



شناسایی جوامع و انجمن ها در شبکه های اجتماعی

Identifying Communities and Associations in Social Networks

استاد راهنما

دکتر علی هارون آبادی

تهیه کننده : بهروز عدالتی مرفه

دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات

گرایش شبکه های کامپیوتری

سال تحصیلی ۱۳۹۵-۹۶

چکیده:

شبکه های اجتماعی به عنوان واقعیتهای جدید در لایه های گوناگون زندگی افراد نفوذ کرده است، این دنیای جدید با رشد سرسام آور و قدرت ماورایی که در ارتباطات بشری به وجود آورده است اکنون به دنبال ارائه خدمات فناوری گوناگون مطابق با سلائیق افراد است. با دقت در این مفهوم به پیچیدگی و تخصصی شدن این محیط جدید پی می بریم.

از طرفی شبکه های اجتماعی با ایجاد فضای تعاملی و نزدیک بین کاربران محیطی مساعد را برای معرفی کالا، خدمات آماده نموده و افراد با ایجاد ارتباط نزدیک و توسعه ارتباطات به شکل روزافزون به دنبال کشف فرصت های ایجاد شده از رشد و توسعه شبکه های اجتماعی می باشند. صاحبان شبکه ها هم با استفاده از فرصت ایجاد شده با بهنگام سازی امکانات و برنامه های ارائه شده در شبکه های اجتماعی سعی در تأمین سلائیق کاربران و مخاطبان خود دارند.

در این شبکه های عمومی و اجتماعی به شکل طبیعی و با استفاده از قوانین حاکم بر جوامع انسانی جوامع کوچکتر و انجمن هایی شکل می گیرند و جوامعی به صورت پنهان ایجاد می گردند. ما می توانیم با استفاده از قوانین حاکم بر جوامع انسانی و همچنین با استفاده از الگوریتم ها و تکنیک های استخراج جوامع انسانی اقدام به شناسایی و جدا نمودن جوامع پنهان در دنیای مجازی بپردازیم.

با شناسایی جوامع می توانیم به صورت دقیق تر و جزئی تر به تحلیل جوامع، شناسایی مخاطبان، شناسایی سلائیق کاربران و مخاطبان بپردازیم. ما می توانیم به دانستن در مورد افراد موجود در جوامع، به شناسایی نیازها و ارائه پاسخ خای دقیق و متناسب با نیازها اقدام نماییم. با توجه به اینکه تعداد روش های شناسایی جوامع بسیار زیاد و متنوع هستند و از طرفی به منظور بهره برداری از بهترین روش با توجه به بزرگی و پیچیدگی شبکه ها باید تحلیلگران با روش ها و تکنیک ها آشنا باشند تا بتوانند با شناخت دقیق جامعه مورد نظر به انتخاب بهترین روش جداسازی انجمن ها در شبکه های مجازی بپردازند. در این تحقیق ابتدا به معرفی ابزارهای ایجاد جوامع و روش های متداول استخراج جوامع می پردازیم، سپس در ادامه به مقایسه دو روش و شناسایی راهکارهای بهبود روش های مذکور می پردازیم.

کلمات کلیدی: شبکه های اجتماعی، انجمن ها، شناسایی انجمن ها و جوامع، ابزارها و تکنیک های شناسایی جوامع، شبکه توئیت، ...

فهرست مطالب :

۱. فصل اول (کلیات تحقیق)

- ۱-۱. مقدمه ۴
- ۲-۱. بیان مسأله ۵
- ۳-۱. اهداف انجام تحقیق ۶

۲- فصل دوم (بستر تحقیق)

- ۱-۲. آشنایی با شبکه های اجتماعی ۷
- ۲-۲. شبکه های اجتماعی در ریاضی و کامپیوتر ۱۱
- ۱-۱-۲. ساختار ماتریس برای مطالعه شبکه های اجتماعی ۱۱
- ۲-۱-۲. ساختار گراف برای مطالعه شبکه های اجتماعی ۱۲
- ۳-۲. انواع روش ها و راه های شناسایی جوامع ۱۳
- ۱-۳-۲. روش بخش بندی گراف ۱۳
- ۲-۳-۲. روش مبتنی بر حذف یال ۱۴
- ۳-۳-۲. روش خوشه بندی ۱۵
- ۴-۳-۲. استفاده از احتمالات و روش های بیزین ۱۶
- ۵-۳-۲. سایر روش ها ۱۶
- ۱-۵-۳-۲. روش طیفی ۱۶
- ۲-۵-۳-۲. روش های سلسله مراتبی ۱۷

۳- فصل سوم (پیشینه تحقیق)

- ۱-۳. نگاهی به الگوریتم های مورد استفاده در شناسایی جوامع ۱۷
- ۲-۳. شناسایی و دسته بندی موضوعات در توئیتز به عنوان یک نمونه از شبکه های اجتماعی ۱۹
- ۱-۲-۳. دنبال کردن شباهت روابط ۱۹
- ۲-۲-۳. شباهت دنبال کنندگان (Followers) ۲۰
- ۳-۲-۳. شباهت دوستان مشترک (Friends) ۲۰

۲۰ شباهت هشتگ ۴-۲-۳
۲۱ شباهت پاسخدهی ۵-۲-۳
۲۱ شباهت های بین کاربران ۶-۲-۳
۲۱ شباهت های کاربران کل ۷-۲-۳
۲۲ خوشه بندی کاربران بر اساس انتشار وابستگی ۸-۲-۳
۲۲ استخراج موضوع و برجسب زدن ۳-۳
۲۲ استخراج موضوعات بر اساس تخصیص پنهان دیریکله (LDA) ۱-۳-۳
۲۳ شناسایی جوامع در شبکه های اجتماعی با استفاده از تکنیک های خوشه بندی ۲-۳-۳
۲۴ شرح مشکل ۲,۱-۳-۳
۲۵ الگوریتم پیشنهاد شده ۲,۲-۳-۳
۲۵ نتایج تجربی ۴-۳
۲۶ کلوب Zachary Karate ۱-۴-۳
۲۸ مدرسه فوتبال آمریکایی ۲-۴-۳
۲۹ شبکه Dolphin Bottlenose ۳-۴-۳

۴ فصل چهارم

۳۰ ۱-۴ نتیجه گیری
۳۲ ۲-۴ فهرست منابع

۱. فصل اول

۱-۱. مقدمه:

هر شبکه اجتماعی شامل مجموعه ای از انسان ها و روابط اجتماعی بین آن هاست. بنابراین می توان گفت شبکه اجتماعی از ابتدای وجود بشر وجود داشته است، اما شبکه های اجتماعی که امروزه به واسطه پدید آمدن دنیای مجازی و ابزارهای دیجیتال و نرم افزارهای متنوع در این حوزه به وجود آمده اند پدیده های نوینی هستند به نام شبکه های اجتماعی مجازی یا اصطلاحاً شبکه های مجازی.

بر اساس تعریف شبکه اجتماعی به تعریف و تحلیل روابط بین انسان ها و سازمان ها و یا هرگونه تعامل مابین موجودیت ها می پردازد. این دو مفهوم یعنی شبکه های اجتماعی و شبکه های اجتماعی مجازی بسیار به هم نزدیک هستند اما از نظر ماهیت تفاوت اساسی با هم دارند. یکی در حیطه جامعه شناسی و دیگری در حوزه فناوری اطلاعات می باشد. در شبکه های اجتماعی افراد را به گره ها یا نودها تعبیر می کنیم.

شبکه های اجتماعی از تودها و یا گره ها و روابط بین آن ها تشکیل شده است. همانطور که شبکه های اجتماعی دارای انواع مختلف هستند، شبکه های مجازی نیز دارای انواع عمومی، خصوصی، تفریحی، علمی-تخصصی و ... می باشند. در میان ابزارها و نرم افزارهای موجود می توان این تقسیم بندی از سلاقی و کاربردها را بهتر مشاهده نمود، به عنوان مثال شبکه اجتماعی توئیتر بیشتر جنبه اطلاع رسانی و زمینه ای متنی داشته در صورتی که شبکه اجتماعی تلگرام کاربردی تفریحی را برای مخاطبانش در نظر گرفته است و یا شبکه اجتماعی ریسرچ گیت دارای مخاطب خاص و ویژه محققان و پژوهشگران می باشد.

همانطور که گفته شد شبکه اجتماعی مجازی زنجیره ای از ارتباطات و گره های شبکه اجتماعی یا حضور غیر فیزیکی افراد در یک محل مجازی است. نیاز به شناخت مخاطبان در شبکه های اجتماعی با توجه به رشد این شبکه ها روز به روز بیشتر می شود. برای تفکیک شبکه به جوامع می توانیم اعضای را که بیشترین تعاملات را با هم دارند به عنوان گروه ها در نظر گرفت.

تحلیل شبکه های اجتماعی یا Social Network Analysis یک استراتژی و راهکار برای تحلیل ساختار و گره های موجود در شبکه های اجتماعی است. SNA به شبکه ها اجتماعی ساختار استاندارد می گویند که بتوان آن را با استفاده از قوانین شناخته شده ریاضیات و احتمالات مطالعه نمود را نسبت داده و و افراد درون شبکه را در قالب عناصر آن ساختار شکل می دهد. به عنوان نمونه در ساختار گراف که یکی از ساختارهای استاندارد برای مطالعه شبکه های اجتماعی می باشد به ساختار شبکه به شکل یک گراف نگاه کرده و افراد درون سازمان را به عنوان رئوس آن گراف و ارتباطات را یال ها در نظر می گیرند. به این گراف اصطلاحاً گراف اجتماعی می گوئیم.

مرکزیت گراف اجتماعی بسیار مهم می باشد. این مرکزیت به گونه های متفاوتی به دست می آید. یک رأس در مرکزیت گراف، صرف نظر از نوع مسأله رأس تأثیرگذاری می باشد. در این مقاله در ابتدا به بررسی شبکه های اجتماعی و مطالعه کلی آن ها پرداخته ایم، سپس در ادامه روش های تحلیل شبکه ها را مورد شناسایی قرار داده و در نهایت انواع الگوریتم های شناسایی شبکه های اجتماعی و افراد

تأثیرگذار در شبکه های اجتماعی را همراه با ریز شدن در یک نمونه عملی از انواع روش های شناسایی شبکه های اجتماعی و افراد تأثیرگذار در آن ها به عنوان مثالی عملی مقاله حاضر را به پایان رسانده ایم.

در این نمونه های مطالعه شده با توجه به تنوع روش های ارائه شده و همچنین تنوع شبکه های اجتماعی معیارهای متفاوتی برای انتخاب روش مورد استفاده وجود دارد که بر اساس تفاوت موجود در شبکه های اجتماعی می توان روش مورد استفاده را انتخاب نمود و یا حتی در برخی از مطالعات با ترکیب روش های موجود به راه های جالبی در خصوص نحوه شناسایی شبکه های اجتماعی و افراد تأثیرگذار در شبکه ها دست یافته اند.

۱-۲. بیان مسأله:

با توسعه روزافزون اینترنت به عنوان شبکه ی شبکه ها ارتباطات بین افراد نسبت به گذشته بزرگتر و روز به روز این روابط پیچیده تر می شود، در این میان شبکه های اجتماعی نیز به عنوان پدیده ای جدید دارای مخاطبان خود بوده و روزه روز پیچیده تر از گذشته خوانند بود. کاربران در شبکه های اجتماعی بر یکدیگر تأثیری گذارند و این تأثیر معمولاً از طریق مشاهده پست های یکدیگر رخ می دهد. به دست آوردن روابط در شبکه های اجتماعی و تحلیل آن ها بخش های مختلفی از زوایای پنهان ارتباطات را روشن خواهد نمود. از طرفی حجم اطلاعات و محتوای تولید شده در شبکه های اجتماعی و توانایی دسته بندی و مدیریت آن اطلاعات به شکلی که بتوان در راستای تبلیغات، معرفی محصولات جدید و یا کاربردهای فراوان دیگر به کار گرفت مطرح می گردد.

یکی از راه های مدیریت و دسته بندی اطلاعات به منظور استفاده بهینه از اطلاعات تولید شده دسته بندی دوستان و ایجاد گروه های مختلف است. این عمل به عنوان مکانیزمی کارآمد می تواند به جداسازی جوامع مختلف برای تولید و به کارگیری اطلاعات هر گروه بسیار کارآمد باشد. از طرف مقابل این گروه ها و انجمن های ایجاد شده در شبکه های اجتماعی برای گروه دیگری از افراد که به دنبال استفاده از ظرفیت ایجاد شده به عنوان گروه های اجتماعی هستند اهمیت می یابد.

دلیل رسیدن به این مفهوم شاید در ابتدا خیلی مهم جلوه ننماید اما در بررسی های بعدی و بررسی های برپایه مطالعات ارتباطات بین جوامع کاربردهای بسیاری دارد. با مطالعه شبکه های اجتماعی و ارتباطات موجود در آن ها می توان به درکی از علایق کاربران رسید، همچنین می توان عقاید و رفتارهای آن ها را قابل بررسی و تحلیل نماید.

شناسایی انجمن ها و جوامع موضوع بسیار مهمی در شبکه های اجتماعی است. انجمن ها گروه هایی از کاربران هستند که بیشتر با هم در ارتباط بوده و به دنبال منافع یکسانی می باشند. شناسایی جوامع و انجمن ها در شبکه های اجتماعی به دلیل اینکه می تواند به عنوان نمونه ای کوچک شده از دنیای واقعی باشد در تحلیل سیستم های پیچیده در دنیای واقعی کاربرد فراوانی دارد. برخی از این شبکه ها در دنیای واقعی شبکه های بیولوژیک مانند شبکه های عصبی، دنیای پر رمز و راز وب (منظور شبکه ای از لینک های صفحات وب می باشد) و شبکه های دوستی، شبکه های اجتماعی و ... می باشد.

با شناسایی جوامع موضوع تعیین افراد تأثیرگذار در جوامع پیش می آید، شناسایی این افراد به منظور استفاده از توانایی ها و موقعیت های اجتماعی آن ها کاربردهای فراوانی دارد. بدیهی است که میزان اهمیت افراد در یک جامعه به یک اندازه نیست، برخی از افراد به دلیل داشتن جایگاه اجتماعی، روابط و دوستان با نفوذ و ... از اهمیت بیشتری برخوردارند. این که چي کسی در جامعه مدنظر مهم تر است یا چه کسی چه قدرتی دارد و درون جامعه چه می گذرد بسیار مهم می باشد، برای شناسایی این افراد مهم نه تنها باید به خود افراد اهمیت داد، بلکه دوستان و نزدیکان این افراد از نظر میزان اهمیت باید مورد بررسی قرار گیرند.

موضوع مهم دیگر در مطالعه شبکه های اجتماعی انواع شبکه های موجود می باشد، به این معنی که شبکه ها دارای خصوصیات مختلف و متفاوتی می باشند، در برخی از آن ها ارتباطات یک طرفه و در برخی دیگر دوطرفه می باشد، شبکه ها می توانند دارای کاربردهای گوناگون باشند، به دلیل این تفاوت در شبکه های اجتماعی، گاهی اوقات دیده می شود که نتایج ارائه شده قابل تعمیم به تمام انواع شبکه نیست و نتیجه گیری ها همواره برای انواع شبکه ها صادق نیستند.

بنابراین برای دست یابی به نتایج دقیق و واقعی برای انواع شبکه ها نیاز به ارائه ترکیب جدیدی از روش ها و شاخص ها بر اساس الگوریتم ها و روش های شناخته شده قبلی می باشد.

۱-۳. اهداف انجام تحقیق :

در این تحقیق سعی داریم در ابتدا با شبکه های اجتماعی به صورت کلی آشنا شده و میزان محبوبیت آن ها را در دنیا بررسی نماییم، سپس با نگاهی کلی به نقش ریاضی و احتمالات در شناسایی شبکه ها به ساختار کلی مطالعات شبکه دست یافته و بدانیم که برای مطالعه شبکه چگونه می توان از ساختار های ریاضی و احتمالات می توان کمک گرفت، پيی برای شناسایی افراد تأثیرگذار در شبکه های اجتماعی به بررسی خود شبکه اجتماعی می پردازیم و با در نظر گرفتن معیارهایی می توان افراد تأثیرگذار در این شبکه ها را شناسایی نمود.

شناسایی جوامع می تواند به روشن کردن خصوصیات شبکه های واقعی و نحوه عملکرد آن ها منجر شود. درواقع شناسایی جوامع نه تنها اهمیت خاص دارد، بلکه کاربردهای جانبی نیز دارد، به عنوان نمونه برار دادن کالاها در شبکه های آنلاین فروش یا گسترش فعالیت های بازاریابی و ویروسی و غیره از نمونه های شناسایی شبکه های اجتماعی می باشد.

این تحقیق با هدف شناخت شبکه های اجتماعی، آشنایی با الگوریتم های شناسایی جوامع در آن ها و در نهایت شناسایی افراد تأثیرگذار در شبکه های اجتماعی با معرفی چند نمونه از تحقیقات صورت گرفته در سطوح مختلف علمی سعی در ایجاد زمینه ای برای انجام یک پژوهش واقعی در آینده ای نزدیک را دارد.

۲. فصل دوم

۲-۱. آشنایی با شبکه های اجتماعی :

شبکه های اجتماعی یکی از انواع رسانه های اجتماعی محسوب می شوند و محلی برای شکل گیری جماعت های مجازی و شبکه سازی کاربران اینترنتی هستند. شبکه های اجتماعی موفق شده اند تعداد قابل توجهی از کاربران اینترنتی را جذب کنند و در سال گذشته میلادی در میان چهار فعالیت اصلی کاربران اینترنتی قرار گرفتند. برخی تحلیلگران وب، آینده فضای مجازی را در اختیار این نوع سایت ها می دانند. صدها سایت شبکه اجتماعی با حوزه های تخصصی متفاوت و همچنین با زبان های متنوع برای کاربران کشورهای مختلف در اینترنت فعالیت می کنند. کاربردها و قابلیت های شبکه های اجتماعی بسیار متنوع است و این سایت ها تمایل دارند حداکثر میزان نیاز آنلاین کاربران خود را برآورده کنند.

همچنین می توان گفت از طرفی شبکه های اجتماعی منبع اصلی برای کسب و کارهای بزرگ و کوچکی است که بدنبال تبلیغ برند خود در اینترنت هستند. پلتفرم شبکه های اجتماعی برای استفاده بسیار ساده است و حتی بعضی از این شبکه ها سرویس های پولی تبلیغات نیز ارائه می دهند، برای کسب و کارهایی که به دنبال مشتری و مخاطبان جدید می گردند. اگرچه کسب و کار شما به شبکه های اجتماعی نیاز دارد اما لزومی ندارد در تمامی آنها حضور داشته باشید! اما برای بهره برداری بهینه از قدرت و ظرفیت شبکه های اجتماعی باید محبوب ترین این شبکه ها را بشناسید و آن یک را که برای شما بهترین است شناسایی کنید.

فیس بوک بزرگترین شبکه اجتماعی موجود در اینترنت است، هم از لحاظ تعداد کاربر و هم از لحاظ اسم و رسم، این شبکه اجتماعی در فوریه ۲۰۰۴ تاسیس شد، و در بیش از ۱۲ سال اداره و پشتیبانی مستمر، نزدیک به ۱٫۶ میلیارد نفر کاربر فعال در ماه را جمع آوری کرده است، و همین امر بصورت خودکار این شبکه را تبدیل به یکی از بهترین شبکه ها برای اتصال مردم از سراسر جهان به کسب و کار شما می کند. تخمین زده می شود که بیش از یک میلیون کسب و کار کوچک و متوسط از این پلتفرم برای تبلیغ بیزنس خود استفاده می کنند.

امروزه توییتر یکی از اصلی ترین شبکه های اجتماعی آنلاین در جهان است. این یک سرویس میکرو بلاگ است که کاربران می توانند پیام های ۱۴۰ کارکتری با عنوان توییت ارسال یا دریافت کنند. توییت ممکن است شامل متن ساده، صفحات URL، تصاویر، اشاره به سایر کاربران (قبل از نماد "@" و هشتگها باشد، هشتگها کلماتی هستند که یک کاربر تصمیم به برجسته کردن آن با قرار دادن نماد "#" در مقابل آنها می کند. موضوع مورد علاقه یک اصطلاح (یک یا چند کلمه) است که در برخی لحظات به نظر می رسد در تعداد زیادی از توییت های یک محل تکرار شده است. این "مکان ها" و "لحظه ها" به دلیل متادیتای اضافی موجود در هر توییت، که مربوط به اطلاعات جغرافیایی، زمان و اطلاعات حساب پخش کننده و غیره است، شناخته شده است. بسیاری از موارد، موضوعات مورد علاقه و طرفدار با هشتگ نوشته می شوند. در توییتر چهار نوع ارتباط وجود دارد: کاربر به کاربر، کاربر به توییت، توییت به توییت و توییت به کاربر. اقدامات معتبر برای هر نوع رابطه در جدول ۲ نشان داده شده است.

جدول شماره ۲) انواع رابطه در شبکه اجتماعی توئیتر

	user	tweet
user	follows / is followed by	posts
	mention	retweets
tweet	replies to	likes
	retweeted to	replies
	posted by	
	retweeted by	replies / is replied from
	liked by	retweets / is retweeted from
	replied by	

توجه داشته باشید که هر دو ارتباط کاربر به توئیتر و ارتباط توئیتر به کاربر متقارن هستند. اقدام ذکر شده قبلاً شرح داده شده است. در توئیتر، زمانی که کاربر A کاربر B را دنبال می کند، می گوئیم A پیرو B است و B دنبال کننده A است. ما همچنین می توانیم به توئیست دیگر کاربران پاسخ دهیم، یا به عبارت دیگر، از طریق یکی از توئیست های کاربر به او پاسخ دهیم. بنابراین، ما می توانیم در نظر بگیریم که عمل پاسخ به چهار نوع روابط تعلق دارد. دو اقدام بسیار قدرتمند شامل به اشتراک گذاری و یا پاسخ به یک توئیست، یا توئیست کاربر دیگر و علامت گذاری یک توئیست به عنوان توئیست مورد علاقه یا لایک کردن یک توئیست. پاسخ دهی به یک توئیست و لایک کردن همچنین شامل تعامل مجازی بین فعالیت کاربر و نویسنده توئیست اصلی است. از سال ۲۰۱۵، توئیتر امکان ویرایش متن در retweets را فراهم کرده است، بنابراین اکنون می توانیم هر دو کار retweet و اشاره به یک کاربر دیگر را در همان توئیست انجام دهیم. به عنوان هر شبکه اجتماعی، توئیتر می تواند توسط دو مؤلفه $G = (V, E)$ ، نشان داده شود، به طوری که V مجموعه ای از گره ها است و E مجموعه ای از لبه های هدایت شده است که نشان دهنده نحوه ارتباط گره ها است. سنتی ترین دیدگاه این است که V مجموعه ای از کاربران توئیتر باشد و لبه های آن بیانگر ارتباط دنبال کردن بین کاربران را نشان می دهند به طوری که یک لبه $(u, v) \in E$ به این معنی است که کاربر u یک follower (یا followee است) بسته به مدل کاربر V. در چنین مواردی، تعاملات میان کاربران مانند بازخوانی، پاسخ، اشاره، و غیره، بخشی از پویایی ضمنی شبکه است.

لینکدین در دسامبر سال ۲۰۰۲ تاسیس و در ماه می سال ۲۰۰۳ راه اندازی شد، در ماه جون ۲۰۱۶ مایکروسافت پیشنهاد خرید این شبکه را به قیمت ۲۶٫۲ میلیارد دلار مطرح کرد. لیندکین براحتی محبوب ترین شبکه اجتماعی برای تبلیغ حرفه ای کسب و کارهاست. این سایت با ۲۴ زبان زنده دنیا و با بیش از ۴۰۰ میلیون کاربر عضو شده در دسترس است. LinkedIn بسیار مناسب برای کسانی است که بدنبال ارتباط با شرکت ها، صنایع و کسب و کارهای مشابه با کار خود هستند.

یوتیوب بزرگترین و محبوب ترین شبکه اجتماعی برای تماشا و به اشتراک گذاری ویدئو می باشد، این شبکه در فوریه ۲۰۰۵ توسط سه نفر از مدیران پیپال راه اندازی شد و در سال بعد به قیمت ۱٫۶ میلیارد دلار توسط گوگل خریداری شد. یوتیوب بیش از ۱ میلیارد بازدید ماهانه دارد و دومین موتور جست و جوی پرطرفدار اینترنت پس از گوگل محسوب میشود

اینستاگرام یک شبکه اجتماعی تصویری و متعلق به فیسبوک ست. سایت در اکتبر ۲۰۱۰ راه اندازی شد و هم اکنون بیش از ۵۰۰ میلیون کاربر فعال دارد. بیشتر کاربران از این شبکه برای ارسال اطلاعات خود همچون مسافرت، فشن، غذا، هنر و موضوعات مشابه استفاده می کنند. همچنین خاصیت منحصر به فرد این پلتفرم وجود فیلترهایی خاص برای ویرایش تصاویر و ویدئو است. تقریباً بیش از ۹۵ درصد کاربران اینستاگرام از فیسبوک نیز استفاده می کنند. (اینستاگرام یکی از محبوب ترین شبکه های اجتماعی در ایران و شاید محبوب ترین بعد از تلگرام باشد).

پینترست شبیه به اینستاگرام یک تابلوی اشتراک گذاری تصویری می باشد. پینترست در مارچ ۲۰۱۰ راه اندازی شد، این شبکه تقریباً یک تازه وارد در بین رقبا محسوب می شود اما یکی از محبوب ترین شبکه های اجتماعی است. این پلتفرم یک تابلو اعلانات دیجیتال است جایی که کسب و کارها می توانند محصولات و داشته های خود را سنجاق کنند، پینترست در سپتامبر ۲۰۱۵ اعلام کرد که بیش از ۱۰۰ میلیون کاربر دارد. کسب و کارهای کوچکی که نیاز به مخاطب خانم دارند مسلماً باید در Pinterest سرمایه گذاری کنند زیرا بیش از نیمی از کاربران این شبکه را خانم ها تشکیل می دهند.

گوگل پلاس پس از توئیتر، فیسبوک و لیندکین شبکه اجتماعی Google+ محبوب ترین شبکه محسوب می شود، ارزش سئو و بهینه سازی این شبکه برای سایت ها، آن را تبدیل به یک شبکه متفاوت و ضروری برای هر کسب و کار می کند. در دسامبر ۲۰۱۱ راه اندازی شد و تا کنون بیش از ۴۲۰ میلیون کاربر جذب کرده است.

در آگوست ۲۰۱۳ تلگرام در روسیه راه اندازی شد اما در حال حاضر مقر اصلی آن در آلمان است. این شبکه در ابتدا فقط یک پیام رسان ساده بود و به تدریج با پدیدار شدن امکاناتی همچون گروه ها و کانال ها میتوان آن را یک شبکه اجتماعی تلقی کرد.

در فوریه ۲۰۱۶ تلگرام اعلام کرد که بیش از ۱۰۰ میلیون کاربر فعال بصورت ماهانه در این شبکه هستند و روزانه ۳۵۰ هزار به این عدد اضافه میشود، همچنین در سال جاری دو شبکه اجتماعی اینستاگرام و پیام رسان واتس اپ که هر دو در مالکیت فیسبوک هستند اقدام به فیلتر و مسدودسازی لینک های تلگرام نمودند.

تامبلر یکی از سخت ترین شبکه ها برای کار کردن می باشد اما در عین حال یکی از جذاب ترین ها نیز است، این پلتفرم نوعی میکرو بلاگ توام با سوشیال مدیا ست و به شما اجازه میدهد که در چندین فرمت مختلف مطالب خود را منتشر کنید، همچون پست نقل قول، چت، پست های تصویری و ویدئویی، بنابراین شما برای انتشار پست های خود هیچ نوع محدودیتی ندارید. شبیه به توئیتر کردن در توئیتر، reblogging در Tumblr بسیار ساده ست، این شبکه اجتماعی در فوریه ۲۰۰۷ توسط David Karp راه اندازی شد و هم اکنون بیش از ۲۰۰ میلیون بلاگ دارد.

فلیکر یک پلتفرم به اشتراک گذاری عکس و ویدئوست که توسط شرکت کانادایی Ludicorp در فوریه ۲۰۰۴ تاسیس شد و بعداً در سال ۲۰۰۵ توسط یاهو خریداری شد، این شبکه بیشتر برای عکاسان و کسانی که به اشتراک گذاری عکس علاقه مندند مناسب است.

طبق آمار اعلام شده در اکتبر ۲۰۱۵ فلیکر بیش از ۱۱۲ میلیون کاربر از بیش از ۶۳ کشور جهان داشته است. و بطور میانگین روزانه ۱ میلیون تصویر در Flickr به اشتراک گذاشته می شود.

ردیت یک شبکه اخبار و سرگرمی اجتماعی است که کاربران در آن قادر به اشتراک لینک ها و متون خود هستند ، همچنین به واسطه سیستم رای گیری برای پست ها کاربران قادر هستند رتبه مطلب خود را در بین دیگر پست های هم رده مشاهده کنند ، مطالب با بیشترین امتیاز مثبت در بالای صفحه اصلی هر بخش نمایش داده می شود . ردیت در ماه جون ۲۰۰۵ توسط دو همکلاسی از دانشگاه Virginia تاسیس شد و اکنون بیش از ۳۶ میلیون کاربر عضو شده و ۲۳۱ میلیون بازدیدکننده ماهانه دارد.

اسنپ چت یک اپلیکیشن پیام رسان تصویری ست که توسط سه دانشجوی دانشگاه استنفورد Reggie Brown, Evan Spiegel و Bobby Murphy در سپتامبر سال ۲۰۱۱ راه اندازی شد و در مدت زمان اندکی و بر طبق آمار می ۲۰۱۵ ، تا کنون موفق به جذب ۱۰۰ میلیون کاربر شدند ، بیش از ۱۸ درصد کاربران تمام شبکه های اجتماعی از اسنپ چت استفاده می کنند.

WhatsApp یک نرم افزار پیام رسان برای تلفن های هوشمند ، رایانه ها و تبلت هاست ، این اپلیکیشن برای ارسال تصاویر ، متون ، صدا ، ویدئو و اسناد به دیگر کاربرانی که نرم افزار را بر روی سیستم شان دارند ، متکی به اینترنت است. این شرکت در ژانویه ۲۰۱۰ تاسیس شد و در فوریه ۲۰۱۴ به قیمت ۱۹,۳ میلیارد دلار توسط فیسبوک خریداری شد. امروزه بیش از یک میلیارد نفر در سراسر جهان برای ارتباط با دیگران از این نرم افزار استفاده می کنند.

سرمایه گذاری بر روی کنجکاوی بشر یک ایده هوشمندانه است که منجر به خلق و تاسیس Quora در جون ۲۰۰۹ شد . این وبسایت توسط دو تن از کارکنان سابق فیسبوک Charlie Cheever و Adam D'Angelo راه اندازی شد ، و حال مدعی ست که بیش از ۸۰ میلیون کاربر یکتا در ماه دارد که بیش از نیمی از آنها از کشور آمریکا هستند. کورا که یک شبکه اجتماعی براساس پرسش و پاسخ است بودجه سرمایه گذاری خود را به ۱۴۱ میلیون دلار رسانده و در حالی که هنوز عمومی نشده ، اما مسلما آینده دار و تماشایی خواهد بود.

با بیش از ۴۰ میلیون کاربر ، Vine یکی از شبکه های اجتماعی اشتراک گذاری ویدئوست که با سرعت در حال پیشرفت و رشد است. این شبکه به کاربران اجازه اشتراک گذاری ویدئوهای ۶ ثانیه ای با فالورهایشان می دهد، در حالی که به نظر ۶ ثانیه مدت کوتاهی برای یک ویدئوست ، کسب و کارهای مختلف در هر سائزی نتیجه شگرفی از این سرویس گرفته اند! شاید یکی از دلایل این موهم اینست که مردم تمایل و وقت کافی به تماشای ویدئوهای طولانی ندارند ، Vine در جون ۲۰۱۲ تاسیس و در اکتبر ۲۰۱۲ توسط تویتر خریداری شد! (قبل از راه اندازی رسمی)

Periscope پریسکوپ یک اپلیکیشن موبایلی برای تماشای آنلاین ویدئو ست که در فوریه ۲۰۱۴ توسط Joe Bernstein و Kayvon Beykpour تاسیس شد و در مارچ ۲۰۱۵ به قیمت ۱۰۰ میلیون دلار توسط تویتر خریداری شد. ۴ ماه بعد پریسکوپ اعلام کرد که که از مرز ۱۰ میلیون کاربر رد شده است و در ماه دسامبر همان سال بعنوان اپلیکیشن برتر سال توسط شرکت اپل انتخاب شد.

BizSugar یک شبکه اجتماعی و یک منبع مناسب برای صاحبان کسب و کارهای کوچک، کارآفرینان و مدیران است. این سایت در سال ۲۰۰۷ توسط شرکت DBH Communications (ارائه دهنده نشریات برگزیده کسب و کار) تاسیس شد و در سال ۲۰۰۹ توسط Small Business Trends LLC خریداری شد. این پلتفرم به کاربران اجازه میدهد به اشتراک ویدئوها، مقالات، پست های بلاگی و .. با دیگر کاربران بپردازند. همچنین به کاربران اجازه میدهد که مطالب دیگر کاربران را ببینند و به آنها امتیاز دهند.

وایبر یک اپلیکیشن پیام رسان اینترنتی و یک VoIP برای گوشی های هوشمند است که توسط Viber Media در دسامبر ۲۰۱۰ تاسیس و منتشر شد. این نرم افزار همچنین اجازه می دهد که صدا، متون، تصویر و ویدئو در بین کاربران به اشتراک گذاشته شود. در آوریل ۲۰۱۴ وایبر اعلام کرد که ۶۰۰ میلیون کاربر عضو شده و ۲۳۰ میلیون کاربر فعال ماهانه دارد.

۲-۲. شبکه های اجتماعی در ریاضی و کامپیوتر

برای شناسایی یک شبکه اجتماعی ابتدا باید ساختار آن را به شکل قابل فهم و محاسبه دریاوریم. ساختارهای قابل فهم در ریاضیات و کامپیوتر برای تحلیل و بررسی شبکه های اجتماعی مناسب می باشند. در این میان به صورت کلی دو روش برای نمایش شبکه های اجتماعی مورد استفاده قرار می گیرد؛ ماتریس و گراف .

۱-۲-۲ ساختار ماتریس برای مطالعه شبکه های اجتماعی :

ماتریس با چیدمانی مربعی از اعداد و ساختاری مرکب از اعدادی که به هم مرتبط هستند می باشد، با استفاده از ماتریس می توانیم شبکه های اجتماعی را به صورت آرایشی از اعداد نشان دهیم. در این صورت هر سطر یا ستون می تواند نماینده ای از گره ها باشد و ارتباط بین آن ها را با انتساب اعداد درون ماتریس نشان می دهیم. به عنوان مثال سه گره به نام های X و Y و Z را در نظر می گیریم، اگر a_{ij} را ارتباط بین دو گره بدانیم، در این صورت ارزش a_{ij} برابر با یک خواهد بود و در غیر اینصورت ارزشی برابر صفر خواهد داشت. این ساختار قابلیت نشان دادن ارتباط دو طرفه را دارد یعنی ماتریس می تواند متقارن باشد و یا نباشد. جدول شماره ۱ ساختار ارتباطی سه گره یا نود را در قالب ماتریس نشان می دهد.

جدول شماره ۱) نمایش ارتباط سه گره در شبکه اجتماعی در قالب ماتریس

گره ها یا نودها	X	Y	Z
X	—	۰	۱
Y	۱	—	۰
Z	۰	۱	—

در این ارتباط می بینیم که هر گره فقط با یک گره ارتباط یک طرفه دارد و اگر نخواهیم نزدیکترین فرد را به هرکس انتخاب کنیم بنابراین Y با X و X با Z و Z با Y به ترتیب ارتباط نزدیک یا دوستانه دارند. اگر ماتریس ما متقارن باشد در این صورت می توان ارتباطات را دو طرفه در نظر گرفت یعنی Y با X و X با Y ارتباط نزدیک دارند.

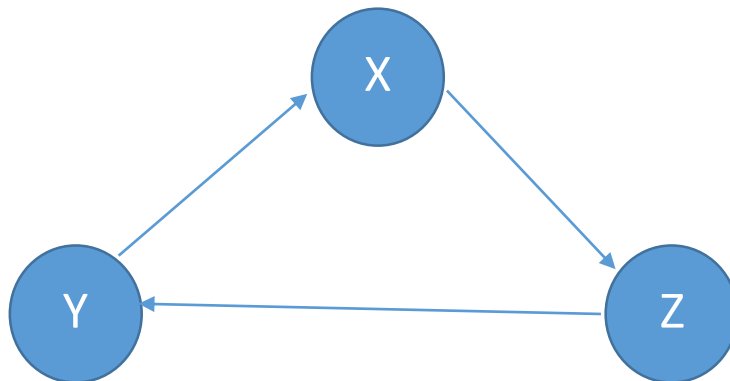
شبکه های اجتماعی مورد استقبال مانند Facebook و Twitter به ترتیب دو طرفه و یک طرفه می باشند و دلیل این انتخاب سنجش نسبت ارتباط دوطرفه به یک طرفه در دنبال کردن ها یا Following می باشد.

۲-۲-۲ ساختار گراف برای مطالعه شبکه های اجتماعی :

یکی دیگر از ابزارهای نمایش ریاضی ارتباطات اجتماعی گراف می باشد، به وسیله این روش نمایش می توان تصویری از ارتباطات را در شبکه های اجتماعی ایجاد نمود که کاربردی بسیار زیاد در بررسی و تحلیل شبکه های اجتماعی دارد.

در یک گراف گره ها یا نودها به عنوان رئوس و ارتباط بین آن ها به عنوان یال ها در نظر گرفته می شوند. در گراف ها نیز می توان از ارتباطات اجتماعی را به صورت یک طرفه یا دو طرفه نمایش داد.

شکل شماره ۱ ارتباط مثال شماره یک را بر اساس جدول شماره ۱ و در قالب گراف نشان می دهد.



شکل شماره ۱) گراف ارتباط سه گره یا نود در شبکه اجتماعی

۲-۳ انواع روش های شناسایی جوامع:

روش های مورد استفاده برای شناسایی جوامع به شرح زیر می باشند؛

۲-۳-۱ روش بخش بندی گراف (Graph Partitioning):

به آن روش افراز گراف نیز می گویند، این روش در ریاضیات بر روی تعدادی داده که شامل یال ها و رأس ها هستند قابل پیاده سازی است، از این روش می توان برای دسته بندی گراف به مؤلفه های کوچکتر که دارای ویژگی های خاص هستند استفاده نمود.

این نوع روش ها به روش های سنتی نیز معروف هستند، در این روش شبکه به محدوده هایی تقسیم می شود، بنابراین شبکه ها همپوشمانی ندارند. در این روش اعضای شبکه با تقسیم دودویی به وجود می آیند، بنابراین برای ایجاد سه شبکه یکی از جوامع تقسیم به دو می شود. یکی از محدودیت های این روش این است که تعداد جوامع و سائز آن ها باید از قبل معلوم باشد.

راه دیگر استفاده از روش سلسله مراتبی است، در این روش در هر مرحله دسته ها بر اساس معیارهای شباهت در هم ادغام و یا بر اساس معیارهای تفاوت از هم تفکیک می شوند، برخی از پارامترهای سنجش یا تفکیک عبارتند از رابطه پیرسون، شباهت کسینوسی، فاصله منهن (manhattan)، برش کمینه و بیشترین جریان. در این روش برای تشخیص بهترین سطح درخت از توابع کیفی استفاده می شود،

توابع کیفی تعیین کننده میزان بخش بندی گراف هستند که هر چه این مقدار بیشتر باشد نشان دهنده بخش بندی بهتر گراف است. از انواع توابع کیفیت عبارتند از $modularity$ ، $conductance$ ، $Normalized\ cut$ ، $Ratio\ cut$ ، $coverage$ ، $performance$.

یکی دیگر از روش های تقسیم به شبکه های مختلف الگوریتم Kernighan-lin است، میزان پیچیدگی این روش از مرتبه $O(n^2)$ است که تقریباً سریع می باشد و جزء الگوریتم های حریصانه است. در این الگوریتم گره ها بر اساس تابع سود به عضویت یک گروه در می آیند. یک ایراد که به این روش وارد است این است که باید تعداد اعضاء در هر گروه مشخص شود. اگر تعداد اعضاء مجهول باشد تقریباً غیرممکن است که بتوان به نتیجه رسید. یکی از کاربردهای مورد قبول این الگوریتم مقایسه صحت روش های جدید و روش های قبلی است، زیرا در این حالت تعداد اعضاء مشخص می باشد.

یکی دیگر از روش ها استفاده از الگوریتم معرفی شده برای حداقل کردن میزان برش ($Minimum\ cut$) و درخت پوشای بهینه برای ایجاد گروه های دوتایی می باشد. میزان پیچیدگی این روش از درجه $O(n^2 \log n)$ است. این روش در گراف های وزن دار که دارای وزن مثبت هستند به کار می رود. این روش نیز محدودیت هایی دارد که برای استخراج جوامع موجود مناسب نیستند.

۲-۳-۲ روش های مبتنی بر حذف یال :

در این روش ها سعی می شود یال ها را شناسایی و سپس به حذف آن ها می پردازند. یعنی در این روش ها بر خلاف روش قبلی که مبتنی بر شباهت و یا تفاوت گره ها بود، با استفاده از یال های درون خوشه ها به شناسایی گروه ها می پردازند.

در این میان شناسایی یال ها و این که چگونه می توان یال های میانی را انتخاب نمود موضوع مورد چالش می باشد. معمولاً در این روش ها از پارامترهای میانی مانند ($Randomwalk\ Edge\ | Current\ flow\ edge$) استفاده می شود. دلیل استفاده از این پارامترها نتایج بهتری بوده که در استفاده از این روش ها در مقایسه با سایر روش ها به دست آمده است.

منظور از میانی بودن یک یال یعنی تعداد مسیرهای جغرافیایی که بین دو رأس یال وجود دارد، به عنوان مثال کوتاهترین مسیر موجود بین دو یال مورد نظر. یک نظریه در خصوص دلیل انتخاب آن ها این است که معتقدند یال های دارای ترافیک بالا معمولاً گلوگاه های ارتباطی بین دو مسیر می باشند. البته این نظریه در تمام موارد موجود صدق نخواهد کرد.

برای رسیده به نتیجه، سه مرحله زیر را در نظر می گیریم؛

۱. محاسبه امتیاز پارامتر انتخاب شده ($Randomwalk\ Edge\ | Current\ flow\ edge$) برای تمام یال ها.

۲. حذف یال دارای بیشترین امتیاز

۳. تکرار مراحل قبلی تا جایی که به تعداد کافی جوامع ایجاد شود.

۲-۳-۳ روش خوشه بندی :

روش خوشه بندی یا Clustering نیز یکی از روش های تشخیص جامعه می باشد، خوشه بندی یعنی جدا کردن مناطق دارای تراکم بالا در مجموعه خصوصیات که توسط مناطق با تراکم کم از هم جدا می شوند. این تعبیر خوشه بندی برای شناسایی جوامع بسیار مناسب می باشد. روش های خوشه بندی به دو دسته سلسله مراتبی و بخش بندی تقسیم می شوند.

در روش سلسله مراتبی نیاز به ماتریس الگو یا ماتریس شباهت داریم، تقسیم به خوشه ها می تواند به صورت بازگشتی صورت پذیرد. به این صورت که به دو شکل تجمعی و تجزیه ای، یعنی در ابتدا کل خوشه ها که در شکل تجزیه ای، کل شبکه بوده اند، در هر مرحله به صورت بازگشتی تجزیه شده و مرحله بعدی را می سازند و به همین ترتیب در شکل ترکیب، که کل خوشه در ابتدا یک نود یا گره بوده در هر مرحله با ترکیب با سایر نودها مرحله ای جدید را آغاز می نماید.

از محاسن این روش این است که ما را از دانستن تعداد گروه ها به عنوان ورودی بی نیاز می کند اما نمی دانیم درخت ایجاد شده در کدام مرحله یا عمق بهترین نتیجه را به ما برمی گرداند. ایجاد دسته های کوچک خوشه ها از دیگر معایب این روش است، این دسته های کوچک معمولاً اطلاعات مهمی را در اختیار ما قرار نمی دهند. الگوریتم های شایع این روش شامل single-linkage و complete linkage می باشند.

در مدل بخش بندی به دانستن سه پارامتر نیاز داریم؛

۱. تعداد گروه ها
۲. بخش بندی اولیه مجموعه
۳. تابع شباهت برای مقایسه و انتخاب اعضای هر خوشه

یکی از پارامترهای مورد استفاده در این روش K-means می باشد که می توان از انواع فازی، همپوشانی و ... استفاده نمود. از جمله مشکلات این روش وابستگی نتایج به گروه بندی اولیه نقاط مرکزی و همچنین تعداد مشخص جوامع (K) که باید به عنوان ورودی به الگوریتم داده شود.

روش خوشه بندی مدل های دیگری نیز دارد که تحلیلگر با توجه به آنچه که نیاز دارد می تواند بر حسب تراکم بالای جوامع و یا ... از آن ها استفاده نماید. برای بهبود کیفیت خوشه بندی می توان از چندین روش برای تفکیک جوامع و یا از یک روش با پارامترهای ورودی متفاوت استفاده نمود.

۲-۳-۴ استفاده از احتمالات و روش های بیزین:

نه تنها شناخت گره ها در شبکه های اجتماعی بلکه اسناد و اطلاعات نیز گاهی نیاز به دسته بندی و شناسایی دارند، این مهم توسط استفاده از متن برای شناخت متونی که توسط بازدیدکنندگان جوامع مجازی مورد استقبال قرار گرفته است محقق می شود. استفاده از متن و برچسب زدن جوامع بر اساس محتوای متون و محتوای مشترک بین افراد و همچنین روابط آشکار بین مستندات بستر مناسبی را برای شناخت جوامع معنایی ایجاد کرده است.

در روش شناخت جوامع معنایی می توان گفته که هر سند از محتوایی تشکیل شده است که شامل چندین واژه می باشد، از طرفی نیز می توان نتیجه گرفت که هر موضوع نیز از تعدادی از واژگان معنا و هویت می گیرد. اساس شناسایی و جداسازی جوامع معنایی تخصیص هر نویسنده، نوشته یا سند به یک موضوع خاص می باشد. کاربرد این روش به این صورت است که می توان با شناسایی موضوعات محبوب مطرح شده در ایمیل ها را شناسایی و جوامع را بر اساس آن ها دسته بندی نموده و استخراج کرد. تشخیص به این روش می تواند به دلیل جدا بودن لیست فرستندگان هرنامه ها از فرستندگان عادی به شناسایی هرنامه ها و در نتیجه ایجاد جوامع منسجم تر کمک شایانی نماید.

در روش ساختاری با استفاده از روش دیریکله و پارامترهای دیریکله و همچنین توزیع چند جمله ای می توان به تخمین احتمال وجود یک کاربر در جامعه پرداخت. در روش احتمالاتی دیگر افراد بر اساس موضوعات مشترک دسته بندی می شوند، فارغ از اینکه افراد به صورت مستقیم با هم ارتباط برقرار کرده باشند. به اینصورت که از نمونه برداری Gibbs که در مطالعات آماری الگوریتمی مبتنی بر زنجیره مارکوف مونا کارلو است استفاده می شود. کاربرد این الگوریتم تولید دنباله ای از مشاهدات از یک تابع توزیع احتمالاتی چند متغیره است که تولید نمونه از آن بصورت مستقیم دشوار است. در این روش برای هر واژه در ایمیل های کاربر یک موضوع یا جامعه به صورت تصادفی انتخاب می شود سپس در ادامه بر اساس همگرایی زنجیره مارکوف در تعدادی مراحل تکرار هر کاربر بر اساس موضوعات مربوطه به عضویت جامعه ای درمی آید.

در مقالات مختلف ترکیبی از روش های مختلف نیز استفاده شده است، به عنوان مثال ترکیبی از چندین روش و الگوریتم تشخیص جامعه را با هم به کار ببریم.

۲-۳-۵ سایر روش ها:

۳,۳,۱. روش طیفی: به دلیل کیفیت بالای این روش بسیار مورد توجه بوده است و در تحقیقات بسیار مورد استفاده قرار گرفته است. این روش در مقایسه با روش شباهت از کیفیت بالاتری برخوردار است. این روش ها با استفاده از ماتریس های شباهت به جداکردن جوامع می پردازد. روش کار به این صورت است که تعداد K عدد از نقاط به عنوان نقاطی از یک فضای K بعدی به عنوان گره های موجود در شبکه تعریف می شوند و سپس الگوریتم های کلاسیک خوشه بندی جوامع آن ها را دسته بندی

می کنند. این روش کیفیت بالایی در جداسازی جوامع در مقایسه با روش های خوشه بندی دارد اما سربار محاسباتی آن بالا می باشد.

۲-۳-۵ روش های سلسله مراتبی:

در این تکنیک، لبه های با محدوده بالا یکی پس از دیگری حذف می شوند به طوری که نمودار به صورت سلسله مراتبی به خوشه ها تقسیم می شود. روش خوشه بندی سلسله مراتبی *agglomerative* را برای شناسایی جوامع پایدار یا طبیعی در شبکه های ارتباطی بزرگ ارائه داد. "یک خوشه به عنوان یک خوشه طبیعی فرض می شود، اگر در فرآیند خوشه بندی زمانی که درصد معینی از پیوندها حذف می شود." خوشه بندی سلسله مراتبی *Agglomerative* یک روش خوشه بندی به صورت پایین به بالا است که در آن خوشه ها دارای زیر خوشه ها هستند که به نوبه خود دارای زیر خوشه ها و غیره هستند.

۳ فصل سوم

۳-۱ نگاهی به الگوریتم های مورد استفاده در شناسایی جوامع

جوامع دارای سابقه طولانی در علوم اجتماعی هستند، الگوریتم های مختلفی در سال های اخیر توسط رشته های مختلفی نظیر ریاضیات، یادگیری ماشین، آمار، داده کاوی، برای تشخیص جوامع در شبکه های اجتماعی و دیگر شبکه های پیچیده پیشنهاد شده اند. در ادامه برخی از الگوریتم های شناخته شده به طور خلاصه مورد بررسی قرار می گیرد. گیروان و نیومن ابتدا ایده ساختار جامعه در شبکه های اجتماعی را بیان کردند. گیروان و نیومن ابتدا ایده ساختار جامعه در شبکه های اجتماعی را بیان کردند. به گفته گیروان و نیومن، بسیاری از شبکه های واقعی دارای زیرگرافی از گره ها هستند که به صورت گروه دیده می شود به طوری که تعداد ارتباطات داخلی بین گره های زیر گراف بزرگتر از اتصالات با گره های باقی مانده در شبکه است. یک روش تقسیم سلسله مراتبی توسط آنها پیشنهاد شده است. در این تکنیک، لبه های با محدوده بالا یکی پس از دیگری حذف می شوند به طوری که نمودار به صورت سلسله مراتبی به خوشه ها تقسیم می شود. *hopcroft* روش خوشه بندی سلسله مراتبی *agglomerative* را برای شناسایی جوامع پایدار یا طبیعی در شبکه های ارتباطی بزرگ ارائه داد. بر طبق نظر *Hopcroft* (۲۰۰۳)، "یک خوشه به عنوان یک خوشه طبیعی فرض می شود، اگر در فرآیند خوشه بندی زمانی که درصد معینی از پیوندها حذف می شود".

خوشه بندی سلسله مراتبی *Agglomerative* یک روش خوشه بندی به صورت پایین به بالا است که در آن خوشه ها دارای زیر خوشه ها هستند که به نوبه خود دارای زیر خوشه ها و غیره هستند

در شبکه های پیچیده، یک روش تشخیص جامعه مبتنی بر *k-means* یا میانگین *k* می باشد، در این روش پس از بررسی الگوریتم تشخیص جامعه در شبکه های پیچیده، یک الگوریتم بهبود یافته بر اساس *k-means* ارائه می شود. در شبکه های اجتماعی برای شناسایی جوامع، نویسندگان دیگری روش خوشه بندی طیفی ارائه دادند. در شبکه های اجتماعی برای شناسایی جوامع، آن نویسندگان روش

خوشه بندی طیفی را ارائه دادند. در این روش، برای استخراج جوامع، نویسندگان از اعضای اصلی استفاده می کنند. نویسندگان از الگوریتم رتبه بندی صفحات برای تشخیص جوامع برای استفاده کامل از ویژگی های شبکه استفاده کردند و ثابت کردند که روش آنها از لحاظ زمان و دقت بهتر است.

به طور کلی شناسایی جوامع در شبکه های اجتماعی از دو رویکرد قابل رسیدن است، رویکرد مبتنی بر توپولوژی و رویکرد مبتنی بر موضوع. با این حال، در بیشتر مقالات پیشنهاد می شود که در تشخیص جامعه باید ساختار گراف و اطلاعات متنی شبکه را در نظر بگیرد، زیرا جوامع شناسایی شده توسط روش های مبتنی بر توپولوژی تمایل دارند موضوعات مختلف در هر جامعه را دربرگیرند، درحالیکه زیر-انجمن های معنی دار مبتنی بر توپولوژی درون جوامع مبتنی بر موضوع قرار دارند. بعلاوه روش های مبتنی بر توپولوژی با اینکه با موفقیت اقدام به شناسایی جوامع می کنند اما هیچگونه اطلاعاتی در خصوص موضوع بحث بین کاربران جوامع ندارند. بسیاری از روش های شناسایی جوامع با ترکیب توپولوژی ارتباطات جوامع و ویژگی های موضوعات اقدام به شناسایی و استخراج موضوعات بامعنی بین کاربران می کنند.

در یکی از تحقیقات با استفاده از الگوریتم های خوشه بندی اقدام به دسته بندی موضوعات کرده است و سپس با استفاده از تحلیل لینک های خوشه ها اقدام به شناسایی جوامع می کنند. در حالیکه از هر دو حالت شباهت موضوعی بین کاربران و ساختار ارتباطی یا پیوند برای اندازه گیری تأثیر کاربران در توئیتر استفاده شده است. در تحقیق دیگر برای شناسایی جوامع توئیتر از روش از علایق کاربر با استفاده از مفهوم شباهت مقادیر متنی و ساختار اجتماعی با استفاده از ساختار خوشه بندی کلاسیک استفاده شده است.

یک چارچوب مبتنی بر یادگیری که می تواند کاربران را بر اساس شباهت محتوای تولید شده در توئیتر شناسایی کند. در مقاله ای دیگر وظیفه شناسایی مجموعه ای از گره هایی که بیشترین تأثیر را بر گره معین و یک موضوع معین دارند معرفی شده است، که نتیجه آن گره هایی است که منبع یا تریگر را برای هر موضوع نشان می دهند.

در یک مقاله دیگر توصیف و طبقه بندی وبلاگ ها و متون کوتاه با استفاده از تحلیل آن ها و شناسایی ویژگی هایی که می توانند آن ها را تشخیص دهند توسعه داده شده است. در یک تحقیق دیگر، نویسنده یک روش واحد را برای دسته بندی کاربران تلویزیون و کلمات توصیف کننده برنامه تلویزیونی که به صورت همزمان مورد استفاده قرار می گیرند مورد استفاده قرار داده است.

گروه دیگری از نویسندگان پیشنهاد داده اند که مطالب و محتوای ارسال شده بر روی سرویس های اجتماعی وب به صورت خودکار اقدام به تخمین شخصیت افراد از روی متن نموده و توصیه های ارزشمندی را متناسب با شخصیت افراد تولید نماید. سایر موارد شامل همپوشانی روش های شناسایی جوامع است و یا یکسان سازی دیدگاه ها است، به عنوان مثال ساختار لبه و مدل ویژگی گره.

در یکی از مقالات یک روش جدید برای کشف معیار فاصله بین کاربران که امکان ترکیب با الگوریتم های خوشه بندی مبتنی بر فاصله را دارد برای شناسایی جوامع در توئیتر ارائه گردیده است. در ابتدا، تعدادی از معیارهای تشابه را بر اساس خواص شبکه توئیتر مانند زیر روابط، هشتک ها و غیره تعریف کرده و در ادامه این معیارها برای محاسبه فاصله (یا معکوس شباهت) بین هر جفت کاربر استفاده می شود. نتیجه این فرآیند یک ماتریس شباهت است که برای دسته بندی کاربران به جوامع با استفاده از گسترش ضعیف، یک الگوریتم

خوشه بندی مبتنی بر شباهت استفاده می شود. سپس با استفاده از یک روش جدید به صورت خودکار موضوعات خوشه ها را با استفاده از مدل تخصیص پنهان دیریکله اقدام به حذف موضوعات بیهوده و شناسایی موضوعات جالب توجه در جوامع توئیتر نموده. این رویکرد بر روی گروهی از کاربران علاقمند به برنامه نویسی با استفاده از یک معیار خوشه بندی با کیفیت بررسی شده است.

۲-۳ شناسایی و دسته بندی موضوعات در توئیتر به عنوان یک نمونه از شبکه های اجتماعی :

در این بخش ما یک روش جدید برای شناسایی جوامع معنی دار ارائه می دهیم. این روش شامل دو مرحله است. ابتدا مفهوم شباهت کاربر بین کاربران توئیتر را تعریف می کنیم و فاصله بین هر جفت کاربر را محاسبه می کنیم. گام دوم شامل فرآیند خوشه بندی کاربران به جوامع معنی دار براساس فاصله آنها می شود. شباهت بین یک جفت کاربر توئیتر از تعاملات ثبت شده در تاریخچه توئیتر آن ها مشتق شده است. بنابراین شباهت را می توان بر اساس محاسبه تمام عناصر قابل اندازه گیری یا شرح تعامل هر کاربر با سایر کاربران منظور نمود: لیست دوستان و (Followers) دنبال کننده ها، میزان هشتگ های قرار داده در توئیت شان و میزان پاسخ به کاربران دیگر که در توئیت خودشان ذکر شده است.

۱-۲-۳ دنبال کردن شباهت روابط :

روش بدیهی برای تعیین میزان مشابه دو کاربر، با بررسی رابطه زیر است. می دانیم که نمودار توئیتر جهت دار است، به این معنی که یک کاربر می تواند کاربر دیگر را دنبال کند بدون اینکه توسط او دنبال شود. اگر u_i و u_j دو کاربر توئیتر باشند، به این ترتیب میزان شباهت این دو کاربر توئیتر بر اساس دنبال کردن یکدیگر به روش زیر به دست می آید: (رابطه شماره ۱)

$$S_1(u_i, u_j) = \begin{cases} 1, & \text{if } u_i \text{ follows } u_j \\ & \text{and } u_j \text{ follows } u_i \\ 0.5, & \text{if only one of the users} \\ & \text{follows the other} \\ 0, & \text{otherwise} \end{cases}$$

۳-۲-۲ شباهت (Followers) دنبال کنندگان مشترک :

جدول زمان بندی کاربران توئیتر یک جریان از توئیت ها را از بین حساب هایی که کاربر دنبال می کند را نشان می دهد، به این ترتیب که افراد معمولاً بر اساس علایق شان اقدام به پیگیری و دنبال کردن سایر حساب ها می کنند. بنابراین، انتظار میرود که کاربرانی که علاقه مند به موضوعات مشابه هستند تعدادی پیروان مشترک داشته باشند. در نتیجه، ما می توانیم شباهت کاربر را براساس دنبال کنندگان مشترک محاسبه کنیم، همانطور که در رابطه ۲ نشان داده شده است، به شکلی که دنبال کننده Si از مجموعه دنبال کنندگان Ui (Followers) است و n تعداد کاربران است. (رابطه شماره ۲)

$$S_2(u_i, u_j) = \frac{|followers_i \cap followers_j|}{\max_{1 \leq l \leq n} followers_l}$$

۳-۲-۳ شباهت (Friends) دوستان مشترک :

در توئیتر مجموعه ای از کاربرانی که یک کاربر دنبال می کند را دوستان آن کاربر می نامیم، همانطور که قبلاً نیز گفتیم می توانیم اینگونه فرض کنیم که کاربران مشترک دارای دوستان یکسانی هستند، بنابراین می توانیم تشابه کاربری را بر اساس رابطه زیر تعریف کنیم. (رابطه شماره ۳)

$$S_3(u_i, u_j) = \frac{|friends_i \cap friends_j|}{\max_{1 \leq l \leq n} friends_l}$$

۳-۲-۴ شباهت هشتگ

استفاده از نماد # بدون فاصله قبل از یک کلمه یا عبارت هشتگ نامیده می شود. هدف از آن دسته بندی توئیت ها و به دنبال آن ساده کردن جستجوی توئیت های مرتبط توسط کاربران است. برای محاسبه هشتگ ابتدا باید ارزش و اهمیت هر کلمه کلیدی را برای هر کاربر خاص تخمین بزنیم. این را می توان با محاسبه وزن های tf-idf به دست آورد، این را باید در نظر گرفت که تمام هشتگ ها در توئیت های یک کاربر یک سند واحد را تشکیل می دهد. به این ترتیب اگر hi بردار tf-idf هشتگ های استفاده شده توسط کاربر Ui باشد، و تابع \cos نیز شباهت بین دو بردار را حساب می کند، محاسبه آن به صورت زیر می باشد؛

$$S_4(u_i, u_j) = \cos(h_i, h_j)$$

۳-۲-۵ شباهت پاسخدهی

یک کاربر تویتر می تواند با پاسخ دادن به توییت های دیگر کاربران، به یک مکالمه پیوندد. فراوانی پاسخ ها بین دو کاربر، و همچنین تعداد کاربرانی که هر دو کاربر به آنها پاسخ داده اند، دو شاخص از شباهت پاسخ کاربران است، که در معادله ۵ تعریف شده است. R_i و R_j مجموعه کاربرانی هستند که کاربران U_i و U_j به آن ها پاسخ داده اند. به ترتیب $|R_i \cap R_j|$ مجموع تعداد کاربرانی است که هر دو به آن ها پاسخ داده اند، nr_{ij} تعداد دفعاتی است که U_i به U_j پاسخ داده است و NR_i تعداد دفعاتی است که U_i به توییت سایر کاربران پاسخ داده است. (رابطه شماره ۵)

$$S_5(u_i, u_j) = \frac{|R_i \cap R_j|}{\sqrt{|R_i|}\sqrt{|R_j|}} + \frac{nr_{ij} + nr_{ji}}{NR_i + NR_j}$$

۳-۲-۶ شباهت بیان های کاربران

با قرار دادن نماد @ در مقابل یک نام کاربری، کاربر می تواند یک کاربر دیگر را در توییت تگ کند یا علامت بزند. یک بیان (mention) هرگونه به روز رسانی شامل "username@" در هرجای از بدنه توییت است که به معنی این است که پاسخ ها نیز ذکر شده اند. شباهت بیان های کاربران از طریق زیر حساب می شود؛ (رابطه شماره ۶)

$$S_6(u_i, u_j) = \frac{|M_i \cap M_j|}{\sqrt{|M_i|}\sqrt{|M_j|}} + \frac{nm_{ij} + nm_{ji}}{NM_i + NM_j}$$

M_i و M_j مجموعه ای از کاربران هستند که U_i و U_j در توییت هایشان آن ها را بیان کرده اند. nm_{ij} تعداد دفعاتی است که U_i ، U_j را بیان کرده است و NM_i تعداد دفعاتی است که U_i سایر کاربران را بیان کرده است.

۳-۲-۷ شباهت کاربران کل

برای به دست آوردن شباهت کلی کاربر، این اقدامات باید ترکیب شوند، در رویکرد ما، یک ترکیب خطی از مقادیر تشابهات فردی را به کار گرفته ایم، همانطور که در معادله زیر نشان داده شده است:

$$S(u_i, u_j) = \sum_{m=1}^6 a_m S_m(u_i, u_j)$$

مقادیر اختصاص داده شده به پارامترهای a_m بین ۰ و ۱ است و مجموع $a_1 + a_2 + a_3 + a_4 + a_5 + a_6 = 1$ است. مقادیر از طریق اعتبار سنجی به دست می آید، همانطور که در بخش ۳-۴ توضیح داده شده است.

۳-۲-۸ خوشه بندی کاربران بر اساس انتشار وابستگی

گام نهایی از تشخیص جامعه خوشه بندی کاربران برای تشکیل جوامع با در نظر گرفتن معیارهای شباهت برای هر جفت از کاربران است. به خوبی شناخته شده است که طیف گسترده ای از الگوریتم های خوشه بندی وجود دارد. از آنجا که موقعیت کلی نقاط داده در دسترس نیست، الگوریتم انتخاب شده برای اندازه گیری شباهت بین جفت نقاط داده را به عنوان ورودی نیاز دارد. انتشار ضعیف (انتشار وابستگی) [۵] یک الگوریتمی است که ملاک ها در میان نقاط داده را شناسایی کرده و خوشه ای از نقاط داده در اطراف این ملاک ها تشکیل می دهد. به طور همزمان تمامی داده ها به عنوان نمونه های بالقوه مورد توجه قرار می گیرد و به طور تکراری پیام های بین نقاط داده را مبادله می کند تا زمانی که یک مجموعه ی خوب از نمونه ها و خوشه ها ظاهر شود. این انتشار (انتشار ضعیف) به عنوان ورودی مجموعه ای از مقادیر واقعی از شباهت های بین نقاط داده را دریافت می کند، به طوری که تشابه $s(i, k)$ نشان دهنده این است که چه میزان نقطه داده با اندیس k برای نقطه داده با اندیس i مناسب است. تعداد نتیجه خوشه ها از پیش تعیین نشده است. اگر تصمیمات برای نمونه ها و مرزهای خوشه ای برای تعدادی تکرار بدون تغییر باقی بماند یا اگر حداکثر تعداد تکرارها به دست آید الگوریتم متوقف می شود. یک راه جایگزین می تواند روش خوشه بندی سلسله مراتبی با استفاده از راه جداسازی متناسب با تراکم برای جداسازی اتوماتیک خوشه ها است.

۳-۳ استخراج موضوع و برجسب زدن :

در این بخش ، برای اینکه بدانیم که آیا کاربرانی که به یک خوشه متعلق هستند، تمایل به گفتگو درباره موضوعات مشابه دارند محتوای مشترک کاربران را مطالعه می کنیم. به شکل خاص ما اقدام به توضیح روند استخراج موضوعات بحث شده در بین کاربران می پردازیم و وروشی را برای حذف موضوعات نامربوط و غیرجالب بیان می کنیم و روندی را برای تولید برجسب به صورت خودکار برای کلمات کلیدی هر موضوع بیان می کنیم. علاوه بر آن فرایندهای ترکیبی استخراج موضوعات و برجسب گذاری را می توان به عنوان آخرین روش جدید کشف موضوع در نظر گرفت. از آنجا که برجسب های موضوعات ایجاد شده لزوماً در تئیت های خوشه ها قرار نمی گیرند.

۳-۳-۱ استخراج موضوعات بر اساس تخصیص پنهان دیریکله (LDA)

در مرحله پیش پردازش، موجودیت هایی مانند هشتک ها، اشارات کاربر و URL ها از متن هر تئیت حذف می گردد. سپس، کلمات با استفاده از نشانه گذاری بازایی می شوند، کلمات متوقف شده حذف می شوند و کلمات با اصطلاح به کلمات ریشه آنها کاهش می

یابد. در نهایت، کلمات از تمام تویت های منتشر شده توسط یک کاربر، درون یک سند جمع می شوند. این فرآیند برای همه کاربران تکرار می شود، بنابراین می توانیم فرض کنیم که کل مجموعه داده ما مجموعه ای از اسناد است.

استخراج موضوعات می تواند توسط تخصیص نهفته دیریکله انجام شود، که مدل احتمالی از ساختاری است که بر اساس آن اسناد بر اساس ترکیب اتفاقی موضوعات پنهان به طوری که هر موضوع دارای یک توزیع احتمال در سراسر کلمات می باشد بیان می شود. مجموعه ای از ابزارها و کتابخانه هایی که تکنیک LDA را پیاده سازی می کنند وجود دارد، ما از MALLET که یک نرم افزار مبتنی بر جاوا و دارای قابلیت یادگیری ماشین به صورت متن می باشد استفاده کردیم. این بسته نرم افزاری دارای قابلیت پیاده سازی تخصیص پنهان دیریکله است. با استفاده از الگوریتم LDA در مجموعه اسناد کاربر، مجموعه ای از موضوعات N (و توزیع کلمات یا کلید ها در هر موضوع) و توزیع موضوعی مربوطه برای هر سند، توزیع موضوعی کاربر (UTD) را بدست می آوریم: (رابطه ۸)

$$UTD_i = [topic_{1,u_i}, topic_{2,u_i}, ..., topic_{N,u_i}]$$

در این رابطه $topic_{p,u_i}$ ارزش احتمالی موضوع p در توزیع موضوع u_i است. هدف ما از این رابطه یافتن توزیع موضوع در هر خوشه با در نظر گرفتن علائق کاربر است. بنابراین ما می توانیم اینطور در نظر بگیریم که تمام کاربران در یک خوشه یک سند را تشکیل می دهند، اسناد خوشه را ادغام می کنند و در نهایت یک مجموعه جمع بندی شده سند خوشه را ارائه می دهند. با محاسبه توزیع موضوع این مجموعه خوشه، در حالی که موضوعات مشابه با مجموعه کاربران را نگه داشته ایم، توزیع هر موضوع برای هر خوشه به دست می آید (۲)، توزیع موضوع محلی به صورت زیر خواهد بود: (رابطه ۹)

$$LTD_r = [topic_{1,C_r}, topic_{2,C_r}, ..., topic_{N,C_r}]$$

در این رابطه $topic_{p,C_r}$ میزان احتمال از موضوع p در سند مربوط به کلاستر (خوشه) C_r است.

۳-۲-۳ شناسایی جوامع در شبکه های اجتماعی با استفاده از تکنیک خوشه بندی:

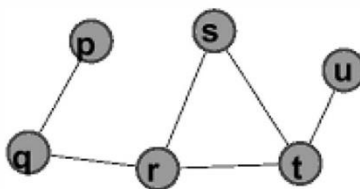
در خوشه بندی نقاط داده، الگوریتم خوشه بندی طیفی با کاهش ابعاد روی مقادیر خاص (طیف) به دست آمده از ماتریس شباهت دسته بندی را انجام می دهد. سپس خوشه بندی را در ابعاد کمتر انجام می دهد. ورودی الگوریتم، ماتریس شباهت است. ماتریس شباهت شامل تشابه نسبی بین هر جفت نقطه در مجموعه داده می باشد. این رابطه نسبی به صورت کمی ارزیابی می شود. با توجه به یک مجموعه D از نقاط داده، ماتریس شباهت ممکن است به عنوان یک ماتریس متقارن W تعریف شود. $W_{jj} = 0$ نشان دهنده اندازه گیری شباهت

بین نقاط داده با شاخص های i و j است. ماتریس لاپلاس گراف لایه نگاشته شده به صورت $L = D - W$ تعریف شده است. بقیه الگوریتم در بخش C این بخش شرح داده شده است.

۳-۲-۱ شرح مشکل

در شبکه ها، هدف تشخیص جامعه، یافتن یک پارتیشن $C = \{c_1, c_2, \dots, c_k\}$ از یک گراف ساده $G = (V, E)$ است، به طوری که V_i ، $C_i \subset V$ و $C_i \cap C_j = \emptyset$ (i, j = 1, ..., k) است. هر C_i یک زیر گراف است که حاوی گروهی از رأسهای G می باشد. این زیر نمودار، C "به عنوان یک جامعه شناخته شده است به گونه ای که چگالی درون خوشه ای لبه ها در زیر نمودار، چگالی بالا و بین خوشه ای لبه ها کم است. در زیر، فرض کنید ما یک گراف بدون جهت و بدون لبه های متعدد است. حرف های i ، j نشان گره ها؛ $e(i, j)$ نشان دهنده یک لبه که اتصال گره های i و j را انجام داده است می باشد. نمایش ماتریس مجاورت نمودار برای حل کردن مشکلات گراف مورد استفاده قرار می گیرد. مجاورت ماتریس A از یک گراف G ، توسط یک ماتریس $N \times N$ که حاوی صفر و یک ها است نشان داده می شود.

$A = (a_{ij})_{n \times n}$ به صورتی که یک لبه بین رئوس i و j وجود داشته باشد $a_{ij} = 1$ و در غیر این صورت $a_{ij} = 0$. ماتریس مجاورت برای یک گراف بدون جهت متقارن است. شبکه های اجتماعی نمونه در شکل ۲ نشان داده شده است. و ماتریس مجاورت آن در جدول شماره ۱ نشان داده شده است.



شکل شماره ۲) یک شبکه اجتماعی

جدول شماره ۱) ماتریس مجاورت شبکه اجتماعی شکل شماره ۲

Node	p	q	r	s	t	u
p	0	1	0	0	0	0
q	1	0	1	0	0	0
r	0	1	0	1	1	0
s	0	0	1	0	1	0
t	0	0	1	1	0	1
u	0	0	0	0	1	0

۳-۲-۲ الگوریتم پیشنهاد شده

در زیر شبه کد الگوریتم خوشه بندی طیفی غیر عادی است که با K-Means Bisecting گنجانده شده است.

- Construct a similarity graph. Let W be its weighted adjacency matrix.
- Compute the unnormalized Laplacian L .
- Compute the first k eigenvectors u_1, u_2, \dots, u_k of L .
- Let $U \in R^{n \times k}$ be the matrix containing the vectors u_1, u_2, \dots, u_k as columns.
- For $i = 1, \dots, n$, let $y_i \in R^k$ be the vector corresponding to the i -th row of U .
- Cluster the points $(y_i)_{i=1, \dots, n}$ in R^k with the **Bisecting k-means** algorithm into clusters C_1, \dots, C_k .

Output: Clusters P_1, \dots, P_k with $P_i = \{j / y_j \in C_i\}$.

In Bisecting k-means initially all objects are placed in single cluster. The mixture of k-means technique and hierarchical clustering technique is the Bisecting k-means.

The pseudo code of the bisecting k-means algorithm is:

Bisecting K-means pseudo code for finding k clusters

- Initialize with one cluster containing all data points
- Pick a cluster randomly to split.
- Use k-means algorithm to divide the cluster into 2 sub-clusters.
(*Bisecting step*)
- Repeat for $ITER$ times
the bisecting step and pick a partition whose overall similarity is high such that it produces the clustering.
- Repeat above three steps until the desired number, k , of clusters is obtained.

۳-۴ نتایج تجربی

در این بخش، نتایج به دست آمده برای برنامه های کاربردی مختلف ارائه شده است که خوشه بندی طیفی غیر متمرکز ارائه شده توسط ما با استفاده از تقسیم میانگین K به دست آمده است. این الگوریتم بر روی داده های مجموعه بازی های زاکاری کاراته، کالج آمریکایی فوتبال و مجموعه داده های شبکه بولتنس دلفین اعمال شده است. در هر یک از این موارد، ما دریافتیم که الگوریتم ما ساختارهای جامعه را به روش قابل اعتماد تشخیص می دهد. جدول شماره ۲ برخی از آمار این سه مجموعه داده را ارائه می دهد.

جدول شماره ۲) آمار آزمایشات تجربی

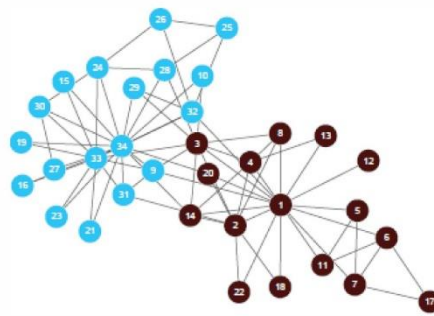
Dataset	V	E	<k>	C	D
Zachary Karate Club	34	78	4.58	0.56	5
American College Football Network	115	616	10.6	0.40	4
Dolphin Network	62	159	5.13	0.26	8

<k> - average degree of the dataset;
 C - clustering coefficient of the dataset;
 D - diameter of the dataset

برای آزمودن، این الگوریتم ده بار برای هر مجموعه داده تکرار شده است و مقادیر به صورت میانگین برای هر ده باز اندازه گیری گزارش شده است.

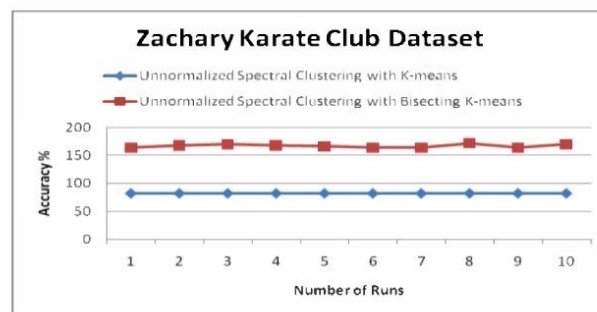
۳-۴-۱ کلوب Zachary Karate

در اوایل دهه ۱۹۷۰، وین زاگری دریک دانشگاه آمریکایی، بر روی اعضای یک باشگاه کاراته به مدت دو سال مطالعه و تعاملات اجتماعی آن ها را ثبت کرد. بر اساس تعاملات اجتماعی آن ها، او یک مجموعه داده شبکه با ۳۴ رأس و ۷۸ لبه ایجاد کرد. در این مجموعه داده ها، دانش آموزان به صورت رأس ها مشخص شده و دو دانش آموز اگر با هم دوست خوب باشند با یک لبه به هم وصل می شوند. به صورت اتفاقی یک مشاجره بین مدیر کلوب و مدرس کاراته در طول مطالعه آن ها رخ داده است. در نتیجه کلوب به دو جامعه کوچکتر به ترتیب با محوریت مدیر کلوب و مدرس کاراته تقسیم می شود. تقسیم باشگاه به دو جامعه در شکل شماره ۳ نشان داده شده است.

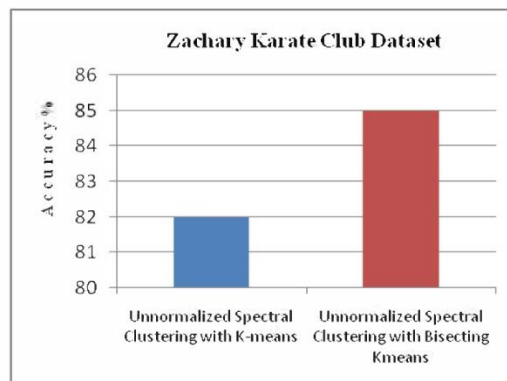


شکل شماره ۳) تقسیم شبکه کلوب Zachary Karate به دو جامعه

شکل ۴ (a) نتایج اعمال رویکرد ارائه شده توسط ما بر شبکه باشگاه کاراته Zachary را نشان می دهد. الگوریتم برای ده بار اجرای مختلف به دست آمده است و شکل ۴ (b) میانگین از این ۱۰ اجرا را نشان می دهد. نتایج با الگوریتم معیار مقایسه می شود.



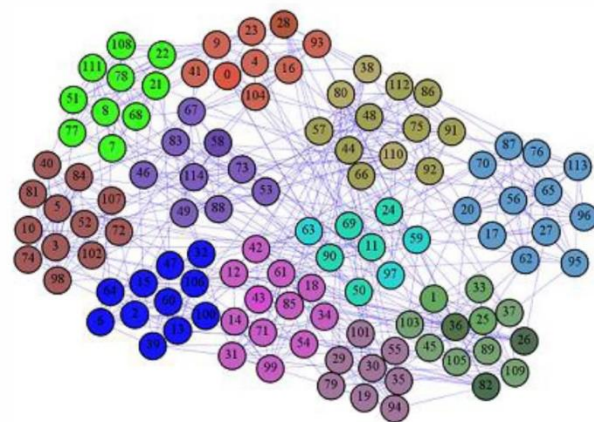
شکل شماره ۴a) نتایج اجرای دو الگوریتم بر روی مجموعه داده های باشگاه Zachary Karate



شکل شماره ۴b) میانگین نتایج ده بار اجرای دو الگوریتم بر روی مجموعه داده های باشگاه Zachary Karate

۳-۴-۲ مدرسه فوتبال آمریکایی

مجموعه داده های شبکه مدرسه فوتبال آمریکایی از داده های بازی های فوتبال مدرسه فوتبال آمریکایی به دست آمده است. جدول بازی های بین مدارس بخش IA در فصل پاییز سال ۲۰۰۰ توسط این شبکه نشان داده شده است. در این شبکه باری ها توسط رئوس و بازی های بین دو تیم در این فصل توسط لبه ها نشان داده شده است. تعداد رئوس در این مجموعه داده ۱۱۵ عدد و تعداد لبه ها ۶۱۶ عدد می باشد. تیم ها به دسته ها تقسیم می شوند. هر تیم در هر گروه به طور متوسط ۴ بازی با تیم های هم گروهش و ۷ بازی با سایر تیم های گروه های دیگر دارد. شکل شماره ۵ ساختار واقعی جوامع این مجموعه داده را قبل از اجرای الگوریتم نشان می دهد.



شکل شماره ۵) مجموعه داده شبکه مدرسه فوتبال آمریکایی با ۱۱ گروه

هنگامی که الگوریتم پیشنهادی خوشه بندی طیفی غیرطبیعی با استفاده از میانگین K دوبعدی در این مجموعه داده استفاده می شود، نتایج خوبی به دست آوردیم. الگوریتم ده بار اجرا شده که در جدول ۳ نتایج به دست آمده نمایش داده شده است. در جدول ۳، نام گروه ها و اندازه آن و تعداد دفعاتی که گروه واقعی با موفقیت توسط الگوریتم شناسایی شده است را را نشان داده ایم. به عنوان مثال، این روش موفق به شناسایی گروه های بزرگ ده در سه اجرا از ۱۰ اجرا نشد. ما مشاهده کردیم که یک تیم به اشتباه به یک گروه دیگر در این سه اجرا انجامید. به طور متوسط، تعداد تیم هایی که جایگزین شده اند نیز در جدول هر زمانی که یک جامعه نامناسب یافت می شود نمایش داده شده است. اطلاعات یکسانی در مورد الگوریتم معیار همچنین در جدول آمده است. حرف "Y" در جدول ۳ نشان می دهد که گروه شناسایی شده، یک گروه واقعی است، در عوض، عدد صحیح نمایش داده شده، تعداد تیم هایی است که توسط الگوریتم به اشتباه به دیگر جوامع اختصاص داده شده است. جدول III همچنین نشان می دهد که در طی ده اجرا، خوشه بندی غیر عادی طیفی کلاسیک با تقسیم میانگین K به اشتباه چهار تیم MidAmerican-WesternAthletic-Conferencing USA و Sunbelt را گروه بندی کرده است. حتی الگوریتم معیار، و همچنین الگوریتم گیروان و نیومن [۱۲] در چنین مواردی شکست خوردند. از آنجایی که در برنامه

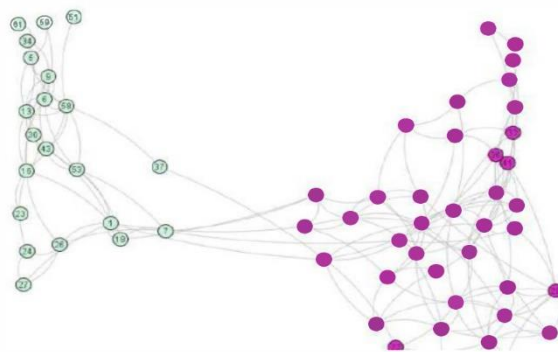
ریزی این بازی ها تفاوت زیادی وجود ندارد، شکست به علت ضعف نگهداری ساختار گروه در این موارد است. نتایج بدست آمده نشان دهنده توانایی خوشه بندی طیفی با روش تقسیم میانگین k برای کمک به تشخیص جامعه در شبکه ها می باشد.

جدول شماره ۳) نتایج مجموعه داده های شبکه مدرسه فوتبال آمریکایی

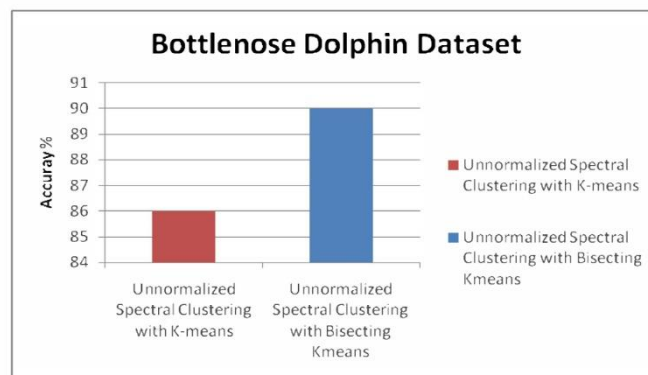
Conference	No.of Teams	Benchmark Algorithm		Proposed Algorithm		
		Number of Correct grouping	Average num of misclassified teams	Number of Correct grouping	Average no. of misclassified teams	
Atlantic Coast	12	10/10	--	10/10	--	Y
Sun Belt	8	0/10	3	3/10	1	3
Big East	8	6/10	3	7/10	1	Y
Pacific Ten	10	9/10	1	9/10	1	Y
Big Ten	11	5/10	1.5	7/10	1	Y
Western Athletic	9	0/10	4	4/10	1.5	2
Big Twelve	12	10/10	--	10/10	--	Y
Southeastern	12	6/10	3	8/10	1	Y
Conference USA	12	2/10	5.5	4/10	2.6	2
Mid-American	12	7/10	3.5	9/10	1.5	5
Mountain West	9	5/10	2	6/10	1	Y

شبکه Dolphin Bottlenose

دیوید Lusseau، یک متخصص زیست شناسی، به مدت هفت سال، رفتار دلفین های bottlenose را که در Sound doubtful (نیوزیلند) زندگی می کنند را تجزیه و تحلیل نموده و این مجموعه داده شبکه را ایجاد کرد. بر اساس ارتباطات مکرر، اگر ارتباط آنها بین دو دلفین برقرار باشد، یک پیوند ایجاد می شود. تعداد کل دلفین هایی که در این مطالعه مورد استفاده قرار گرفتند، ۶۲ عدد و تعداد ۱۵۹ لبه بین این دلفین ها وجود داشت، بین آنهایی که با هم احتمالا بیشتر دیده می شوند. شکل ۶ (a) ساختار اصلی جامعه شبکه دلفین ها را نشان می دهد و شکل ۶ (b) نتایج حاصل از رویکرد ما را نشان می دهد.



شکل ۶ (a)؛ ساختار اجتماعی جامعه دلفین های bottlenose. شبکه به دو گروه تقسیم می شود که در بالا نشان داده شده است. دایره های صورتی یک گروه را نشان می دهند و دایره های سبز گروه دیگر را نشان می دهد.



شکل ۶ (b)؛ نتایج اجرای دو الگوریتم در مجموعه داده های دلفین های Bottlenose به طور میانگین برای ۱۰ بار اجرای الگوریتم

۴ فصل چهارم

نتیجه گیری کارهای آینده:

در این مقاله به معرفی شبکه های اجتماعی پرداختیم و با بررسی انواع شبکه های اجتماعی که در سطح جهان و در جوامع انسانی گوناگون مورد استقبال قرار گرفته اند به یک نگاه کلی در مورد شبکه های اجتماعی رسیدیم، سپس با بیان مسأله و اهداف تحقیق موضوع جوامع و شناسایی جوامع و افراد تأثیرگذار در شبکه های اجتماعی را در نظر گرفته و مورد توجه قرار دادیم.

در ادامه به شناخت انواع روش های شناسایی جوامع و تکنیک هایی که می توان از طریق آن ها جوامع را در شبکه های اجتماعی شناسایی نمود مورد مطالعه قرار دادیم، در این بین روش های مختلفی توسط پژوهشگران در سطح مجامع علمی در دنیا و کشورمان مورد استفاده

قرار گرفته است. این روش ها بر اساس مطالعات و تجربیات به دست آمده در شناسایی جوامع مختلف اعم از جوامع انسانی و غیره در طی سال ها مطالعه و پژوهش به دست آمده است و با بروز و ظهور شبکه های مجازی به عنوان معیاری مورد قبول در شناسایی جوامع مجازی در تحقیقات به کار گرفته شده است.

لازم به توضیح است که این روش های استاندارد در تحقیقات و پژوهش های مختلف گاهی به صورت استاندارد و گاهی نیز به صورت ترکیبی از انواع روش ها و الگوریتم ها به منظور دستیابی به بهترین نتایج به کار گرفته شده است. در اکثر این تحقیقات انجام شده به این نتیجه رسیده بودند که به کارگیری این الگوریتم ها به صورت ترکیبی نتایج مثبت و قابل قبولی را نسبت به روش های استاندارد به دست آورده است، این الگوریتم ها و نتایج به دست آمده به همان نسبت که تعداد جوامع و ترکیب آن ها متفاوت و متنوع می باشد قابل به کارگیری بوده و نتایج متفاوتی مورد انتظار می باشد.

در ادامه این تحقیق به بررسی دو نمونه از تحقیقات انجام شده در خصوص شناسایی جوامع و افراد تأثیرگذار در شبکه های اجتماعی در شبکه اجتماعی توئیتر به عنوان یک شبکه اجتماعی پرطرفدار در جهان پرداختیم، در این دو مقاله به بررسی روش های شناسایی جوامع در توئیتر توسط الگوریتم های شباهت شامل شباهت های Friends، Followers، هشتگ، پاسخدهی و شباهت بین کاربران و همچنین استفاده از الگوریتم های خوشه بندی برای شناسایی جوامع پرداختیم و نتایج تجربی به دست آمده از این دو نمونه عملی انجام شده را مورد توجه قرار دادیم.

در ادامه قصد داریم با معرفی بیشتر و دقیق تر الگوریتم های شناسایی جوامع و افراد تأثیرگذار در سطح شبکه های اجتماعی به انتخاب دو الگوریتم متناسب با ساختار جامعه پرداخته و ضمن مقایسه عملکرد آن دو الگوریتم اقدام به ترکیب توانایی های آن ها نموده و نتایج به دست آمده از روش جدید را با نتایج قبلی مقایسه و به عنوان یک نتیجه گیری کلی بیان نماییم.

فهرست منابع :

- al., A. e. (۲۰۱۵). OLFinder: Finding opinion leaders in online social networks. Journal of Information Science, ۱۶. doi:۰۱۶۵۵۵۱۵۱۵۶۰۵۲۱۷/۱۰,۱۱۷۷
- al., E. V. e. (۲۰۱۵). Mining Interesting Topics in Twitter Communities. Springer International Publishing Switzerland ۲۰۱۵. doi:۱۲۵-۲۴۰۶۹-۳۱۹-۳-۹۷۸/۱۰,۱۰۰۷
- al., L. L. e. (۲۰۱۱). Leaders in Social Networks, the Delicious Case. plos one, ۶
- al., M. e. (۲۰۰۸). Identifying User Behavior in Online Social Networks. {mmaia, jussara, virgilio}@dcc.ufmg.br .
- al., M. Z. S. e. (۲۰۱۳). Identifying Leaders and Followers in Online Social Networks. IEEE International .
- al., S. A. M. e. (۲۰۱۶). Community detection in social networks using user frequent pattern mining. Springer-Knowledge and Information Systems. doi:۱۰,۱۰۰۷/5۸-۰۹۷۰-۰۱۶-۱۰۱۱۵
- E. Raju, e. a. (۲۰۱۵). Detecting Communities in Social Networks using Unnormalized Spectral Clustering incorporated with Bisecting K-means. IEEE International. doi:-۱-۹۷۸ ۱۵/۹-۶۰۸۵-۴۷۹۹/۳۱\$. . .
- Gonz´alez-Cantergiani, F. a. R. a. P. (۲۰۱۶). Measuring user influence on Twitter: A survey. {fabian.riquelme.c, pablo.gonzalezca}@usach.cl .
- Kusters, F. W. T. W. A. (۲۰۱۱). IdentifyingProminentActorsinOnline SocialNetworksusingBiasedRandomWalks. www.facebook.com/press/info.php?statistics .
- Rossi, A. C. a. S. (۲۰۱۴). Users Ranking in Online Social Networks to Support POI Selection in Small Groups. Proceedings of UMAP ۲۰۱۴posters .
- Yanmei Hu, B. Y. (۲۰۱۵). Enhanced link clustering with observations on ground truth to discover social circles. elsevier Knowledge-Based Systems, ۹. doi:۷۰۵۱-۰۹۵۰
- محمودرضا بابایی. (۱۳۹۰). تحلیل ساختار ارتباط بر روند انتشار اطلاعات در شبکه‌های اجتماعی.
- همکاران, ا. م. ح. آ. و. (۱۳۹۴). شناسایی گره‌های قدرت در شبکه‌های اجتماعی به کمک داده کاوی (۱۹JRM), ۲(۲۱).

- همکاران, م. آ. د. و. (۱۳۹۱). شبکه های اجتماعی و روش های استخراج جوامع در آن ها. همایش ملی علوم و مهندسی کامپیوتر-دانشگاه آزاد اسلامی اصفهان واحد نجف آباد.
- شریف زاده, فائزه؛ سمیه کافی و مرتضی براری، ۱۳۹۳، ارائه روشی جدید برای شناسایی گره های فعال و تاثیرگذار در شبکه های اجتماعی، همایش ملی مهندسی رایانه و مدیریت فناوری اطلاعات، تهران، شرکت علم و صنعت طلوع فرزین.