

تشخیص اجتماعات در شبکه‌های اجتماعی

محدثه قاسم‌پور¹، عباس سیفی²، حسین علیزاده³

¹دانشکده مهندسی صنایع دانشگاه صنعتی امیرکبیر، m.ghasempour@aut.ac.ir

²دانشکده مهندسی صنایع دانشگاه صنعتی امیرکبیر، a.seifi@aut.ac.ir

³دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران، halizadeh@iust.ac.ir

شماره مقاله: 1788

چکیده

در عصر حاضر، به علت در دسترس بودن اینترنت، امکان استفاده، تولید و نشر اطلاعات در وب افزایش یافته است. همچنین با گسترش روزافزون استفاده از ارتباطات در دنیای مجازی و رشد تعداد کاربران شبکه‌های اجتماعی، مطالعه و تحلیل این شبکه‌ها ضروری به نظر می‌رسد. یکی از زمینه‌های مورد توجه تحلیل‌گران شبکه‌های اجتماعی، تشخیص اجتماعات در این شبکه‌ها است. هدف تشخیص اجتماعات، کشف زیرساختارهایی است که ممکن است در شبکه‌ها وجود داشته باشند. تشخیص این گروه‌ها در شبکه‌های اجتماعی کاربردهای فراوانی در بازاریابی، علوم اجتماعی، اقتصادی و... دارد. لذا با توجه به اهمیت و نو بودن موضوع، در این مقاله به مساله تشخیص اجتماعات در شبکه‌های اجتماعی به عنوان یکی از جدیدترین حوزه‌های علم کامپیوتر می‌پردازیم و ضمن معرفی مفاهیم بنیادی شبکه‌های اجتماعی و کاربردهای فراوان مساله تشخیص اجتماعات، روش‌های مختلف تشخیص اجتماعات در شبکه‌ها را بررسی می‌کنیم. در این مقاله ضمن تشریح هر سه دسته، به تشریح آخرین مطالعات انجام گرفته در این زمینه نیز می‌پردازیم.

کلمات کلیدی

شبکه‌های اجتماعی، تشخیص اجتماعات، روش‌های تقسیم‌کننده، روش‌های تجمعی، بیشینه‌سازی پیمانه‌بندی.

Community Detection in Complex Networks

Mohadeseh Ghasempour, Abbas Seifi, Hosein Alizadeh

ABSTRACT

Nowadays, producing and distributing information on the Net are enhanced due to the widespread access to the Internet. In addition, by exploring virtual communication and increase in social networks' users, analyzing such networks seems inevitable. One hot topic in this area is community detection. It aims to discover sub-structures in the network that have applications in marketing, social science, economics, etc. Due to the importance of this topic, we discuss community detection problem in this paper. In addition to clarifying essential concepts and applications, we discuss different methods of community detection in complex networks. In this paper, we touch on latest researches in this area besides discussing methods of community detection in complex networks.

KEYWORDS

Social networks, Community detection, Divisive methods, Agglomerative methods, Modularity maximization

¹محدثه قاسم‌پور، دانشکده مهندسی صنایع دانشگاه صنعتی امیرکبیر، تلفن: 09119093441، پست الکترونیکی: m.ghasempour@aut.ac.ir

1- مقدمه

در دنیای پیشرفته امروز با رشد تکنولوژی و افزایش قدرت جمع‌آوری داده‌ها، بشر با حجم عظیمی از داده‌ها سر و کار دارد. همچنین با توجه به گسترش روزافزون استفاده از اینترنت، خدمات وب و شبکه‌های اجتماعی مجازی نقش مهمی در زندگی واقعی افراد پیدا کرده است. به طوری که مدیریت داده‌های حاصل از میلیون‌ها تراکنش ارتباطی بین بازیگران این شبکه‌ها و کشف دانش نهفته در ساختار شبکه‌های اجتماعی برای بهره‌برداری هر چه موثرتر و کم‌هزینه‌تر از آن در اهداف بازاریابی و ... را بدون مطالعه و بررسی دقیق علمی و استفاده از ابزارهای پیشرفته غیر ممکن ساخته است. به همین خاطر در سال‌های اخیر تلاش‌های علمی قابل توجهی برای توسعه ابزارهای ریاضی و کامپیوتری برای تحلیل شبکه‌های اجتماعی صورت گرفته است. در حوزه شبکه‌های اجتماعی، تشخیص و تمایز گروه‌هایی از کاربران که با هم ارتباطات تنگاتنگی دارند از اهمیت ویژه‌ای در مباحث تجاری، بازاریابی، اهداف سیاسی و اجتماعی برخوردار است. به عنوان مثال گروه‌هایی از افراد که علائق مشترکی از خود نشان داده‌اند و یا گروه‌های دوستی و یا تشخیص سایت‌ها و وبلاگ‌های حامی مواضع سیاسی مشترک و غیره [1]. به دلیل اهمیت موضوع به نظر می‌رسد مطالعه و بررسی در زمینه روش‌های تشخیص اجتماعات در شبکه‌های اجتماعی و تلاش برای بهبود روش‌های موجود ضروری باشد.

2- شبکه‌های اجتماعی

1-2 شبکه‌های پیچیده

شبکه‌های پیچیده یک نمایش انتزاعی از ساختار سیستم‌های پیچیده هستند. یک سیستم پیچیده، سیستمی است که از ترکیب اجزای بسیار زیادی تشکیل شده است که با هم در تعامل هستند به گونه‌ای که رفتار کلی سیستم لزوماً از رفتار تک‌تک اجزا پیروی نمی‌کند. مثال‌هایی از سیستم‌های پیچیده شامل اکوسیستم‌های طبیعت، بورس و سیستم‌های اقتصادی، سیستم‌های تکاملی بیولوژیکی، شبکه‌های کامپیوتری، و البته سیستم‌های اجتماعی در جوامع انسانی می‌باشند [3].

2-2 شبکه‌های اجتماعی

نخستین بار مبحثی با عنوان شبکه‌های اجتماعی در سال 1960 در دانشگاه ایلی نیوز در ایالت متحده آمریکا مطرح شد که بعد از آن در 1997 نخستین سایت شبکه اجتماعی به آدرس اینترنتی SixDegrees.com راه‌اندازی شد. یک شبکه اجتماعی ساختاری اجتماعی است، متشکل از گره‌هایی که به صورت منفرد یا گروهی (سازمانی) مرتبط شده‌اند، روابطی همچون مبادلات مالی، دوستی، بازرگانی، پیوندهای وب و سرگرمی دارند [4 و 5]. چنین شبکه‌هایی مردم را با انواع علائق مختلف مرتبط می‌کند. استفاده از این شبکه‌ها فقط محدود به گروه‌های دوستی نبوده و امروزه شرکت‌ها به منظور کمک به کارمندان و یا کمک به مشتریان و دست‌یافتن به اطلاعات، به استفاده از شبکه‌های اجتماعی روی آورده‌اند.

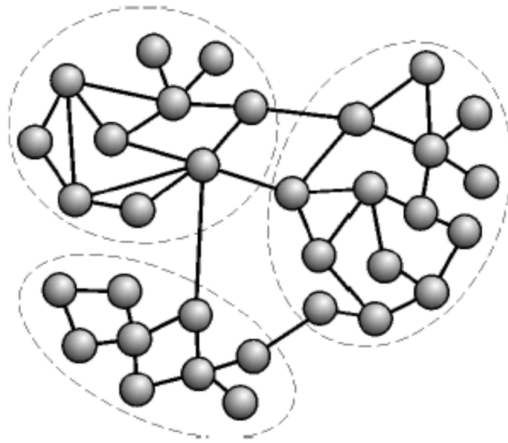
3-2 تعریف اجتماع در شبکه‌های اجتماعی

در شبکه‌های اجتماعی برخی گره‌ها (افراد یا سازمان‌ها) در مقایسه با کل گره‌های شبکه، ارتباط بیشتری با هم دارند که به آن‌ها اجتماع (گروه)² گفته می‌شود [2]. یکی از چالش‌های مهم در تحلیل شبکه‌های اجتماعی، تشخیص اجتماعات می‌باشد. به عنوان مثال در شبکه‌های اجتماعی، افرادی را می‌بینیم که به دفعات مختلف با افراد و گروه‌های دیگری ارتباط برقرار می‌کنند. برخی افراد اجتماعی‌تر هستند و با گروه‌های بیشتری مرتبطند برخلاف برخی دیگر که محتاط و منزوی عمل می‌کنند و مایل به برقراری ارتباطات جدید نیستند. مثالی از نمایش گرافی یک شبکه و عمل تشخیص اجتماع روی آن در شکل (1) نمایش داده شده است.

در مراجع مختلف تعاریف متفاوتی از اجتماع ارائه شده است که برخی از آنها در ادامه ذکر می‌شوند:

² Community

- یک اجتماع وب مجموعه‌ای از صفحات وب است که توسط افراد یا سازمان‌های مختلف که علایق مشترک درباره یک موضوع خاص دارند، ایجاد شده‌اند. مانند صفحات علاقه‌مندان یک تیم بیس‌بال و یا صفحات تولید کنندگان تجهیزات کامپیوتری [6].
- در تعریف دیگری از اجتماعات آمده است در یک اجتماع چگالی یال‌ها داخل آن بیشتر از چگالی یال‌های بین آن با سایر گره‌های شبکه است [7].
- یک اجتماع مجموعه‌ای از گره‌های مرتبط به یکدیگر است که پیوندهای بین آنها متراکم است [5].



شکل (1): نمونه‌ای از یک شبکه که اجتماعات با خط‌چین از هم جدا شده‌اند.

به طور دقیق‌تر در تشخیص اجتماعات با داشتن یک شبکه $G(V,E)$ شامل گره‌ها و یال‌ها (که در یک شبکه اجتماعی می‌توانند معرف کاربران و ارتباطاتشان باشند) می‌خواهیم ارتباطات معنادار را پیدا کنیم. در واقع به دنبال یک افراز از شبکه هستیم که پیوندهای درون آن‌ها متراکم باشد.

4-2 کاربردهای تشخیص اجتماعات

- تشخیص اجتماعات، تقسیم‌بندی‌های موجود در شبکه را نشان می‌دهد و گروه‌های آن را از هم مجزا می‌کند. تشخیص اجتماعات به ما کمک می‌کند تا دید بهتری نسبت به ساختار شبکه پیدا کنیم. به طور کلی برخی از کاربردهای تشخیص اجتماعات عبارتند از [8]:
- **درک ساختار شبکه:** در شبکه‌های پیچیده‌ای مانند وب و یا شبکه‌های اجتماعی که با حجم بالایی از گره‌ها روبرو هستیم، تشخیص اجتماعات اولین و مهم‌ترین قدم در راستای شناخت و تحلیل این شبکه‌ها محسوب می‌شود.
 - **یافتن اجتماعاتی با ویژگی خاص:** گاهی در تشخیص اجتماعات به دنبال اجتماعاتی هستیم که ویژگی خاصی دارند. مثلاً اجتماع افرادی که در یک تیم خاص عضو هستند، یا به طور کلی گروه افرادی که به محصول خاصی علاقه‌مندند هستند. یافتن اجتماعات خاص می‌تواند اهداف متفاوتی داشته باشد، که از آن جمله می‌توان به اهداف تجاری، بازاریابی، اهداف سیاسی و اجتماعی اشاره نمود.
 - **پیمایش موضوعی:** پیمایشگرهای موضوعی کار خود را با یک مجموعه اولیه از صفحات آغاز می‌کنند و با استفاده از پیوندهای موجود در این صفحات، صفحات دیگر را پیمایش کرده و این روند را تا رسیدن تعداد صفحات پیمایش شده به حدی مشخص انجام می‌دهند. پیمایشگرهای موضوعی در روند پیمایش خود به صورت انتخاب‌گر عمل می‌کنند و صفحاتی را برای پیمایش انتخاب می‌کنند که تا حد ممکن در ارتباط با موضوعی خاص باشند.
 - **قابل مشاهده کردن گراف:** با تشخیص اجتماعات می‌توانیم نمای کامل گراف یا بخشی از آن را مشاهده کنیم. با مشاهده اجتماعات موجود در یک گراف می‌توان اجزای آن را به طور مجزا از کل گراف مورد تحلیل قرار داد.
 - **بهبود موتورهای جستجو:** موتورهای جستجو برای بازبانی اسناد مرتبط با پرس‌وجوی کاربران، اسناد را رتبه‌بندی می‌نمایند. روش‌های رتبه‌بندی میزان ارتباط پرس‌وجوی کاربر را با هر صفحه از یک مجموعه صفحات محاسبه می‌کنند و به هر صفحه رتبه‌بندی اختصاص می‌-

دهند. برای این منظور پرس‌وجو و اسناد به صورتی که قابل مقایسه با یکدیگر باشند، بازنمایی می‌شوند. از تشخیص اجتماعات می‌توان در خوشه‌بندی موضوعی وب سایت‌ها نیز استفاده کرد. این روش به بهبود عملکرد موتورهای جستجو کمک شایان توجهی می‌کند. با توجه به کاربردهای فراوان تشخیص اجتماعات در شبکه‌های اجتماعی، دست یافتن به روش‌هایی برای اجرایی کردن آن از اهمیت ویژه‌ای برخوردار است. در این راستا بررسی و مطالعه جدیدترین روش‌های موجود در این زمینه ضروری به نظر می‌رسد. لذا در ادامه به دسته‌بندی، معرفی و بررسی جدیدترین روش‌های تشخیص اجتماعات در شبکه‌های اجتماعی می‌پردازیم.

3- روش‌های تشخیص اجتماعات در شبکه‌های اجتماعی

به‌طور کلی روش‌هایی که در ادبیات موضوع برای یافتن اجتماعات در شبکه‌های پیچیده استفاده می‌شوند به سه دسته کلی روش‌های تقسیم‌کننده، روش‌های تجمعی و بیشینه‌سازی پیمانه‌بندی تقسیم می‌شوند.

3-1 روش‌های تقسیم‌کننده^۳

این روش از جمله روش‌هایی است که توپولوژی شبکه را تغییر می‌دهد. در واقع در این روش به دنبال یال‌هایی هستیم که اجتماعات مختلف را به هم وصل می‌کند و این یال‌ها را در یک فرم تکرار شونده حذف می‌کنیم تا شبکه به اجتماعات مجزایی از گره‌ها تقسیم شود. معروف‌ترین روش تقسیم‌کننده روش گیروان-نیومن^۴ [9] است. چون اجتماعات مختلف توسط تعداد کمی یال به هم متصل شده، این روش گلوگاه‌هایی که در یال‌ها قرار دارند و دو اجتماع را به هم متصل می‌کند مورد توجه قرار می‌دهد. از طریق این یال‌ها همه مسیرهای کوتاه عبور می‌کنند. این روش از مرکزیت بینابینی^۵ استفاده می‌کند. روش کار بدین صورت است که مرکزیت بینابینی برای تک‌تک یال‌ها محاسبه می‌شود و یال‌هایی که بیشترین مقدار را دارند به عنوان یال‌های بین اجتماعات شناسایی شده و حذف می‌شوند و این کار تا تشخیص کل اجتماعات ادامه می‌یابد. اگر چه این روش قدرتمندی در تشخیص اجتماع است ولی معایبی دارد که مهم‌ترین آن‌ها هزینه بالای محاسباتی است. برای بهبود این روش، تایلر در^۶ [10] روشی ارائه کرده که در آن یک عنصر آشوبی (بی‌نظمی) به این روش اضافه شده است که محاسبه B (مقدار مرکزیت بینابینی یال‌ها) را به مجموعه جزئی از یال‌ها محدود می‌کند و از محاسبات آماری برای تخمین B واقعی استفاده می‌کند. پس از این روش، الگوریتم‌های دیگری ارائه شده که از دقت و سرعت بیشتری برخوردارند. در رویکرد تقسیم‌کننده در بسیاری از روش‌های ارائه شده، تعداد قابل توجهی از یال‌ها مورد پردازش قرار می‌گیرند. تعداد یال‌ها در شبکه‌ها زیاد است و محاسباتی که روی آنها انجام می‌شود، به طبع آن بالاست. بنابراین روش‌های تقسیم‌کننده از پیچیدگی زمانی و محاسباتی زیاد رنج می‌برند.

3-2 روش‌های تجمعی^۷

این روش‌ها بر این حقیقت بنا شده‌اند که گره‌های یک اجتماع ویژگی‌های مشترکی دارند و می‌توان از این ویژگی‌های مشترک برای گروه‌بندی استفاده کرد. در برابر روش‌های تقسیم‌کننده، روش تجمعی در ابتدا همه گره‌ها را جدا از هم و غیر متصل در نظر می‌گیرد و آن‌ها را بر اساس ویژگی‌های مشترک به هم متصل می‌کند تا به اجتماعات برسد. در واقع روش‌های تجمعی بر خلاف روش‌های تقسیم‌کننده رویکرد پایین به بالا یا رسیدن از جزء به کل دارند. نحوه عملکرد این رویکرد بدین صورت است که یال‌های بین اجتماعات خود به خود حذف شده و تنها یال‌های داخل اجتماع باقی می‌مانند.

یک گروه مهم از روش‌های تجمعی، خوشه‌بندی سلسله‌مراتبی^۸ است، که با N گره غیر متصل و بدون یال شروع می‌شود. یال‌ها در طول روش بر اساس کاهش شباهت به شبکه اضافه می‌شوند. در ابتدا یال‌هایی که بیشترین شباهت را دارند، اضافه می‌شوند. سپس به مرور سایر گره‌ها مشابه به اجتماع اضافه می‌شوند تا به تدریج اجتماع را شکل دهند. در این روش‌ها تعداد اجتماعات می‌تواند توسط تحلیل گر تعیین شود. یکی از روش‌های اندازه‌گیری شباهت، فاصله اقلیدسی است که برای گراف‌های وزن دار قابل استفاده است. در [11] که توسط نیومن ارائه شده روشی تجمعی برای تشخیص اجتماع به کار گرفته شده است. در این مقاله از اصول روش گیروان-نیومن

³Divisive

⁴ Girvan-newman(GN)

⁵ Betweenness Centrality

⁶ Tyler

⁷ Agglomerative

⁸ Hierarchical clustering

استفاده شده، با این تفاوت که زمان اجرای آن کاهش پیدا کرده است. این روش با معیار پیمان‌بندی ارزیابی می‌شود. روش کار بدین صورت است که در ابتدا هر گره یک اجتماع را تشکیل می‌دهد. اجتماعات به شرطی که بیشترین افزایش در پیمان‌بندی یا Q (یا کم‌ترین کاهش در Q) را داشته باشند به هم می‌پیوندند این فرایند را می‌توان با یک دندروگرام نشان داد. تغییرات در Q بر اثر اتصال دو اجتماع به صورت زیر است:

$$\Delta Q = e_{ij} + e_{ji} - 2a_i a_j = 2(e_{ij} - a_i a_j) \quad (1)$$

این روش بجای محاسبه مکرر و زمان‌بر Q از ΔQ استفاده می‌کند که به طور محلی تغییرات Q را محاسبه می‌کند و باعث کاهش زمان روش می‌شود. روش به کار رفته قابل تبدیل به گراف وزن‌دار است و از سری روش‌های حریصانه محسوب می‌شود. مقاله [12] روش تجمعی سلسله مراتبی دیگری ارائه کرده است. این روش به نوعی تکمیل‌کننده این روش می‌باشد. در [11] از ماتریس مجاورتی برای ذخیره کردن گراف استفاده شده در حالی که [12] به جای ذخیره ماتریس مجاورتی، ماتریس مقادیر ΔQ ذخیره می‌شود. اجتماعاتی که ارتباطی بینشان نیست تغییری در ΔQ آن‌ها ایجاد نمی‌شود. در نتیجه ذخیره نمی‌شوند و این باعث ایجاد یک ماتریس اسپارس می‌شود.

3-3 پیشینه سازی پیمان‌بندی⁹

نیومن در [13] معیار پیمان‌بندی را برای کیفیت ارزیابی اجتماعات شبکه معرفی کرد که به‌زودی به یکی از معیارهای پرکاربرد برای ارزیابی کیفیت اجتماعات مبدل گشت. با داشتن یک افراز از شبکه، معیار پیمان‌بندی یال‌های درون اجتماعات را با گراف تصادفی فرضی با توزیع درجه یکسان با شبکه مورد نظر مقایسه و از این طریق کیفیت خوشه‌ها را ارزیابی می‌کند.

رابطه معیار پیمان‌بندی به صورت زیر تعریف می‌شود:

$$Q(C) := \frac{1}{2m} \sum_{u,v} (a_{u,v} - \frac{d_u d_v}{2m}) \cdot \delta(\gamma(u), \gamma(v)) \quad (2)$$

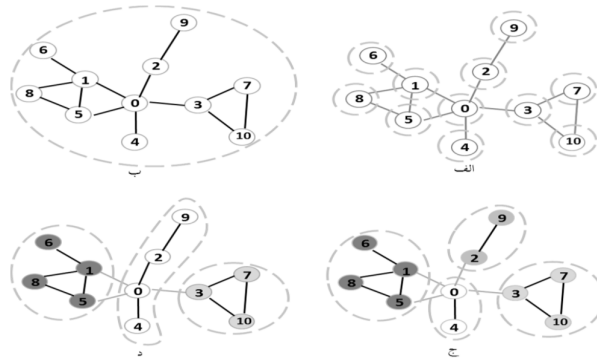
در این رابطه δ تابع کرونیگر¹⁰ است که اگر آرگمان‌های آن با هم مساوی باشند مقدار 1 را می‌پذیرد و در غیر این صورت مقدارش صفر است.

درایه‌های ماتریس M که ماتریس پیمان‌بندی است $m_{u,v} := a_{u,v} - \frac{d_u d_v}{2m}$ هستند. اگر کل گراف شامل تنها یک اجتماع باشد یا گره‌ها

به صورت تصادفی در بین اجتماعات قرار گرفته باشند $Q = 0$ خواهد بود. هر چه مقدار Q به 1 نزدیک‌تر باشد اجتماعات بهتر جدا شده‌اند ولی هیچ‌گاه مقدار آن یک نمی‌شود. در عمل مقدار $Q > 0$ ساختار گروهی مناسبی را نشان می‌دهد. این معیار می‌تواند مقادیر منفی را هم بپذیرد.

مقدار پودمانی ممکن است برای یک شبکه در بهترین حالت گروه‌بندی از عدد بیشتر نشود. این بدان معناست که لزوماً نمی‌توان گفت اگر پودمانی مقدار 0.4 دارد، تشخیص اجتماع به درستی انجام نشده است. ممکن است 0.4 بیشترین مقداری باشد که پودمانی در آن شبکه دارد. در شکل (2) پودمانی را برای یک گراف آزمایشی در اجتماعات مختلف نشان داده‌ایم. این گراف دارای 11 گره و 13 لبه است. در شکل (2-الف) هر گره در یک اجتماع قرار داده شده است. این نوع گروه‌بندی دارای خطای زیادی است و در نتیجه پودمانی این گروه‌بندی منفی و برابر $Q = -0.115$ است. در شکل (2-ب) کل گره‌های گراف در یک اجتماع قرار داده شده است. رابطه پودمانی به گونه‌ای است که در این حالت انگار گروه‌بندی انجام نشده است و پودمانی گراف در این حالت برابر صفر است. دو شکل (2-ج) و (2-د) دو گروه‌بندی دیگر از گراف آزمایشی را نشان می‌دهد. بهترین گروه‌بندی مربوط به شکل (2-د) است که بیشترین پودمانی را بدست می‌دهد. پودمانی بیشینه برای این گراف برابر با $Q = 0.451$ است. همان‌طور که مشاهده می‌کنید، بیشترین مقدار پودمانی برای این گراف به 0.5 نمی‌رسد. این بدان معناست که لزوماً بهترین گروه‌بندی برای یک گراف پودمانی نزدیک به 1 ندارد.

⁹Modularity maximization
¹⁰ Kronecker Delta



شکل (2): نمایی از یک گراف آزمایشی در گروه‌بندی‌های مختلف

جستجو برای یافتن پیمان‌بندی بهینه (بیشینه) از نوع مسائل بسیار سخت به نظر می‌آید. چراکه فضای تقسیم‌بندی‌هایی که از گراف امکان‌پذیر است با بزرگ شدن اندازه گراف به سرعت در حال افزایش است. روش‌هایی که هدف آن بیشینه‌سازی پیمان‌بندی است، معمولاً با رویکردهای تقسیم‌کننده، تجمعی و یا طیفی ترکیب می‌شود و یا از روش‌های مکاشفه‌ای استفاده می‌کند تا به پیمان‌بندی بیشینه دست یابد. اخیراً برنوس و همکاران در [14] نشان داده‌اند که برای یک گراف مشخص، یافتن خوشه‌هایی که دارای بیشترین مقدار پیمان‌بندی هستند یک مساله NP-کامل¹¹ است بدین مفهوم که یافتن الگوریتمی کارا که بتواند همیشه جواب بهینه این مساله را در زمان چندجمله‌ای بیابد بعید به نظر می‌رسد.

روش‌هایی که در ادبیات موضوع در زمینه یافتن خوشه‌هایی با حداکثر مقدار معیار پیمان‌بندی وجود دارد به طور کلی به چند دسته تقسیم می‌شوند که در ادامه کارهای انجام شده در هر مورد را بررسی خواهیم کرد.

3-1-3 بهینه‌سازی اکسترمم

این روش که توسط داچ و ارنانز¹² [15] پیشنهاد شده است، جستجوی مبتنی بر تجربه برای بهینه‌سازی مقدار پیمان‌بندی است. به طور پیش فرض، شبکه به دو بخش تصادفی تقسیم می‌شود که تعداد گره‌ها در آن‌ها مساویند. در هر مرحله، سیستم با حرکت دادن گره‌هایی با کم‌ترین برآزش، از یک بخش به بخش دیگر خودسازماندهی می‌کند. این فرآیند هنگامی که پیمان‌بندی بیشینه شود متوقف می‌شود. اکنون شبکه دارای دو اجتماع است. حال یال‌ها بین اجتماعات را حذف می‌کنیم. در اینجا یک اجتماع را برداشته و روند پیشین را برای آن تکرار می‌کنیم تا به دو اجتماع جدید برسیم. فرآیند بالا را آن قدر تکرار می‌کنیم تا به پیمان‌بندی بیشینه برسیم. اگرچه این روش سریع نیست، ولی می‌تواند مقادیر پیمان‌بندی بالایی بدست آورد.

3-2-3 روش‌های مبتنی بر الگوریتم‌های تکاملی

روش‌های مختلفی در حوزه الگوریتم‌های تکاملی بر روی تشخیص اجتماعات انجام شده است. تارگین و بینگل¹³ [16] روشی ارائه کردند که از پیمان‌بندی برای سنجش استفاده می‌کند. در این روش هر کروموزوم کل گره‌های موجود در شبکه را شامل می‌شود. در ابتدا به هر گره به صورت تصادفی یک اجتماع نسبت داده می‌شود و سپس CROSSOVER که تغییراتی روی آن صورت گرفته، انجام می‌شود. گاهی نیز جهش¹⁴ که گره‌های دو اجتماع را جابجا می‌کند اعمال می‌شود و در نهایت هر کروموزوم توسط معیار پیمان‌بندی سنجیده می‌شود. سپس پیمان‌بندی مقادیر همه اجتماعات را محاسبه می‌کند که با به‌کارگیری ساختار سلسله مراتبی بدست آمده‌اند و به عنوان نتیجه، اجتماعی که بیشترین مقدار پیمان‌بندی را دارد، بر می‌گرداند.

مشکل این روش آن است که در گراف‌های بزرگ نمی‌توان از آن استفاده کرد زیرا نیاز به حافظه بسیار زیادی دارد که بتواند در هر کروموزوم

¹¹ NP-Complete
¹² Duch and Arenas
¹³ Targin and Bingol
¹⁴ mutation

آن، کل گره‌های شبکه را جا دهد. پیزوتی¹⁵ [17] در روش خود که آن را GA-Net نامیده است، از ماتریس مجاورتی استفاده می‌کند. در این روش کل گراف با ماتریس نشان داده می‌شود و اجتماعات زیرماتریس‌هایی از ماتریس اصلی هستند که در ابتدا به صورت تصادفی انتخاب می‌شوند. برای سنجیدن درستی کار خود پیزوتی معیار جدیدی را ارائه کرده است که در آن به اجتماعات با چگالی بیشتر امتیاز بالاتری داده شود. روش دیگری که از الگوریتم ژنتیک استفاده کرده، روش ACGA¹⁶ است که توسط لیپزاک و میلیوس¹⁷ [18] ارائه شده است. این روش به حافظه زیادی نیاز ندارد زیرا هر کروموزوم آن تنها بخشی از پاسخ مسئله است و نه تمام آن و کل کروموزوم‌ها با هم، تشکیل جواب یا نتیجه را می‌دهند. این روش را می‌توان به راحتی برای گراف‌های بزرگ نیز استفاده کرد. از سایر الگوریتم‌های ابتکاری که برای بهینه‌سازی معیار پیمان‌بندی به‌کار گرفته شده‌اند می‌توان به الگوریتم‌های حریمانه¹⁸ [12] و استفاده از بهینه‌سازی شبیه‌سازی ذوب فلزات [19] اشاره کرد.

3-3 روش‌های طیفی¹⁹

این روش‌ها بر مبنای تحلیل ماتریس‌های بردار ویژه‌ای است که از شبکه بدست آمده است. در این روش بر روی مقادیر ویژه ماتریس‌های مرتبط با ماتریس مجاورت کار می‌شود، که می‌تواند ماتریس لاپلاسیان باشد. نیومن در [20] مساله یافتن ساختار اجتماعات در شبکه‌ها را با استفاده از بردارهای ویژه ماتریس بهینه می‌کند. وی تابع پیمان‌بندی را به فرم ماتریسی استفاده می‌کند تا مساله را به شکل یک مساله طیفی²⁰ در جبر خطی بررسی کند. ادعا می‌کند بیشینه‌سازی معیار پیمان‌بندی با استفاده از تجزیه برداری این ماتریس می‌تواند همانند نقش ماتریس لاپلاس در خوشه‌بندی گراف عمل کند. وی از این طریق الگوریتمی برای تقسیم شبکه به دو اجتماع ارائه می‌کند. وقتی در روش طیفی دو بخشی دو اجتماع داریم (C=2)، تقسیم‌بندی شبکه با نسبت دادن مقادیر مثبت بردار ویژه به یک اجتماع و مقادیر منفی به اجتماع دیگر بدست می‌آید. در [21] روشی ارائه شده است که مفهوم پیمان‌بندی را به شکل بردار ویژه فرمول‌بندی مجدد می‌کند و آن را ماتریس پیمان‌بندی می‌نامد. برای هر زیر گراف g ماتریس پیمان‌بندی $B^{(g)}$ درایه‌های زیر را دارد:

$$b_{ij}^{(g)} = a_{ij} - \frac{k_i k_j}{2M} - \delta_{ij} \sum_{u \in N(g)} \left[a_{iu} - \frac{k_i k_u}{2m} \right] \quad (3)$$

که برای گره‌های i و j در زیر گراف g تعریف شده است. δ_{ij} در صورت وجود یال بین گره‌های i و j یک و در غیر این صورت صفر است. رابطه (12)، M تعداد کل یال‌های گراف و $N(g)$ مجموعه گره‌های موجود در زیر گراف g است. k_i, k_j, k_u درجات گره‌های i, j و u است. بنابراین برای جداسازی شبکه به اجتماعات تشکیل‌دهنده آن ابتدا باید ماتریس پیمان‌بندی ساخته شود و مثبت‌ترین مقدار ویژه و بردار ویژه متناظر با آن تعیین شود. بر اساس علامت درایه‌های این بردار، شبکه به دو قسمت تقسیم می‌شود (گره‌های با مقادیر مثبت در یک اجتماع و مقادیر منفی در اجتماع دیگر). این فرایند به طور بازگشتی برای هر اجتماع تکرار می‌شود. در [22] از خوشه‌بندی طیفی ماتریسی²¹ برای خوشه‌بندی شبکه استفاده شده است. آزادسازی طیفی یک روش آزادسازی معروف است که در خوشه‌بندی طیفی استفاده می‌شود که در آن خوشه‌بندی بر اساس تجزیه بردار ویژه ماتریس لاپلاس نرمال شده عمل می‌کند. همچنین در این مقاله از برنامه‌ریزی ماتریسی²² برای خوشه‌بندی گراف به خوشه‌های هم‌اندازه استفاده شده است که شرط لازم برای برقراری شرایط دوگانی قوی برقرار است. به طور کلی روش کار خوشه‌بندی طیفی ماتریسی بر مبنای تجزیه بردارهای ویژه ماتریس شدنی بهینه محاسبه شده توسط برنامه‌ریزی ماتریسی استوار است.

¹⁵ Pizzuti

¹⁶ Agglomerative Clustering Genetic Algorithm

¹⁷ Lipszak and Millios

¹⁸ Greedy Algorithms

¹⁹ Spectral

²⁰ Spectral Problem

²¹ Semidefinite Spectral Clustering

²² Semidefinite programming

4- نتیجه

در این مقاله شبکه‌های اجتماعی، به عنوان یکی از پرکاربردترین منابع اطلاعاتی معرفی شد و ضرورت شناخت، مطالعه و تحلیل این شبکه‌ها مورد بررسی قرار گرفت. همچنین مساله تشخیص اجتماعات در شبکه‌های اجتماعی به عنوان یکی از ابزارهای کشف دانش از این شبکه‌ها بررسی شده و کاربردهای فراوان آن برشمرده شد. علاوه بر آن روش‌های موجود برای تشخیص اجتماعات در شبکه‌ها در سه دسته کلی دسته‌بندی شده و مورد بررسی قرار گرفت. روش‌های تقسیم‌کننده که با حذف یال‌های رابط اجتماعات را نمایان می‌سازد، روش‌های تجمعی که طی یک ساختار سلسله‌مراتبی و با توجه به شباهت گره‌ها اجتماعات را شناسایی می‌کند و در آخر روش‌های مبتنی بر بیشینه‌سازی پیمان‌بندی که با مقایسه شبکه با یک شبکه تصادفی دارای توزیع درجه مشابه، سعی در یافتن ارتباطات معنادار و تشخیص اجتماعات در شبکه دارد. با توجه به اهمیت موضوع تشخیص اجتماعات در شبکه‌های پیچیده (به‌خصوص شبکه‌های اجتماعی) به نظر می‌رسد مطالعه عمیق روش‌های موجود تشخیص اجتماعات و سعی در بهبود آن‌ها از کارهای مهم پیش رو باشد.

مراجع

- [1] محدثه قاسم‌پور، تشخیص اجتماعات در شبکه‌های پیچیده، سمینار کارشناسی ارشد، دانشکده مهندسی صنایع، دانشگاه صنعتی امیرکبیر، 1391.
- [2] رسول حسین‌زاده، حسین علیزاده و اسلام ناظمی، تشخیص اجتماعات با رویکرد ترکیبی در شبکه‌های اجتماعی، یازدهمین کنفرانس سیستم‌های هوشمند ایران، 2013.
- [3] M. E. J. Newman, "Complex systems: A survey", American Journal of Physics, Vol. 79, pp. 800-810, 2011.
- [4] Ahn, Y.-Y., Bagrow, J. P., and Lehmann, S., Link communities reveal multiscale complexity in networks. Nature 466, 761-764 (2010).
- [5] M. E. J. Newman, "Communities, modules and large-scale structure in networks", Nature Physics, Vol. 8, 2012.
- [6] B. Liu, "Web Data Mining Exploring Hyperlinks, Contents, and Usage Data", Springer, 2007.
- [7] M. A. Porter, J. P. Onnela and P. J. Mucha, "Communities in networks", Notices of the American Mathematical Society, Vol. 56, No. 9, pp. 1082-1097, 2009.
- [8] S. Fortunato, "Community detection in graphs", Physics Reports, Vol. 486, No. 3-5, pp. 75-174, 2010.
- [9] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks", National Academy of Sciences of the United States of America, Vol. 99, No. 12, p. 7821, 2002.
- [10] J. R. Tyler, D. M. Wilkinson and B. A. Huberman, "E-mail as spectroscopy: Automated discovery of community structure within organizations", The Information Society, Vol. 21, No. 2, pp. 143-153, 2005.
- [11] M. E. J. Newman, "Fast algorithm for detecting community structure in very large networks", Physical review E, Vol. 69, 2004.
- [12] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks", Physical Review E, Vol. 70, No. 6, 2004.
- [13] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks", Physical review E, Vol. 69, No. 2, 2004.
- [14] U. Brandes, D. Delling, M. Gaertler, R. Görke, M. Hoefer, Z. Nikoloski, and D. Wagner. "On modularity clustering", IEEE Transactions on Knowledge and Data Engineering, Vol. 20, No. 2, pp. 172-188, 2008.
- [15] J. Duch and A. Arenas, "Community detection in complex networks using extremal optimization", Physical Review E, Vol. 72, No. 2, 2005.
- [16] M. Tasgin and H. Bingol, "Community detection in complex networks using genetic algorithm", Arxiv preprint cond-mat/0604419, 2006.
- [17] C. Pizzuti, "Ga-net: A genetic algorithm for community detection in social networks" Parallel Problem Solving from Nature-PPSN X, pp. 1081-1090, 2008.
- [18] M. Lipczak and E. Milios, "Agglomerative genetic algorithm for clustering in social networks", 11th Annual conference on Genetic and evolutionary computation, pp. 1243-1250, 2009.
- [19] J. Reichardt and S. Bornholdt, Statistical Mechanism of Community Detection, Phys. Rev. E, Vol. 74, 2006.
- [20] M. Newman, "Finding community structure in networks using the eigenvectors of matrices", Physical Review E, 74, 2006.
- [21] M. E. J. Newman, "Modularity and community structure in networks", National Academy of Sciences, Vol. 103, No. 23, 2006.
- [22] J. Kim, S. Choi, "Semidefinite spectral clustering", Pattern Recognition 39, pp. 2025 - 2035, 2006.



دومین کنفرانس ملی مهندسی صنایع و سیستمها
6 و 7 اسفند 1392 - دانشگاه آزاد اسلامی واحد نجف آباد
گروه مهندسی صنایع

