التدریس دانشگاه‌ها و مؤسسات	۱۰۳۶۰	۸۳	تیم تخصصی تناسب اندام آکالااین	۱۳۱۵۱
۳۴	بزرگ‌ترین پارک مشاوره‌ای تخصصی کودکان	۱۵۳۶۱	۸۴	گروه تخصصی نمایش	۱۳۵۱۵
۳۵	Temporary Resident Visas (Canada)	۱۶۷۷۹	۸۵	طیب طب ایرانی	۱۰۰۴۵
۳۶	آموزش تخصصی میکروبلیدینگ	۱۵۱۳۸	۸۶	مشاور طب سنتی	۱۱۱۴۲
۳۷	کمپین شهرسازان نوین	۱۵۷۰۷	۸۷	گروه روانشناسی و پرسش و پاسخ	۱۲۰۹۴
۳۸	گروه پرسش و پاسخ کامپیوتری	۱۲۷۹۸	۸۸	گروه تحلیلی و آموزش بورس آریا	۱۲۱۲۹
۳۹	سؤالات تخصصی فقط پوست	۱۱۵۵۴	۸۹	گروه نقد و تحلیل ارزهای دیجیتال	۱۵۷۸۹
۴۰	گروه سنگین خودرو	۱۱۲۴۹	۹۰	پایانه اعلام بار سراسری نیسان. سه تن.	۱۰۷۸۲
۴۱	اعلام بار کمپرسی و اطاق‌دار بدون گپ و گف	۱۱۸۰۲	۹۱	صنعت نوین ساختمان	۱۶۴۷۴
۴۲	اعلام بار خاوری (سراسری) شمال مازندران	۱۶۲۷۲	۹۲	نورپردازی و عکاسی پرتره	۱۱۱۷۵
۴۳	پایگاه تخصصی ورزشی نشاط	۱۴۵۶۶	۹۳	سوپر گروه برق الکترونیک	۱۶۳۸۵
۴۴	سامانه کاریاب استان یزد	۱۲۹۵۰	۹۴	گروه بررسی تخصصی سلاح‌های بادی	۱۶۹۷۶
۴۵	مشاور تخصصی داروهای گیاهی ارگانیک	۱۳۸۷۳	۹۵	مهندسی آثار باستانی	۱۲۰۵۹
۴۶	گروه تحلیل سهم بورس	۱۳۰۳۸	۹۶	سوپر گروه عمده‌فروشان تهران	۱۵۱۱۳
۴۷	صنف ابزار و لوازم ساختمانی	۱۳۶۳۷	۹۷	اعلام بار پیمان پیام میبد، یزد	۱۹۳۹۷
۴۸	گروه اعلام بار سراسری ایرانیان	۱۳۲۳۹	۹۸	اندیشکده تخصصی رسانه	۱۵۴۶۰
۴۹	گروه مهندسی معماری مهان دیزاین	۱۸۴۵۹	۹۹	گروه تخصصی پکیج و موتورخانه	۱۶۹۳۷
۵۰	پرسش و پاسخ الوبورس	۱۸۲۰۴	۱۰۰	پارچه سرای کریمی	۱۴۲۶۹

## Archive of SID

طبق جدول ۴، این گروه در گام اول، توانسته است با افزودن کاربران ۳ گروه، حداکثر به ۱۰۲۸۴۶ کاربر برسد. برای این گروه در گام اول، خطای  $RMSE$ ، ۰٫۹۱۶، و خطای  $MAE$ ، ۰٫۸۱۴، است. شکل ۱۵ نتایج میانگین دو خطای  $RMSE$  و  $MAE$  را بر روی ۱۰۰ گروه مورد ارزیابی بر اساس نتایج حاصل از جدول ۴ و ۵ نشان می‌دهد. با توجه به شکل ۱۵، گام چهارم در مقایسه با دیگر گام‌ها، خطای  $RMSE$  و  $MAE$  کمتری دارد. این گام توانسته است نسبت به گام اول که روشی با نام گروه‌های مشابه است، خطای  $RMSE$  را به ۰٫۰۸۱ و خطای  $MAE$  را به ۰٫۱۲۸ کاهش دهد. با توجه به کاهش خطاها در گام چهارم، این نکته مهم حاصل می‌شود که رتبه‌بندی گروه‌های کاربران ورودی بر اساس درصد اشتراک موجب بالا آمدن رتبه گروه‌های با تعداد عضو کم می‌شود و در نتیجه گروه‌های به‌دست‌آمده، پیش‌بینی درستی از کاربرانی مشابه با کاربران ورودی خواهند داشت. علاوه بر گام چهارم، به ترتیب گام‌های پنجم، ششم، هفتم، دوم و سوم نیز نسبت به گام اول خطای  $RMSE$  را به ۰٫۰۰۶، ۰٫۰۰۲، ۰٫۰۰۹، ۰٫۰۰۷ و خطای  $MAE$  را به ۰٫۱۲۲، ۰٫۰۹۹، ۰٫۰۰۴، ۰٫۰۱۸ و ۰٫۰۱۲ کاهش داده‌اند. در بین هشت گام مورد بررسی، گام اول در جایگاه هفتم از میزان خطا و پیش‌بینی درست کاربران قرار دارد؛ گام هشتم تنها گام از روش پیشنهادی است که در مقایسه با گام اول خطای  $RMSE$  را به ۰٫۰۰۸ و خطای  $MAE$  را به ۰٫۱۵ افزایش داده است. البته گام هشتم در مقایسه با گام اول کیفیت گروه‌ها را نیز تضمین می‌نماید. با توجه به افزایش خطای گام هشتم، این نکته مهم حاصل می‌شود که در صورتی که گروه‌ها بر اساس نسبت نزولی سابق به اعضا رتبه‌بندی شوند، میزان پیش‌بینی کاهش و خطا افزایش می‌یابد. به‌طور کلی از گام اول تا چهارم میزان خطاها کاهش و از گام چهارم تا هشتم، میزان خطاها افزایش یافته است. با توجه به اینکه از گام ششم تا گام هشتم کیفیت افزایش می‌یابد، می‌توان گفت در گام ششم تا هشتم، خطا با کیفیت رابطه مستقیم دارد و با افزایش کیفیت، میزان خطا افزایش می‌یابد.

### ۴٫۴ بحث و تجزیه و تحلیل نتایج

در این زیر بخش، نتایج جامع به‌دست‌آمده در جدول ۴ و ۵ تجزیه و تحلیل می‌شود. طبق نتایج حاصل از جدول ۴ و ۵، در گروه‌های شماره ۵، ۶، ۱۴، ۴۴، ۴۷، ۵۴، ۶۶، ۷۳ و ۸۰ گام اول و دوم با یکدیگر برابر شده است. دلیل برابری این دو گام این است که هنگامی که فهرست گروه‌های حاصل از گام اول به‌وسیله روش مهاجرت با گروه‌های دیگری جایگزین می‌شوند، برخی از گروه‌های جدید در گروه‌های حاصل از گام اول وجود دارند و همین دلیل باعث می‌شود تا پس از بررسی بیشترین اشتراک با مجموعه کاربران ورودی، فهرستی از همان گروه‌ها تشکیل شود.

در این مقاله برای بررسی میزان کارایی روش پیشنهادی تنها از معیار خطا استفاده شده است. استفاده از معیارهای دقت<sup>۱</sup>، صحت<sup>۲</sup> و فراخوانی<sup>۳</sup> برای این روش امکان‌پذیر نیست؛ دلیل آن هم این است که ما طبق آزمایش‌هایی مشاهده کردیم به‌عنوان مثال برای بررسی معیار دقت با افزایش تعداد گروه‌ها و تعداد کاربران، دقت کاهش می‌یابد. دلیل کاهش دقت این است که تعداد کاربر درست پیش‌بینی شده وقتی تقسیم بر کل تعداد (درست و نادرست) کاربران پیش‌بینی شده می‌شود، مقدار صورت نمی‌تواند به‌اندازه مخرج افزایش یابد و رفته‌رفته باعث کاهش دقت می‌شود.

معیارهای استفاده شده در این مقاله جهت بررسی مدل پیشنهادی، معیار خطای جذر میانگین مربع‌ها و خطای میانگین قدرمطلق‌ها است. خطای جذر میانگین مربع‌ها و میانگین قدرمطلق‌ها روش‌هایی جهت بررسی پیش‌بینی مدل هستند که با بیان میزان خطا، اثربخشی آن را نشان می‌دهند. خطای جذر میانگین مربع‌ها و خطای میانگین قدرمطلق‌ها به ترتیب طبق رابطه (۸) و (۹) محاسبه می‌شود.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}} \quad (8)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Predicted_i - Actual_i| \quad (9)$$

در رابطه (۸) و (۹)،  $Predicted_i$  مجموعه پیش‌بینی و شامل مجموعه‌ای از صفرها و یک‌ها است. صفر بیان‌گر این است که پیش‌بینی مدل درست بوده است و یک بیانگر عکس آن است.  $Actual_i$  مجموعه واقعی و شامل مجموعه‌ای از صفرها است که بیانگر مجموعه کاربران نگه‌داری شده (آزمون) است.  $N$  مجموعه خطاهای کاربران پیشنهادی است. برای هر یک از ۱۰۰ گروه مورد ارزیابی، با توجه به نتایج حاصل از اجرای گام‌ها، مجموعه پیش‌بینی و مجموعه واقعی به‌صورت جداگانه ایجاد و سپس هر یک از دو خطا محاسبه می‌شود.

### ۳٫۴ نتایج آزمایش‌ها

هر یک از ۱۰۰ گروه به‌صورت جداگانه آزمایش شده است. بخشی از نتایج ارزیابی گام‌های روش پیشنهادی بر روی ۱۰۰ گروه، در جدول ۴ و ۵ نشان داده شده است. به‌دلیل حجم بالای نتایج ارزیابی گام‌ها، تنها شمایی از نتایج در جدول ۴ و ۵ نشان داده شده است. با توجه به جدول ۴ و ۵ برای هر گام از روش پیشنهادی، چهار ستون در نظر گرفته شده است. ستون اول بیانگر تعداد گروه است. ستون دوم، تعداد کاربران به‌دست‌آمده بر اساس کاربران گروه‌ها را نشان می‌دهد. ستون‌های سوم و چهارم به ترتیب بیانگر خطای  $RMSE$  و  $MAE$  هستند. به‌عنوان مثال گروه شماره ۶ از جدول ۴ را در نظر بگیرید. طبق جدول ۳، ۸۰ درصد تعداد اعضای این گروه، ۱۱۴۷۲ است. هدف ارزیابی در این گروه، رسیدن به تعداد ۱۱۴۷۲۰ کاربر است.

<sup>۱</sup> Recall

<sup>۱</sup> Precision

<sup>۲</sup> Accuracy

جدول ۴. شمایی از نتایج ارزیابی گام‌های ۱ تا ۴ بر روی ۱۰۰ گروه (NG: تعداد گروه و NU: تعداد کاربر)

گام ۴				گام ۳				گام ۲				گام ۱ [۱]				شماره گروه
MAE	RMSE	NU	NG	MAE	RMSE	NU	NG	MAE	RMSE	NU	NG	MAE	RMSE	NU	NG	
۰,۴۰۸	۰,۶۳۹	۱۲۰۰۴۰	۵۵	۰,۵۰۶	۰,۷۱۱	۱۲۵۶۹۲	۴	۰,۵۰۶	۰,۷۱۱	۱۲۵۶۹۲	۴	۰,۸۱	۰,۹	۱۲۲۲۵۲	۳	۱
۰,۷۳	۰,۸۵۴	۱۰۲۵۴۷	۲۲	۰,۸۷	۰,۹۳۳	۸۶۴۳۱	۳	۰,۸۷	۰,۹۳۳	۸۶۴۳۱	۳	۰,۹۶۷	۰,۹۸۳	۱۹۶۰۵۶	۱	۲
۰,۶۸۳	۰,۸۲۶	۹۳۸۶۳	۲۷	۰,۷۱۸	۰,۸۴۷	۹۱۸۱۰	۴	۰,۷۷۱	۰,۸۷۸	۸۵۹۶۰	۲	۰,۶۸۳	۰,۸۲۶	۱۹۶۰۵۶	۱	۳
۰,۶۶	۰,۸۱۲	۷۶۴۰۰	۴۰	۰,۸۵	۰,۹۲۲	۷۰۰۳۷	۳	۰,۸۵	۰,۹۲۲	۷۰۰۳۷	۳	۰,۸۹۲	۰,۹۴۴	۵۵۸۱۷	۱	۴
۰,۶۷۹	۰,۸۲۴	۸۱۸۵۵	۱۷	۰,۹۱۲	۰,۹۵۵	۸۲۶۷۱	۱	۰,۹۱۲	۰,۹۵۵	۸۲۶۷۱	۱	۰,۹۱۲	۰,۹۵۵	۸۲۶۷۱	۱	۵
۰,۷۷	۰,۸۷۸	۹۸۵۰۷	۳۵	۰,۸۰۶	۰,۸۹۸	۱۱۳۴۲۴	۹	۰,۸۴	۰,۹۱۶	۱۰۲۸۴۶	۳	۰,۸۴	۰,۹۱۶	۱۰۲۸۴۶	۳	۶
۰,۷۰۱	۰,۸۳۷	۱۱۱۴۷۲	۲۳	۰,۷۴۶	۰,۸۶۴	۱۰۴۵۴۰	۷	۰,۷۴۲	۰,۸۶۱	۱۱۰۹۸۲	۷	۰,۷۰۹	۰,۸۴۲	۱۱۰۷۵۷	۷	۷
۰,۵۸۵	۰,۷۶۴	۱۲۴۸۸۱	۴۰	۰,۷۳۴	۰,۸۵۷	۱۰۴۷۲۷	۵	۰,۶۳۳	۰,۷۹۶	۶۰۰۸۹	۳	۰,۷۶۹	۰,۸۷۶	۱۰۸۶۶۲	۲	۸
۰,۳۸۵	۰,۶۲۱	۱۱۹۶۸۹	۳۷	۰,۴۴۴	۰,۶۶۶	۱۱۸۹۸۱	۱۴	۰,۴۵	۰,۶۷	۱۲۰۸۹۲	۹	۰,۴۶۳	۰,۶۸	۱۲۲۵۶۸	۱۰	۹
۰,۱۶	۰,۴	۶۴۵۴۴	۳۰	۰,۶۷۸	۰,۸۲۳	۳۹۲۸۳	۲	۰,۹۸۹	۰,۹۹۴	۴۲۹۲۴	۴	۰,۹۹۸	۰,۹۹۹	۲۵۸۰۴	۱	۱۰
...																
۰,۶۲۲	۰,۷۸۸	۱۰۶۶۴۸	۶۴	۰,۸۴	۰,۹۱۶	۱۰۴۷۰۲	۲	۰,۷۴۲	۰,۸۶۱	۷۴۸۹۸	۳	۰,۹۹۷	۰,۹۹۸	۱۶۱۹۶۱	۱	۱۰۰
۰,۶۴۹	۰,۷۹۳	۱۰۵۳۲۲,۲	۴۵,۸	۰,۷۶۵	۰,۸۶۷	۹۱۵۶۲,۹	۵,۰۵	۰,۷۵۹	۰,۸۶۵	۹۰۲۶۸,۱	۴,۱	۰,۷۷۷	۰,۸۷۴	۹۹۰۵۰,۲	۳,۱	میانگین

جدول ۵. شمایی از نتایج ارزیابی گام‌های ۵ تا ۸ بر روی ۱۰۰ گروه (NG: تعداد گروه و NU: تعداد کاربر)

گام ۸				گام ۷				گام ۶				گام ۵				شماره گروه
MAE	RMSE	NU	NG	MAE	RMSE	NU	NG	MAE	RMSE	NU	NG	MAE	RMSE	NU	NG	
۰,۹۳۸	۰,۹۶۸	۱۲۹۳۶۳	۱۱	۰,۵۱۷	۰,۷۱۹	۱۲۵۴۹۸	۱۳	۰,۴۷	۰,۶۸۵	۱۳۳۱۱۱	۲۷	۰,۳۹۸	۰,۶۳۱	۱۳۱۲۷۳	۴۸	۱
۰,۹۲۱	۰,۹۵۹	۱۰۲۱۰۹	۸	۰,۷۴۱	۰,۸۶۱	۸۴۳۹۳	۱۲	۰,۷۳۶	۰,۸۵۸	۱۰۱۲۵۸	۱۳	۰,۷۲۸	۰,۸۵۳	۱۰۸۳۲۱	۱۹	۲
۰,۹۰۲	۰,۹۵	۹۴۵۴۷	۶	۰,۶۹۴	۰,۸۳۳	۹۱۸۲۵	۸	۰,۷۱۶	۰,۸۴۶	۸۲۱۶۴	۱۳	۰,۶۸۹	۰,۸۳	۸۷۴۱۱	۱۶	۳
۰,۹۹۷	۰,۹۹۸	۷۵۲۲۶	۴	۰,۸۵۵	۰,۹۲۴	۵۰۰۱۳	۶	۰,۷۸۵	۰,۸۸۶	۷۰۰۷۵	۷	۰,۶۸۱	۰,۸۲۵	۶۸۴۸۵	۲۴	۴
۰,۹۵۹	۰,۹۷۹	۷۰۸۷۷	۶	۰,۹۰۹	۰,۹۵۳	۱۸۷۶۸	۱	۰,۸۳۱	۰,۹۱۱	۸۳۹۰۱	۷	۰,۶۶۸	۰,۸۱۷	۸۵۸۱۸	۱۱	۵
۰,۹۹۷	۰,۹۹۸	۱۰۷۴۵۲	۵	۰,۷۷۲	۰,۸۷۸	۱۱۱۸۰۲	۷	۰,۷۷۸	۰,۸۸۲	۱۱۳۹۵۸	۱۲	۰,۷۵۸	۰,۸۷۱	۱۱۳۵۶۱	۲۱	۶
۰,۹۹۸	۰,۹۹۹	۱۰۵۲۲۶	۶	۰,۷۸۷	۰,۸۸۷	۱۰۶۴۵۵	۱۰	۰,۷۱۵	۰,۸۴۶	۱۱۰۶۳۳	۱۳	۰,۷۰۹	۰,۸۴۲	۱۰۷۴۰۸	۱۷	۷
۰,۸۴۶	۰,۹۱۹	۱۱۵۵۵۶	۹	۰,۶۰۴	۰,۷۷۷	۱۲۴۳۱۲	۸	۰,۵۹	۰,۷۶۸	۱۲۱۴۷۰	۲۳	۰,۵۸۵	۰,۷۶۵	۱۱۷۸۰۶	۳۱	۸
۰,۷	۰,۸۳۶	۱۲۲۱۳۱	۲۲	۰,۴۴۴	۰,۶۶۷	۱۱۹۴۲۳	۱۷	۰,۳۹۷	۰,۶۳	۱۱۶۴۴۴	۲۰	۰,۳۸۶	۰,۶۲۱	۱۱۷۵۰۸	۳۴	۹
۰,۷۳	۰,۸۵۴	۷۷۳۰۲	۶	۰,۱۶۳	۰,۴۰۳	۶۴۹۷۳	۱۳	۰,۱۵۸	۰,۳۹۸	۷۵۷۲۸	۲۴	۰,۱۶	۰,۴	۷۹۹۹۱	۲۹	۱۰
...																
۰,۹۵۳	۰,۹۷۶	۱۱۰۰۰۳	۱۱	۰,۶۶۶	۰,۸۱۶	۸۷۷۳۹	۷	۰,۶۱۹	۰,۷۸۷	۸۷۵۰۱	۸	۰,۶۰۷	۰,۷۷۹	۱۰۴۵۶۱	۴۶	۱۰۰
۰,۹۲۸	۰,۹۶۱	۱۰۵۰۷۳,۰۶	۸,۹	۰,۷۳۳	۰,۸۴۷	۹۶۷۲۵,۹	۸,۳	۰,۶۷۸	۰,۸۱۲	۱۰۱۹۳۶,۸	۱۵,۱	۰,۶۵۵	۰,۷۹۷	۱۰۲۴۲۰,۲	۳۱,۳	میانگین

در گروه‌های شماره ۵، ۴۷ و ۷۳ گام اول تا سوم با یکدیگر برابر شده‌اند. با توجه به برابر شدن گام اول و دوم که در پاراگراف قبل بیان شد، دلیل آنکه این دو گام با گام سوم برابر شده‌اند، این است که در این گروه‌ها، هیچ کدام از کاربران ورودی دارای وضعیت ادمین نیستند. همچنین برابری گام دوم و سوم در گروه‌های ۱، ۲، ۴، ۵، ۱۳، ۱۶، ۱۹، ۲۰، ۲۳، ۲۹، ۳۴، ۳۷، ۴۱، ۴۵، ۴۷، ۴۸، ۵۱، ۵۵، ۶۰، ۶۱، ۶۷، ۷۰، ۷۳، ۷۴، ۷۸، ۷۹، ۸۱، ۸۲، ۸۸، ۸۹، ۹۱، ۹۸ و ۹۹ نیز مشاهده شده است.

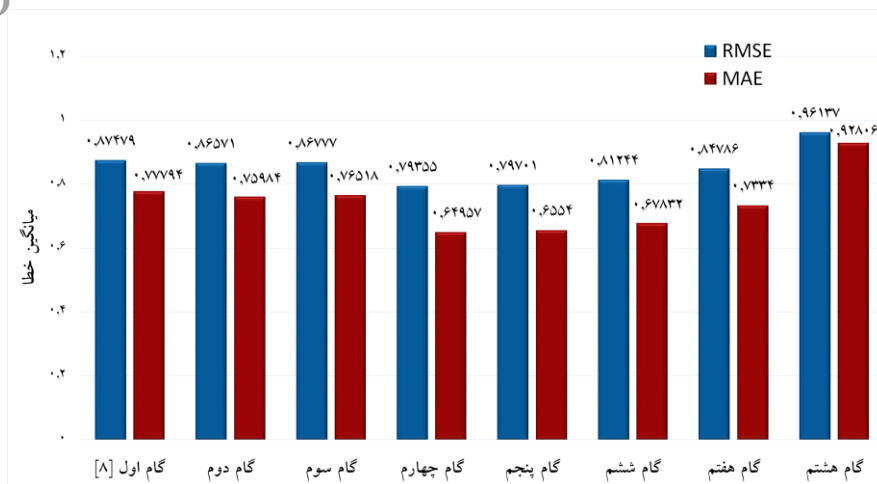
گروه‌های شماره ۲، ۳، ۱۲، ۱۶، ۱۹، ۲۷، ۳۰، ۴۱، ۴۸، ۵۳، ۵۶، ۵۹، ۶۲، ۸۱، ۸۷، ۹۰، ۹۷ و ۱۰۰ در گام اول و گروه‌های شماره ۲۸، ۳۹

در گروه‌های شماره ۵، ۴۷ و ۷۳ گام اول تا سوم با یکدیگر برابر شده‌اند. با توجه به برابر شدن گام اول و دوم که در پاراگراف قبل بیان شد، دلیل آنکه این دو گام با گام سوم برابر شده‌اند، این است که در این گروه‌ها، هیچ کدام از کاربران ورودی دارای وضعیت ادمین نیستند. همچنین برابری گام دوم و سوم در گروه‌های ۱، ۲، ۴، ۵، ۱۳، ۱۶، ۱۹، ۲۰، ۲۳، ۲۹، ۳۴، ۳۷، ۴۱، ۴۵، ۴۷، ۴۸، ۵۱، ۵۵، ۶۰، ۶۱، ۶۷، ۷۰، ۷۳، ۷۴، ۷۸، ۷۹، ۸۱، ۸۲، ۸۸، ۸۹، ۹۱، ۹۸ و ۹۹ نیز مشاهده شده است.

گروه‌های شماره ۲، ۳، ۱۲، ۱۶، ۱۹، ۲۷، ۳۰، ۴۱، ۴۸، ۵۳، ۵۶، ۵۹، ۶۲، ۸۱، ۸۷، ۹۰، ۹۷ و ۱۰۰ در گام اول و گروه‌های شماره ۲۸، ۳۹



## Archive of SID



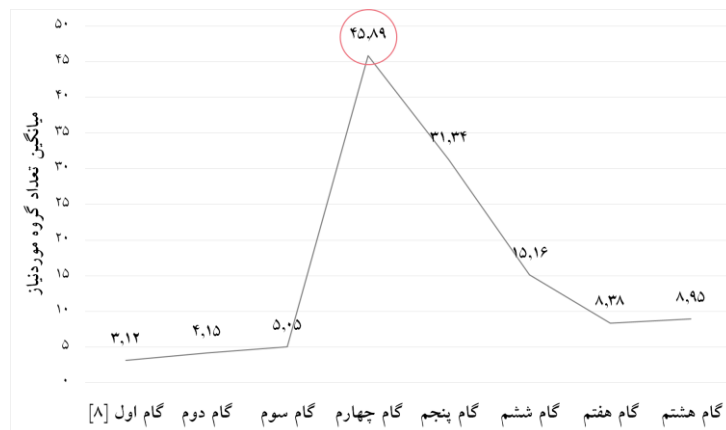
شکل ۱۵. میانگین خطای RMSE و MAE برای ۱۰۰ گروه (مقایسه گام‌های روش پیشنهادی با روش مقاله [A])

۹۲ به‌عنوان گروهی با بیشترین خطا در بین تمامی گروه‌ها، خطای  $RMSE$  را به ۰,۰۵۸ و خطای  $MAE$  را به ۰,۰۸۵ افزایش داده است. کیفیت گروه‌ها نیز از گام ششم تا هشتم افزایش می‌یابد و همچنین خطا در این گام‌ها به ترتیب افزایش یافته است. با توجه به این نکته می‌توان به این نتیجه دست‌یافت که در گام‌های ششم تا هشتم کیفیت با خطا، رابطه مستقیم دارد.

با توجه به جدول ۴ و ۵، علاوه بر خطا، هر یک از گام‌ها بر اساس تعداد گروه مورد نیاز و تعداد کاربر به‌دست‌آمده نیز بررسی شده‌اند. طبق شیوه اجرای هر یک از گام‌ها که در بخش روش پیشنهادی بیان شد، خروجی هر گام، فهرستی از گروه‌های رتبه‌بندی شده می‌باشد که کاربران آن گروه‌ها به‌ترتیب تا رسیدن به فهرستی از کاربران با تعداد موردنظر با یکدیگر ادغام می‌شوند. هر یک از گام‌ها جهت دستیابی به فهرستی از کاربران (با تعداد موردنظر)، تعداد گروه‌های متفاوتی را با یکدیگر ادغام می‌نمایند. شکل ۱۶ گام‌ها را بر اساس تعداد گروه موردنیاز جهت دستیابی به فهرست کاربران با تعداد موردنظر، نشان می‌دهد. تعداد گروه‌های موردنیاز در هر یک از گام‌ها، هرچقدر کمتر باشد، آن گام توانسته است با تعداد گروه کمتری به فهرست کاربران نهایی (با تعداد موردنظر) دست یابد. البته اگر روشی با کمترین گروه مورد نیاز به فهرستی از کاربران دست یابد، ملاک بر

که رتبه‌بندی گروه‌ها بر اساس فهرست مهاجرت یافته نسبت به فهرست گروه‌های گام اول اثر رتبه‌بندی گروه‌های با تعداد عضو زیاد را کاهش می‌دهد. دلیل اینکه در دیگر گام‌ها همچنین رخدادی مشاهده نمی‌شود، این است که از گام چهارم که تعداد مشترک، بر تعداد اعضا تقسیم می‌شود، اثر رتبه‌بندی گروه‌های با تعداد عضو زیاد خنثی شده است. طبق جدول ۳، گروه‌های شماره ۲، ۳، ۱۶، ۱۹، ۳۰، ۴۱، ۴۸، ۵۳، ۵۶، ۵۹، ۸۱، ۸۷، ۹۰، ۹۷ دارای نام‌های مشابهی در حوزه برابری و اعلام بار هستند. با توجه به نتایج ارزیابی گام اول در این گروه‌ها، مشخص است که اکثر اعضای این گروه‌ها در گروهی با تعداد اعضای ۱۹۶۰۵۶ عضو هستند؛ چراکه تمامی گروه‌های مذکور در گام اول با یک گروه و تعداد اعضای ۱۹۶۰۵۶ متوقف شده‌اند. این نکته به همان تعریف گام اول که گروه‌ها بر اساس بیشترین تعداد عضو مشترک با کاربران ورودی رتبه‌بندی می‌شوند، اشاره دارد که نتیجه آن نیز بالا آمدن رتبه گروه‌های با تعداد عضو زیاد است.

با بررسی میانگین خطای گام‌ها در جدول ۴ و ۵، گروه شماره ۱۰ از گام اول تا چهارم، به‌عنوان گروهی با کمترین خطا در بین تمامی گروه‌ها، خطای  $RMSE$  را به‌اندازه ۰,۵۹۹ و خطای  $MAE$  را به‌اندازه ۰,۸۳۸ کاهش داده است. همچنین از گام اول تا چهارم، گروه شماره

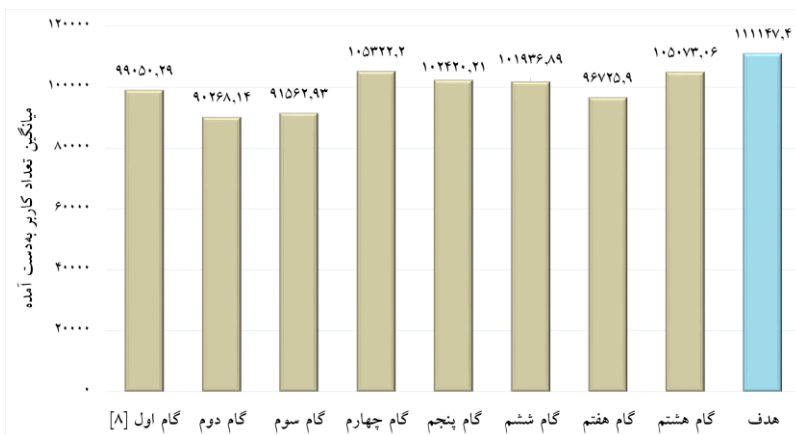


شکل ۱۶. میانگین تعداد گروه موردنیاز هر گام

هشتم صعود یافته است. دلیل نزولی بودن نمودار شکل ۱۶ در گام‌های ششم و هفتم، این است که در فهرست گروه‌های رتبه‌بندی شده در این دو گام، گروه‌های با تعداد عضو کم پس از رتبه‌بندی گروه‌ها بر اساس درصد  $G_C$  و  $G_A$  به مراتب کاهش می‌یابند. همچنین دلیل این که در گام هشتم تعداد گروه موردنیاز افزایش یافته است این می‌باشد که در فهرست گروه‌های رتبه‌بندی شده در این گام گروه‌های با تعداد عضو کم، با رتبه‌بندی گروه‌ها بر اساس درصد  $G_F$  کاهش یافته‌اند.

مقایسه‌ای دقیق‌تر از میانگین تعداد گروه موردنیاز، میانگین تعداد کاربر به‌دست‌آمده توسط هر یک از گام‌های روش پیشنهادی است. میانگین حداکثر تعداد کاربر به‌دست‌آمده توسط هر گام در شکل ۱۷ نشان داده شده است. در شکل ۱۷، علاوه بر میانگین تعداد کاربر به‌دست‌آمده توسط هر یک از گام‌ها، میانگین هدف (تعداد کاربر موردنظر) برای ۱۰۰ گروه نیز نشان داده شده است. تعداد کاربران به‌دست‌آمده در هر گام، هرچقدر به تعداد کاربر هدف نزدیک‌تر باشد، آن گام بهتر توانسته است مجموعه کاربر نهایی (با تعداد موردنظر) را به‌دست آورد. در گام اول، به دلیل وجود برخی گروه‌های با تعداد عضو بیشتر از هدف، انتظار بر این بود تا شاهد بیشترین میانگین تعداد کاربر به‌دست‌آمده در این گام باشیم که این‌طور نیست. این گام پس از گام ششم در جایگاه پنجم از این مقایسه قرار دارد. گام‌های دوم و سوم به‌ترتیب با بررسی مهاجرت و مهاجرت کاربران ادمین ورودی، دارای کمترین میانگین تعداد کاربر به‌دست‌آمده، در بین دیگر گام‌ها هستند. گام چهارم، نزدیک‌ترین گام به مجموعه هدف است که دلیل آن نیز رتبه‌بندی گروه‌ها بر اساس تعداد عضو کم است. بر طبق شکل ۱۷، پس از گام چهارم تا گام هفتم، نمودار نزولی است که بیانگر دست‌یابی به مجموعه کاربر کمتری نسبت به هدف، در این گام‌ها است. سپس در گام هشتم، نمودار نزدیک به میانگین تعداد کاربر در گام چهارم شده است و توانسته است پس از گام چهارم بیشترین مجموعه‌ی کاربر، نزدیک به مجموعه هدف را به‌دست آورد.

خوب یا بد بودن آن روش نیست، چراکه تعداد اعضا در هر گروه متغیر است. در صورتی که روشی بتواند با کمترین تعداد گروه موردنیاز به بیشترین تعداد کاربر در فهرست نهایی نزدیک به تعداد کاربر موردنظر دست یابد، آن روش، روشی بهینه است. گام اول در مقایسه با تمامی گام‌ها توانسته است با تعداد گروه کمتری و با میانگین تعداد ۳،۱۲ گروه به فهرست کاربران نهایی با تعداد موردنظر دست یابد. در گام اول گروه‌های با تعداد اعضای زیاد همیشه در ابتدای فهرست گروه‌های رتبه‌بندی شده، قرار دارند و همین دلیل باعث می‌شود تا این گام، گروه کمتری نسبت به دیگر گام‌ها نیاز داشته باشد. در گام دوم و سوم با توجه به اینکه مهاجرت و مهاجرت کاربران ادمین ورودی به‌دست‌آمده است، اما باز هم گروه‌های با تعداد عضو زیاد در ابتدای فهرست قرار می‌گیرند. به‌دلیل اینکه در فهرست گروه‌های رتبه‌بندی شده گام‌های دوم و سوم، گروه‌های با تعداد عضو زیاد از ابتدای فهرست کاهش می‌یابند باعث می‌شود تا این دو گام نسبت به گام اول به تعداد گروه‌های بیشتری جهت دست‌یابی به فهرست کاربران نهایی با تعداد موردنظر نیاز داشته باشند. بدو از بین بردن رتبه‌بندی گروه‌های با تعداد عضو زیاد در گام چهارم است. در شکل ۱۶، ناحیه‌ای که با دایره قرمز رنگ نشان داده شده است، بیان‌کننده تعداد گروه موردنیاز در گام چهارم است که نسبت به تمامی گام‌ها به تعداد گروه بیشتری و با میانگین ۴۵،۸۹ جهت دست‌یابی به فهرست کاربران نهایی با تعداد موردنظر نیاز دارد. دلیل افزایش تعداد گروه در گام چهارم این است که هنگامی که مقدار  $H$  بر  $M$  تقسیم می‌شود، گروه‌ها بر اساس درصد موجود بودن کاربر از هر گروه رتبه‌بندی می‌شوند و این باعث رتبه‌بندی بهتر گروه‌های با تعداد عضو کم می‌شود. در گام پنجم به‌دلیل اینکه کسر حاصل از گام چهارم ( $\frac{H}{M}$ ) بهبود یافته است، در فهرست گروه‌های رتبه‌بندی شده، گروه‌های با تعداد عضو کم از ابتدای فهرست کاهش یافته است و این گام به میانگین تعداد ۳۱،۳۴ گروه جهت دست‌یابی به فهرست کاربران نهایی با تعداد موردنظر نیاز دارد. در گام‌های ششم، هفتم و هشتم که رتبه‌بندی گروه‌های با کیفیت را تضمین می‌نمایند، تعداد گروه‌های مورد نیاز در گام ششم و هفتم رشد نزولی دارد و در گام



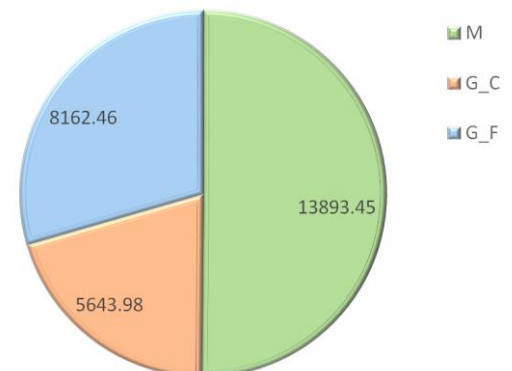
شکل ۱۷. میانگین تعداد کاربر به‌دست‌آمده توسط هر یک از گام‌ها

## Archive of SID

می‌باشد. این روش شامل گام‌هایی جدید در ادامه روش گروه‌های مشابه [۸] که جهت توصیه گروه ارائه شده بود، ایجاد شده است. هر یک از گام‌ها، گام قبل از خود را تکمیل می‌نماید و همچنین هر گام را می‌توان روشی جداگانه جهت توصیه کاربران در نظر گرفت. گام‌های روش پیشنهادی در این مقاله با روش گروه‌های مشابه که از گراف عضویت کاربران جهت توصیه گروه بهره برده است، ارزیابی و مقایسه شده‌اند. روش پیشنهادی تمامی معایب روش گروه‌های مشابه را رفع و با افزودن مهاجرت بین‌گروهی کاربران و بهبود آن، توصیه‌هایی با کیفیت را ارائه داده است. مهاجرت نیز بدین معنا است که حداکثر کاربرانی که هر گروه را ترک کرده‌اند، در حال حاضر عضو چه گروه‌هایی هستند. روش مهاجرت بر اساس تحلیل رفتار بین گروهی کاربران، ترجیحات آنان را جهت بهبود توصیه در نظر می‌گیرد. روش پیشنهادی با مدل‌سازی پارامترهای عددی استخراج‌شده از گروه‌های کاربران در تلگرام ایجاد شده است؛ همچنین این روش، کیفیت گروه‌هایی که کاربران از آنان توصیه می‌شوند را تضمین می‌نماید. پارامترهای کیفیت، گروه‌ها را بر اساس تعداد کاربران ادمین کم، سابق کم و جاری زیاد، رتبه‌بندی می‌نمایند. ما پس از افزودن پارامترهای کیفیت به روش پیشنهادی، مشاهده کردیم که خطاها با افزودن کیفیت افزایش می‌یابد که این بیانگر رابطه مستقیم کیفیت و خطا در روش پیشنهادی است. نتایج حاصل از ارزیابی بیانگر آن است که روش پیشنهادی قادر به فراهم کردن توصیه کاربرانی دقیق با خطای کم مطابق با رفتار و علاقه‌مندی کاربران ورودی است. همچنین ما پس از تحلیل نتایج ارزیابی به این نکته دست‌یافتیم که در صورتی که کاربران ورودی به سامانه توصیه‌گر، با تحلیل رفتار و در نظر گرفتن علاقه‌شان، بر اساس درصد اشتراک با گروه‌ها رتبه‌بندی شوند، خطای کاربران توصیه شده نسبت به روش گروه‌های مشابه که تنها از گراف عضویت استفاده می‌نماید در معیار  $RMSE$  به  $0,0812$  و در معیار  $MAE$  به  $0,128$  کاهش می‌یابد؛ همچنین این روش، توانایی پیش‌بینی مجموعه کاربران توصیه‌شده نهایی سامانه توصیه‌گر را نسبت به روش گروه‌های مشابه، به‌طور میانگین به  $6271,9$  کاربر افزایش داده است. درنهایت، اگر که کاربران را بر اساس سه وضعیت عضو سابق، جاری و ادمین در هر گروه در نظر بگیریم و با استفاده از مدل‌سازی رفتار آنان و بهره‌گیری از گراف، کاربرانی متناسب با علاقه‌شان توصیه نماییم، عملکرد توصیه‌گر بهبود و نتایج حاصل، خطای کمتری خواهند داشت. ارزیابی روش پیشنهادی در این مقاله به‌وسیله  $100$  گروه تخصصی تلگرام انجام شده است. ما با میانگین گرفتن از تعداد اعضا و وضعیت متفاوت کاربران، برای  $100$  گروه مورد ارزیابی به این نتیجه دست‌یافتیم که هرچقدر تعداد اعضای هر گروه افزایش می‌یابد، به ترتیب تعداد اعضای با وضعیت سابق، جاری و ادمین در هر گروه افزایش می‌یابد. همچنین در این پژوهش با تحلیل داده‌های استخراج شده از تلگرام

با توجه به شکل ۱۷، در صورتی که گام‌های روش پیشنهادی را با گام اول (روش گروه‌های مشابه) بر اساس تعداد کاربران به‌دست‌آمده مقایسه نماییم، این نکته مهم حاصل می‌شود که در بین گام‌های روش پیشنهادی، گام چهارم توانسته است با افزایش تعداد میانگین  $6271,9$  کاربر نسبت به گام اول تعداد کاربران توصیه‌شده نهایی سامانه توصیه‌گر را افزایش دهد. البته گام چهارم که بهترین گام جهت به‌دست آوردن کاربران نهایی است، هنوز به‌طور میانگین تعداد  $5825$  کاربر تا رسیدن به هدف اختلاف دارد. علاوه بر گام چهارم، گام‌های پنجم، ششم و هشتم نیز به‌ترتیب به‌طور میانگین  $3369,9$ ،  $2886,6$  و  $6022,7$  کاربر، نسبت به گام اول تعداد کاربران نهایی توصیه‌گر را افزایش داده‌اند. اما گام‌های دوم، سوم و هفتم به‌ترتیب به‌طور میانگین  $8782,1$ ،  $7487,4$  و  $2324,4$  نسبت به گام اول تعداد کاربران نهایی توصیه‌گر را کاهش داده‌اند.

در ادامه تحلیلی دیگر با بررسی  $100$  گروه مورد ارزیابی بیان می‌شود. شکل ۱۸ میانگین  $M$ ،  $G_C$  و  $G_F$  را برای  $100$  گروه مورد ارزیابی نشان می‌دهد. ما با تحلیل  $M$ ،  $G_C$  و  $G_F$  در هر یک از  $100$  گروه مورد ارزیابی به این نتیجه دست‌یافتیم که با افزایش  $M$  در هر گروه، به ترتیب  $G_C$ ،  $G_F$  و  $G_C$  افزایش می‌یابد. دلیل افزایش کم  $G_A$  در هر گروه این است که تعداد ادمین هیچ‌وقت همانند تعداد اعضا، تعداد جاری و تعداد سابق افزایش نمی‌یابد. میانگین  $G_A$  برای  $100$  گروه  $9,938$  است و به دلیل مقدار بسیار کم نسبت به مقادیر  $M$ ،  $G_F$  و  $G_C$  در شکل ۱۸ نشان داده نشده است. بر طبق شکل ۱۸، به‌طور میانگین از  $13,893$  اعضای گروه، میانگین  $5643$  کاربر در وضعیت  $C$  و  $8162$  کاربر در وضعیت  $F$  قرار دارند.



شکل ۱۸. میانگین  $M$ ،  $G_C$  و  $G_F$  برای  $100$  گروه مورد ارزیابی

## ۵. نتیجه‌گیری و کارهای آینده

در این مقاله با استفاده از تحلیل گراف عضویت و مدل‌سازی مهاجرت بین گروهی کاربران، روشی برای توصیه کاربران ارائه شده است. روش پیشنهادی از سوابق عضویت، وضعیت جاری، وضعیت ادمین بودن کاربران در گروه‌ها و همچنین تعداد اعضای هر گروه برای تولید فهرستی از کاربران جهت توصیه استفاده می‌نماید و در این راستا قادر به مدل‌سازی رفتار کاربران و همچنین تضمین کیفیت توصیه‌ها

[۶] بهشتی‌نژاد، راحله. سمیع، محمدابراهیم. (۱۳۹۶) «بهبود سیستم‌های توصیه‌گر با کمک وب معنایی»، دوفصلنامه فناوری اطلاعات و ارتباطات ایران، دوره ۹، شماره ۳۱ (۱۳۹۶): ۴۵-۵۶.

[۷] صابری، نفیسه. منتظر، غلامعلی. (۱۳۸۹) «شخصی سازی محیط یادگیری الکترونیکی به کمک توصیه‌گر فازی مبتنی بر تلفیق سبک یادگیری و سبک شناختی»، دوفصلنامه فناوری اطلاعات و ارتباطات ایران، دوره ۲، شماره ۳ (۱۳۸۹): ۹۱-۱۰۹.

[۸] هاشمی، سیدعلی. زارع چاهوکی، محمدعلی. (۱۳۹۷)، «توسعه بازاریابی با توصیه‌گر گروه‌های پیام‌رسان‌ها»، دوفصلنامه محاسبات و سامانه‌های توزیع شده، سال اول، شماره اول (۱۳۹۷): ۲۱-۳۰.

[۹] G. Ramakrishnan, V. Saicharan, K. Chandrasekaran, M. V. Rathnamma, and V. Venkata Ramana. "Collaborative Filtering for Book Recommendation System", *Advances in Intelligent Systems and Computing*, Vol. ۱۰۵۷, Springer, Singapore, ۲۰۲۰, pp. ۳۲۵-۳۳۸.

[۱۰] A. B. Melchiorre, E. Zangerle, and M. Schedl, "Personality Bias of Music Recommendation Algorithms", in *Fourteenth ACM Conference on Recommender Systems*, New York, NY, USA, ۲۰۲۰, pp. ۵۳۳-۵۳۸.

[۱۱] محمدرضایی لرکی، رضوان. روان‌مهر، رضا. امراللهی، میلاد. (۱۳۹۸) «سیستم توصیه‌گر زمینه آگاه سفر با بهره‌گیری از اطلاعات عکسهای برچسب‌دار جغرافیایی»، دوفصلنامه فناوری اطلاعات و ارتباطات ایران، دوره ۱۱، شماره ۴۱ (۱۳۹۸): ۹۶-۷۵.

[۱۲] L. Sang et al., "Context-Dependent Propagating based Video Recommendation in Multimodal Heterogeneous Information Networks", *IEEE Transactions on Multimedia*, July ۲۰۲۰, p. ۱.

[۱۳] A. Zare, M. R. Motadel, and A. Jalali, "Presenting a hybrid model in social networks recommendation system architecture development", *AI & SOCIETY*, Vol. ۳۵, No. ۲, June ۲۰۲۰, pp. ۴۶۹-۴۸۳.

[۱۴] D. M. JIMÉNEZ-BRAVO, J. F. D. Paz, and G. Villarrubia, "Twitter's Experts Recommendation System Based on User Content," In: *International Symposium on Distributed Computing and Artificial Intelligence*, Springer, Cham, ۲۰۱۸, pp. ۲۵۱-۲۵۸.

[۱۵] X. Cai et al., "Learning Collaborative Filtering and Its Application to People to People Recommendation in Social Networks," ۲۰۱۰

این نتیجه حاصل شد که تعداد بسیاری از گروه‌های این پیام‌رسان بین ۱۰۰۰ تا ۲۰۰۰ عضو دارند؛ این نتایج با میانگین تعداد اعضای ۹۲۰۸۱۰ سوپرگروه در تلگرام به‌دست‌آمده است. داده‌های این مقاله بخش بزرگی از اطلاعات عمومی پیام‌رسان تلگرام را در نظر گرفته است؛ این داده‌ها شامل بیش از ۱۲۰ میلیون کاربر و ۹۰۰,۰۰۰ سوپرگروه است.

جهت توسعه و بهبود این پژوهش در آینده می‌توان تعداد بیشتری از کاربران و گروه‌ها را در نظر گرفت. در این پژوهش از گراف کاربران به‌تنهایی استفاده شده است، حال آنکه می‌توان از اطلاعات محتوایی کاربران جهت برچسب‌زدن به هر گره کاربر در گراف بهره برد. جهت بهبود کارایی توصیه‌گر کاربر می‌توان از پیام‌های کاربران، تاریخ و ساعت ارسال پیام در گروه‌ها، استفاده نمود. همچنین در پژوهش‌های آتی می‌توان با در نظر گرفتن تعداد پیام‌های کاربر در هر گروه، امتیازی جهت مشارکت اعضا در نظر گرفت. چهارچوب این پژوهش مختص داده‌های تلگرام نبوده و می‌توان با استفاده از داده‌های دیگر پیام‌رسان‌ها و شبکه‌های اجتماعی که امکان ایجاد گروه را در بر دارند و همچنین رابط برنامه‌نویسی آنان در دسترس است، علاوه بر اثبات مؤثر بودن این روش بر روی آن محیط‌ها، به نقاط ضعف و قوت این روش و بهبود آن دست‌یافت.

## مراجع

[۱] Durov, P. ۲۰۲۰. ۴۰۰ Million Users, ۲۰,۰۰۰ Stickers, Quizzes ۲,۰ and ۴۰۰K for Creators of Educational Tests, accessed ۱۰ January ۲۰۲۱, <<https://telegram.org/blog/۴۰۰-.-million>>.

[۲] Iqbal, M. ۲۰۲۰. Telegram revenue and usage statistics, accessed ۱۳ May ۲۰۲۱, <<https://www.businessofapps.com/data/telegram-statistics/>>.

[۳] Clement, J. ۲۰۲۰. Most popular global mobile messenger apps as of October ۲۰۲۰, based on number of monthly active users, accessed ۱۰ January ۲۰۲۱, <<https://www.statista.com/statistics/۲۵۸۷۴۹/most-popular-global-mobile-messenger-apps/>>.

[۴] میلانی، فرزانه. زنوزی، سیدجعفر. (۱۳۹۸) «بررسی تاثیر استفاده از انواع استراتژی‌های بازاریابی در شبکه‌های اجتماعی بر جلب اعتماد مشتریان شورایی»، دوفصلنامه فناوری اطلاعات و ارتباطات ایران، دوره ۱۱، شماره ۳۹ (۱۳۹۸): ۹۱-۱۰۴.

[۵] فتحیان، محمد. حضرتقلی‌زاده، رحیم. (۱۳۹۵) «ارائه روشی مناسب برای دسته‌بندی نامه‌های الکترونیکی تبلیغاتی بر مبنای پروفایل کاربران»، دوفصلنامه فناوری اطلاعات و ارتباطات ایران، دوره ۸، شماره ۲۷ (۱۳۹۵): ۲۱-۳۶.

Archive of SID

- [۲۲] H. J. Jeong, K. H. Lee, and M. H. Kim, "DGC: Dynamic group behavior modeling that utilizes context information for group recommendation", *Knowledge-Based Systems*, Vol. ۲۱۳, December ۲۰۲۰, p. ۱۰۶۶۵۹.
- [۲۳] S. Guo, S. J. Ji, C. Zhang, X. Wang, and J. Zhao, "A New Multi-criteria Recommendation Algorithm for Groups of Users", In: *International Conference on Genetic and Evolutionary Computing, Springer*, Singapore, ۲۰۱۹, pp. ۷۳۵-۷۴۲.
- [۲۴] F. Ortega, R. Hurtado, J. Bobadilla, and R. Bojorque, "Recommendation to groups of users using the singularities concept", *IEEE Access*, Vol. ۶, July ۲۰۱۸, pp. ۳۹۷۴۵-۳۹۷۶۱.
- [۲۵] D. Karimpour, M. A. Zare Chahooki and A. Hashemi, "User recommendation based on Hybrid filtering in Telegram messenger," *۲۶th International Computer Conference, Computer Society of Iran (CSICC)*, ۲۰۲۱, pp. ۱-۷.
- [۲۶] E. Yalcin, F. Ismailoglu, and A. Bilge, "An entropy empowered hybridized aggregation technique for group recommender systems", *Expert Systems with Applications*, Vol. ۱۶۶, October ۲۰۲۰, p. ۱۱۴۱۱۱.
- [۲۷] H. Yin, Q. Wang, K. Zheng, Z. Li, and X. Zhou, "Overcoming Data Sparsity in Group Recommendation", *IEEE Transactions on Knowledge and Data Engineering*, September ۲۰۲۰.
- [۲۸] A. Hashemi and M. A. Z. Chahooki, "Telegram group quality measurement by user behavior analysis", *Social Network Analysis and Mining*, Vol. ۹, No. ۱, July ۲۰۱۹, p. ۳۳.
- IEEE International Conference on Data Mining*, ۲۰۱۰, pp. ۷۴۳-۷۴۸.
- [۱۶] L. Berkani, "A semantic and social- based collaborative recommendation of friends in social networks", *Software: Practice and Experience*, Vol. ۵۰, No. ۸, August ۲۰۲۰, pp. ۱۴۹۸-۱۵۱۹.
- [۱۷] P. Kumar and G. R. Mohana Reddy, "Friendship Recommendation System Using Topological Structure of Social Networks," in *Proc. of Progress Intelligent Computing Techniques: Theory, Practice, and Applications*, Singapore, Aug. ۲۰۱۸, pp. ۲۳۷-۲۴۶.
- [۱۸] S. Huang, J. Zhang, L. Wang, and X.-S. Hua, "Social Friend Recommendation Based on Multiple Network Correlation," *IEEE Transactions on Multimedia*, vol. ۱۸, no. ۲, pp. ۲۸۷-۲۹۹, Feb. ۲۰۱۶.
- [۱۹] A. Alshammari, S. Kapetanakis, R. Evans, N. Polatidis, and G. Alshammari, "User modeling on twitter with exploiting explicit relationships for personalized recommendations", In *International Conference on Hybrid Intelligent Systems*, Springer, Cham, ۲۰۱۸, pp. ۱۳۵-۱۴۵.
- [۲۰] D. F. Gurini, F. Gasparetti, A. Micarelli, and G. Sansonetti, "A Sentiment-Based Approach to Twitter User Recommendation," in *Proc. of Recommender System and the Social Web (CEUR Workshop)*, Hong kong, China, Oct. ۲۰۱۳.
- [۲۱] K. Xu et al., "Improving user recommendation by extracting social topics and interest topics of users in uni-directional social networks," *Knowledge-Based Systems*, vol. ۱۴۰, pp. ۱۲-۱۳۳, Jan. ۲۰۱۸.