

کشف اجتماعات در شبکه‌های اجتماعی یک لایه‌ی پویا با رویکرد بیشینه سازی اهمیت

علی سرآبادانی^۱، سید مرتضی پورنقی^۲، خیراله رهسپار^۳

۱- دانشجوی مقطع دکتری تخصصی، رشته مهندسی فناوری اطلاعات، دانشگاه قم

۲- استادیار گروه رایانش امن، دانشکده رایانه، شبکه و ارتباطات، دانشگاه جامع امام حسین (ع)

۳- عضو هیئت علمی گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه قم

(دریافت: ۱۴۰۲/۱۲/۰۱، بازنگری: ۱۴۰۳/۰۲/۱۳، پذیرش: ۱۴۰۲/۰۲/۱۷، انتشار: ۱۴۰۳/۰۳/۱۳)

4- DOR: <https://dorl.net/dor/>

-۵-

چکیده

دنیای کنونی، عصر شبکه و اینترنت است عصری که در آن با شکل‌گیری شبکه‌های گوناگون اجتماعی، شیوه‌های ارتباطی و اطلاع‌رسانی نوینی به عرصه‌ی گسترده ارتباطات اجتماعی معرفی شده‌اند. شبکه‌های اجتماعی از محبوب‌ترین مقاصد سال‌های اخیر کاربران اینترنت شده است [۱]. شبکه‌ی اجتماعی گروهی از افراد یا سازمان‌های دارای منافع و سلايق مشترک هستند که برای دستیابی به اهداف مشخصی گرد هم می‌آیند. دلایل عمده ایجاد شبکه‌های اجتماعی شامل روابط فردی، روابط کاری، روابط علمی، سلیقه و علایق و تفریحات مشترک و انگیزه‌های اجتماعی سیاسی می‌باشد. آنالیز یا تحلیل شبکه‌های اجتماعی به معنایی مطالعه‌ی ویژگی‌های شبکه‌ی اجتماعی و روابط بین افراد و بخش‌های یک شبکه با رویکرد تئوری شبکه‌ای یا گراف است. تحلیل شبکه‌های اجتماعی نوعی مطالعه میان رشته‌ای در حوزه‌های مختلف است از جمله جامعه‌شناسی، ریاضیات، علوم رایانه و سایبری می‌باشد. یکی از چالش‌های اساسی در تحلیل شبکه‌های اجتماعی ناپایدار بودن این شبکه‌ها است زیرا در هر ثانیه ممکن است افرادی یا سازمانی‌هایی به عضویت این شبکه‌ها درآیند و یا از آن خارج شده و یا رابطه‌های جدیدی شکل گیرد.

ما در این مقاله یک روشی ابتکاری برای کشف اجتماعات پیچیده در شبکه‌های اجتماعی با یک لایه‌ی پویا و با تمرکز بر بیشینه‌سازی معیار اهمیت ارائه می‌دهیم. سپس با استفاده از مجموعه داده‌ی تولید شده توسط LFR Benchmark به بررسی نتایج طرح پیشنهادی خود با سایر طرح‌ها پرداخته و نتایج بدست آمده حاکی از بهبود عملکرد طرح پیشنهادی ما در دقت شناسایی روابط پیچیده در شبکه‌های اجتماعی و زمان کمتر آن در مقایسه با سایر طرح‌ها است.

کلمات کلیدی

کشف اجتماعات، شبکه‌های اجتماعی پویا، شبکه یک لایه، اهمیت، اجتماع، بیشینه سازی

۱- مقدمه

به هر ساختار اجتماعی از افراد که بر اساس یک رابطه‌ی اجتماعی ایجاد می‌شود، یک شبکه‌ی اجتماعی گفته می‌شود. بنابراین هر شبکه‌ی اجتماعی شامل مجموعه‌ای از انسان‌ها و روابط اجتماعی بین آن‌هاست. لذا هر شبکه اجتماعی از دو عنصر موجودیت‌های شرکت کننده در ارتباط و ارتباط بین این موجودیت‌ها تشکیل شده است [۲].

شبکه‌های اجتماعی از قرن نوزدهم مورد توجه دانشمندان حوزه جامع‌شناسی و علوم رایانه قرار گرفتند. پژوهش‌ها در این حوزه از دهه‌ی چهل به بعد با تعریف ابزارهایی چون گراف اجتماعی [۱] شتاب بیشتری گرفت. در سال ۱۹۹۴ میلادی واسرمن^۱ با چاپ کتاب تحلیل شبکه‌های اجتماعی [۲] این زمینه از علم را وارد دوره‌ی جدیدی کرد، و پس از آن شبکه‌های اجتماعی به صورت جدی در زیرمجموعه‌های علوم اجتماعی و ریاضی مورد بررسی قرار گرفت.

رشد شبکه‌های اجتماعی برخط، باعث شده که راهکارهایی جهت استخراج جوامع بوجد بیاید. آنچه که در تشخیص جوامع مدنظر می‌باشد و از اهداف استخراج آن به شمار می‌رود تعاملاتی است که افراد یک جامعه با هم دارند، به طوری که بیشترین ارتباط درون جامعه و روابط خارج از جامعه آن‌ها کمترین مقدار ارتباط را به خود اختصاص می‌دهد. پس با استفاده از تفکیک شبکه به جوامع گوناگون با توجه به شباهت‌های هر گروه به مطالعه آن‌ها پرداخته می‌شود. هرچند استخراج و مطالعه‌ی جوامع در شبکه‌های اجتماعی موضوع جدیدی نیست ولی محدودیت‌ها، روش‌های تحقیق بیشتری را در این زمینه می‌طلبد.

مشکل مهمی که در روش‌های موجود وجود دارد عدم توجه به این نکته است که در یک شبکه‌ی اجتماعی ممکن است چندین نوع رابطه بین افراد وجود داشته باشد. برای مثال در یک شبکه‌ی

اجتماعی علاوه بر رابطه‌ی دوستی بین افراد روابطی مانند فامیل بودن نیز وجود دارد. این مساله در سال‌های اخیر در مبحث کشف اجتماعات در شبکه‌های اجتماعی، مورد توجه پژوهشگران قرار گرفته است.

یکی دیگر از مشکلات مطرح در الگوریتم‌های کشف اجتماعات، پویا بودن شبکه‌هاست که در دنیای واقعی امری متداول است. در واقع نیاز به الگوریتم‌های کارایی برای به روز رسانی ساختار اجتماعات بعد از ایجاد تغییرات در روابط المان‌های شبکه ضروری به نظر می‌رسد. ما در این مقاله موفق به ارائه یک روش پویا برای شناسایی روابط پیچیده در شبکه‌های اجتماعی شده‌ایم.

۲- شبکه اجتماعی

در این بخش به بررسی برخی از شبکه‌های پیچیده در دنیای واقعی می‌پردازیم. شبکه‌های رایانه‌ای شامل رایانه‌ها هستند که اتصالاتی مانند کابل یا ماهواره به هم متصل شده‌اند. انواع مختلفی از این شبکه‌ها مانند شبکه‌ی تلفن و شبکه‌ی حمل و نقل وجود دارند. در بسیاری از موارد اتصال بین گره‌ها فیزیکی نیست، برای مثال شبکه‌ی WWW یک شبکه‌ی جهت‌دار از صفحات وب است که اگر یک صفحه به صفحه‌ی دیگر پیوند داشته باشد به هم متصل خواهند بود. مثال دیگر، شبکه‌ی پست الکترونیک (رایانامه) است که در آن هر فرد با یک گره نمایش داده می‌شود و یک یال جهت‌دار زمانی وجود دارد که یک فرد پست الکترونیک فرد دیگری را در لیست آدرس خود داشته باشد. نوع دیگری از شبکه‌ها، گراف تماس بین افراد است که معمولاً توسط اپراتورهای شبکه‌ی تلفن ایجاد می‌شود. در این گراف، هر فرد با یک گره نمایش داده می‌شود و یک یال جهت‌دار زمانی وجود دارد که در یک بازه‌ی معین یک فرد با فرد دیگر تماس برقرار کرده باشد.

نمونه‌ی دیگر از شبکه‌های پیچیده، شبکه‌های اجتماعی هستند که تمرکز ما در این مقاله بر روی این شبکه‌ها است. این شبکه‌ها

¹ Wasserman

ماتریس‌های شبکه‌های اجتماعی لزوماً متقارن نیستند. یعنی ممکن است رابطه‌ی بین افراد یک طرفه یا دو طرفه باشد.

۲-۱-۲- گراف

گراف‌ها یکی از ابزارهای نمایش بصری و تحلیل شبکه‌های اجتماعی هستند. در گراف هر گره به عنوان یک فرد در نظر گرفته می‌شود و وجود یال بین دو گره بیانگر وجود رابطه بین دو فرد متناظر با این دو گره است. گراف مورد استفاده برای یک شبکه بر اساس نوع شبکه می‌تواند گراف جهت‌دار یا بدون جهت باشد.

۳- ویژگی‌های ساختاری شبکه‌های اجتماعی

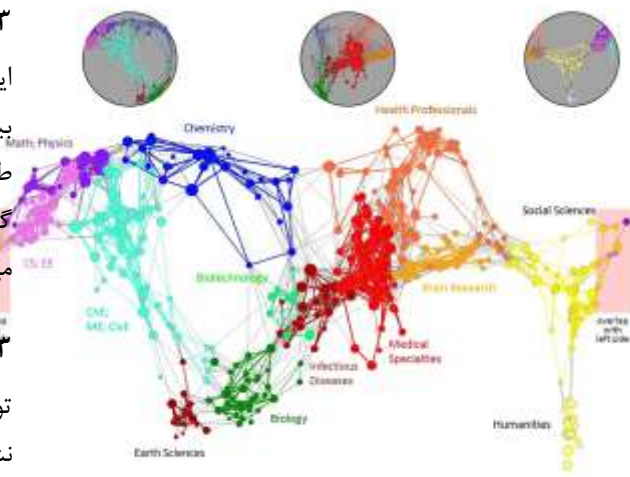
۳-۱- قطر

این ویژگی روی گراف شبکه‌های اجتماعی تعریف می‌شود. مسیر بین دو گره از گراف دنباله‌ای از یال‌های بین این دو گره است و طول مسیر به معنای تعداد یال‌های بین آن دو گره می‌باشد. قطر گراف برابر با طولانی‌ترین مسیر موجود بین دو گره از گراف می‌باشد.

۳-۲- توزیع درجه

توزیع درجه در یک گراف یک تابع $p(k)$ می‌باشد که نشان‌دهنده‌ی نسبت تعداد گره‌های با درجه‌ی k نسبت به تعداد کل گره‌های گراف است. توزیع درجه یکی از ویژگی‌های مهم برای تحلیل گراف می‌باشد. توزیع درجه رئوس در گراف شبکه‌های اجتماعی متفاوت با گراف‌های تصادفی است. در بسیاری از شبکه‌های واقعی از جمله شبکه‌های اجتماعی، توزیع درجه‌ی رئوس گراف متناظر شبکه از توزیع قاعده‌ی توانی^۳ پیروی می‌کند. در صورتی که در گراف‌های تصادفی که وجود یال میان هر دو زوج راس دلخواه دارای احتمال یکسان است، توزیع درجه رئوس از توزیع دوجمله‌ای پیروی می‌کند. در شکل ۲-۲ توزیع درجه‌ی یک گراف با توزیع قاعده‌ی توانی نمایش داده شده است که بیشتر رئوس آن دارای درجه‌ی پایین و تعداد محدودی از رئوس دارای درجه‌ی بالا هستند. به شبکه‌هایی که در آن‌ها از قاعده‌ی توزیع توانی پیروی می‌شود شبکه‌های مقیاس آزاد^۴ گفته می‌شود.

بیانگر ارتباط بین افراد مانند رابطه‌ی دوستی یا رابطه‌ی کاری هستند. برای مثال یکی از این شبکه‌ها شبکه‌ی بازیگران است و دو بازیگر زمانی با هم مرتبط هستند که حداقل در یک فیلم مشترک بازی کرده باشند. یک مثال دیگر از شبکه‌های اجتماعی، شبکه‌ی همکاری پژوهشگران است. در این شبکه هر پژوهشگر با یک گره نمایش داده می‌شود و دو پژوهشگر زمانی به هم متصل هستند که حداقل یک مقاله‌ی مشترک با هم منتشر کرده باشند. یکی دیگر از شبکه‌های اجتماعی، گراف رجوع^۲ است که در آن هر پژوهشگر با یک گره نمایش داده می‌شود و زمانی بین پژوهشگران ارتباط وجود دارد که یک پژوهشگر در مقاله‌ی خود به مقاله‌ی فرد دیگر ارجاع داده باشد. شکل ۱-۲ نمونه‌ای از این نوع شبکه را نشان می‌دهد.



شکل ۱-۲ شبکه‌ی ارجاع مقاله‌ها [۳]

۲-۱-۱- نمایش شبکه‌های اجتماعی

برای مطالعه و تحلیل یک شبکه‌ی اجتماعی نیاز به ساختارهایی قابل فهم است که افراد شبکه و روابط آن‌ها را نمایش دهد. معمولاً دو شیوه برای نمایش شبکه‌های اجتماعی مورد استفاده قرار می‌گیرد: گراف‌ها و ماتریس‌ها.

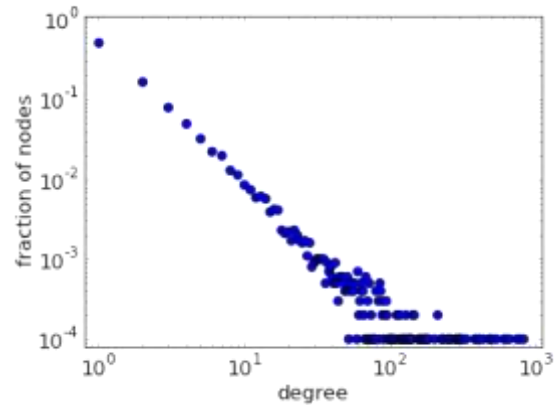
۲-۱-۱-۲- ماتریس

می‌توان شبکه‌های اجتماعی را به صورت آرایش مربعی از اعداد نشان داد. در ماتریس هر سطر و ستون متعلق به یک فرد در شبکه است. در ماتریس مربوط به یک شبکه، a_{ij} برابر یک است اگر ارتباط موجود باشد، در غیر این صورت این مقدار صفر است.

³ power law
⁴ scale free

² citation network

(سه‌گانه‌ی بسته) یال به هم وصل می‌شوند. یک مثلث از سه‌گانه‌ی بسته تشکیل شده است. ضریب خوشه‌بندی کلی برابر است با نسبت سه‌گانه‌های بسته، یعنی ۳ ضربدر تعداد مثلث‌ها بخش بر تعداد کلی سه‌گانه‌ها که هم شامل سه‌گانه‌های بسته و هم سه‌گانه‌های باز می‌باشند. در واقع ضریب خوشه‌بندی نشان دهنده‌ی میزان اتصال همسایه‌های یک گره و در نتیجه احتمال تشکیل خوشه می‌باشد. این ضریب عددی حقیقی بین صفر و یک است، زمانی که هیچ خوشه‌بندی وجود نداشته باشد این مقدار برابر با صفر و در صورتی که گراف، کامل باشد این مقدار برابر با یک می‌شود [۶].



شکل ۳-۱ توزیع درجه‌ی یک گراف با قاعده‌ی توانی [۴]

۳-۳- اثر دنیای کوچک^۵

تراورس و میلگرام [۴] در سال ۱۹۶۹ آزمایشی طراحی کردند تا میانگین طول مسیر شبکه‌ی اجتماعی انسان در آمریکا را بررسی کنند. در این آزمایش افرادی انتخاب و از آن‌ها خواسته شده بود که نامه‌ای را به فردی در شهر موردنظر با آغاز از یک شهر ارسال کنند. در پایان ۶۴ نامه دریافت شد و میانگین مسیر پیموده شده بین افراد مختلف حدود ۶ بود. در واقع میانگین کمترین فاصله‌ی رئوس در شبکه بسیار پایین بود. این ویژگی که اثر دنیای کوچک نام دارد به این معنی است که بیشتر زوج رئوس حاضر در شبکه‌های واقعی فاصله‌ی بسیار کمی با یکدیگر دارند و این میزان با لگاریتم تعداد رئوس موجود در شبکه رابطه دارد.

۳-۴- ضریب خوشه‌بندی

در نظریه‌ی گراف ضریب خوشه‌بندی اندازه‌ای است که بیانگر درجه‌ای است که گره‌ها در یک گراف گرایش به تشکیل خوشه در کنار یکدیگر دارند. شواهد بیانگر این هستند که در بیشتر شبکه‌های موجود در دنیای واقعی و در شبکه‌های اجتماعی خاص گره‌ها تمایل دارند تا به صورت قوی گروه‌هایی را تشکیل دهند که از طریق تراکم بالای یال‌ها مشخص و معلوم می‌شوند. در شبکه‌های واقعی این همسایگی از میانگین احتمال این که یک یال به صورت تصادفی بین دو گره اتصال برقرار نماید بیشتر است. دو نوع از این اندازه‌ها وجود دارند یکی اندازه‌ی کلی و دیگری اندازه‌ی محلی است. ضریب کلی خوشه‌بندی بر مبنای گره‌های سه‌تایی محاسبه می‌شود. یک حالت سه‌گانه از سه گره تشکیل شده است که از طریق دو (سه‌گانه‌ی باز) و یا سه

۳-۵- همبندی

در گراف بدون جهت، دو راس را متصل می‌نامیم اگر حداقل یک مسیر بین آنها وجود داشته باشد. اگر دو راس با مسیری به طول یک یال به هم متصل شده باشند، آنگاه این دو راس همسایه نامیده می‌شوند. یک گراف همبند نامیده می‌شود اگر هر دو راس دلخواه آن به یکدیگر متصل باشند. یک مولفه‌ی همبند یک زیرگراف همبند بیشینه از گراف است. تعداد مولفه‌های همبند نشان‌دهنده‌ی قابلیت اتصال گراف است. هرچه تعداد مولفه‌های همبند گراف کمتر باشد کل گراف دارای اتصالات قوی‌تری است.

۳-۶- مرکزیت^۶

مرکزیت یکی از معیارهای مهم ساختاری شبکه می‌باشد. مرکزیت معمولاً برای گره‌های گراف تعریف می‌شود و بیانگر میزان اهمیت یک گره از جنبه‌های مختلف می‌باشد. سه جنبه‌ی مهم از مرکزیت، درجه، نزدیکی^۷ و بینابینی^۸ می‌باشد.

- **درجه** : درجه تعداد رئوسی می‌باشد که یک راس به آن متصل است. اگر گراف آشنایان را در نظر بگیریم، مرکزیت درجه برابر با تعداد افرادی می‌باشد که یک فرد خاص می‌شناسد.

- **نزدیکی** : مجموع فاصله‌ی یک گره از گراف تا سایر رئوس شبکه را نزدیکی می‌نامند. زمانی که یک راس امتیاز نزدیکی کمی داشته باشد در واقع مرکزیت بالایی دارد و در جریان اکثر اطلاعات شبکه قرار

⁶ centrality

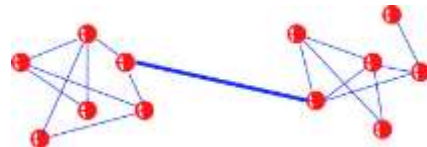
⁷ closeness

⁸ betweenness

⁵ small world

می‌گیرد و از آنجایی که با بیشتر رئوس فاصله‌ی کمی دارند، اطلاعات را سریعتر دریافت می‌کنند.

- **بینابینی** : بینابینی، تعداد کوتاه‌ترین مسیر بین هر جفت از رئوس گراف است که از یال مورد نظر عبور می‌کند. این معیار اهمیت یک یال را در فرآیندی مثل انتشار اطلاعات نشان می‌دهد. زیرا در اغلب اوقات اطلاعات از نزدیکترین مسیر انتقال می‌یابند. به طور شهودی قابل درک است که بینابینی یال‌های بین دو اجتماع بیش از سایر یال‌هاست. زیرا کوتاه‌ترین مسیرهای بین دو اجتماع از این یال‌ها عبور خواهند کرد. البته این فرض زمانی درست است که تعداد گره‌ها به نسبت برابر بین دو جامعه تقسیم شده باشد. شکل ۲-۳ بیانگر این موضوع است.



شکل ۲-۳ زیاد بودن میزان بینابینی یال بین دو جامعه در یک شبکه [۷]

۴- اجتماع و تعاریف آن

انواع مختلفی از شبکه‌ها در جهان وجود دارد، اما در این نوشته، تمرکز ما روی شبکه‌های اجتماعی است که در واقع بیانگر روابط بین افراد می‌باشد. برای مثال یکی از شبکه‌های اجتماعی، شبکه‌ی همکاری^۹ و مقالات مشترک بین تمام محققان است. در این شبکه هر پژوهشگر با یک گره نمایش داده می‌شود و در صورتی که دو پژوهشگر حداقل یک مقاله‌ی مشترک داشته باشند، بین آنها یک یال وجود دارد. برخی گره‌های شبکه‌های واقعی با هم ارتباط بیشتری دارند و جوامع را تشکیل می‌دهند به طوری که داخل جوامع یال‌های به مراتب بیشتری نسبت به یال‌های بین جوامع وجود دارد. هدف شاخه‌ی کشف اجتماعات^{۱۰} یافتن جوامع موجود در یک شبکه و احتمالاً ساختار سلسله مراتبی بین آنهاست.

کشف اجتماعات کاربردهای مهمی دارد. برای مثال یافتن جوامع در شبکه‌ی همکاری پژوهشگران می‌تواند به یافتن گروهی از پژوهشگرانی که در حوزه‌های مشترک کار می‌کنند کمک کند. به عنوان مثالی دیگر، در شبکه‌ی اجتماعی فیس بوک هر کاربر با افراد زیادی رابطه‌ی دوستی دارد و دوستی با افراد مختلف باعث

اشتراک فرد در جوامع مختلف مانند جامعه‌ی خانواده، دانشگاه، محل کار و ... می‌شود. اکنون اگر برای مثال فیس بوک بخواهد به یک کاربر لیستی از افراد مشابه برای دوستی را پیشنهاد دهد، کشف اجتماعات می‌تواند به انتخاب افراد مشابه کمک کند.

یکی از سخت‌ترین کارهای مربوط به کشف اجتماعات در همان گام نخست روی می‌دهد. اولین مساله در کشف اجتماعات شبکه‌های پیچیده، یافتن تعریفی کمی از جامعه است. شکل ۲-۴ یک گراف را همراه با ساختار اجتماعات آن نشان می‌دهد. هرچند به لحاظ بصری یافتن اجتماعات این گراف راحت به نظر می‌رسد، اما به لحاظ علم ریاضی و نظریه گراف ارائه‌ی تعریف و یافتن این اجتماعات پیچیده می‌نماید. در گراف شکل ۲-۴ توده‌های مجتمع راس‌ها، انجمن‌ها را تشکیل می‌دهند. شکل ۲-۵ یک گراف تصادفی با تعداد راس‌ها و یال‌های مساوی با گراف شکل ۲-۴ است که برای هر یال دو راس را به طور تصادفی انتخاب کرده و به هم وصل می‌کند. همانطور که مشخص است به لحاظ شهودی و تصویری گراف‌های مربوط به شبکه‌ی اجتماعی از گروه‌های متراکم زیادی تشکیل شده است اما گراف مربوط به گراف تصادفی تنها از یک توده تشکیل شده است. به صورت شهودی توزیع یال‌ها در گراف‌های واقعی همانند گراف فیس‌بوک نه به صورت سراسری که به صورت محلی است. بدین معنی که تعداد یال‌های توزیع شده میان گروهی از راس‌ها بسیار بیشتر از تعداد یال‌های توزیع شده بین این گروه از راس‌ها با بقیه راس‌های گراف است. این ویژگی که در گراف‌های مربوط به داده‌های واقعی می‌توان مشاهده کرد، جامعه خوانده می‌شود. در برخی منابع به جامعه انجمن یا خوشه نیز گفته می‌شود.

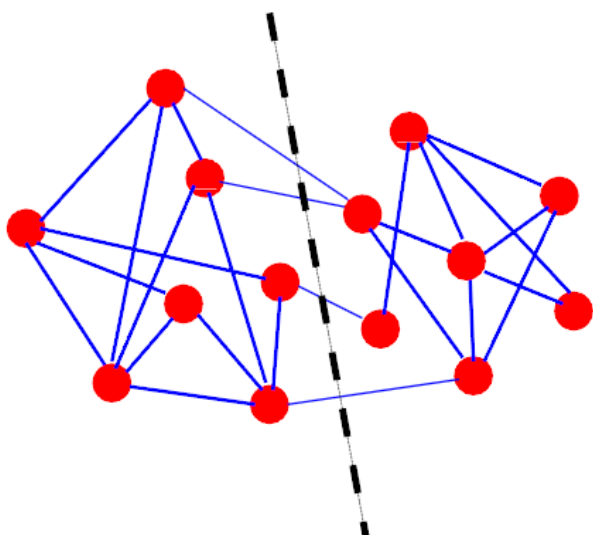
از کاربردهای کشف اجتماعات می‌توان به تبلیغات و بازاریابی اشاره کرد. از آنجایی که افراد حاضر در اجتماعات تشکیل شده در یک شبکه‌ی اجتماعی به احتمال زیاد علائق مشترکی دارند، می‌توان با یافتن علائق آنها از این اطلاعات به منظور تبلیغ محصولات خاص استفاده کرد. کاربردهای زیاد دیگری نیز می‌توان برای اجتماعات نام برد. کاربرد فراوان در زمینه‌هایی همچون زیست‌شناسی، مهندسی کامپیوتر، اقتصاد این شاخه از علم شبکه‌های اجتماعی و تئوری گراف را به زمینه‌ای محبوب برای پژوهشگران جهت تحقیق تبدیل کرده است.

⁹ collaboration network

¹⁰ community detection

گروه‌ها با عنوان اندازه‌ی برشی^{۱۱} شناخته می‌شود. شکل ۱-۶ حل این مساله برای یک گراف با ۱۴ راس و ۲ گروه با اندازه‌های مساوی را نشان می‌دهد.

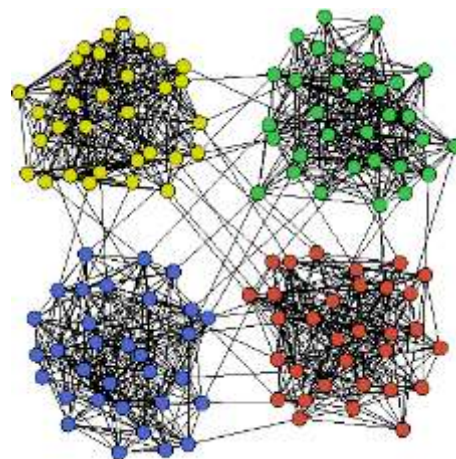
تعیین تعداد خوشه‌های موردنظر برای تقسیم‌بندی ضروری است. زیرا اگر تعداد خوشه‌ها مهم نباشد و هدف کمینه کردن مقدار اندازه‌ی برشی باشد جواب بدیهی است، تمام یال‌ها باید در یک خوشه قرار بگیرند. مشخص بودن اندازه‌ی خوشه‌ها نیز ضروری است زیرا در غیر این صورت راه حل این است که راس با کمترین درجه را از گراف جدا کنیم که نتیجه مطلوب نخواهد بود. این مشکل با بهینه‌سازی یک معیار دیگر برای خوشه‌بندی که اندازه‌ی خوشه را کنترل کند قابل حل است. تعیین تعداد خوشه‌های موردنظر برای تقسیم‌بندی ضروری است.



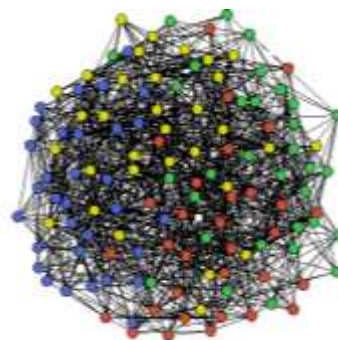
شکل ۱-۵ تقسیم‌بندی گراف به دو گروه هم‌اندازه

زیرا اگر تعداد خوشه‌ها مهم نباشد و هدف کمینه کردن مقدار اندازه‌ی برشی باشد جواب بدیهی است، تمام یال‌ها باید در یک خوشه قرار بگیرند. مشخص بودن اندازه‌ی خوشه‌ها نیز ضروری است زیرا در غیر این صورت راه حل این است که راس با کمترین درجه را از گراف جدا کنیم که نتیجه مطلوب نخواهد بود. این مشکل با بهینه‌سازی یک معیار دیگر برای خوشه‌بندی که اندازه‌ی خوشه را کنترل کند قابل حل است.

مساله‌ی افزاز کردن گراف یک مساله‌ی NP-Hard است. اگرچه، روش‌های زیادی برای حل این مساله وجود دارد هرچند ممکن است بهینه نباشند. بسیاری از روش‌های موجود گراف را به دو



شکل ۱-۴ یک گراف با خوشه‌های معین [۸]



شکل ۲-۴ یک گراف با ساختار تصادفی [۸]

تعاریف متعددی برای جامعه ارائه شده است. هیچ تعریفی به صورت کلی پذیرفته شده نیست. در حقیقت، این تعاریف معمولاً به سیستم خاص در دست بررسی و یا کاربردی که در ذهن برای آن خوشه‌بندی در نظر گرفته شده بستگی دارد. اساس بیشتر تعاریف موجود برای جوامع این نکته است که تعداد یال‌ها در داخل جامعه نسبت به یال‌هایی که رئوس آن جامعه را به بقیه‌ی گراف متصل می‌کنند بیشتر است.

۵- روش‌های کشف اجتماعات

در سال‌های گذشته روش‌های متعددی برای کشف اجتماعات در شبکه‌ی های اجتماعی ارائه شده است. در این بخش دسته بندی جامعی از این روش‌ها ارائه خواهد شد.

۵-۱- الگوریتم‌های افزاز کننده

مساله‌ی افزاز کردن گراف به چند بخش برابر است با تقسیم گره‌ها به گروه‌های با اندازه‌ی از پیش تعریف شده به طوری که تعداد یال‌های بین گروه‌ها حداقل باشد. تعداد یال‌های بین

^{۱۱} cut size

بخش تقسیم می‌کنند و برای ایجاد چند خوشه باید به صورت بازگشتی الگوریتم را تکرار کرد.

در این روش‌ها هر گره مختصات معینی دارد و شباهت گره‌ها بر اساس فاصله بین گره‌ها تعریف می‌شود. هدف این است که گره‌ها طوری به چند خوشه تقسیم شوند که یک معیار هدف، بر اساس فاصله‌ی بین گره‌ها کمینه شود. برخی از معیارهایی که برای این کار استفاده می‌شوند عبارتند از :

- **Minimum k-clustering**: تابع هزینه در اینجا قطر

خوشه است، که به معنی بیشترین فاصله بین دو گره در یک خوشه است. گره‌ها به شکلی خوشه‌بندی می‌شوند که بزرگترین قطر خوشه‌ها کمترین مقدار ممکن باشد. ایده‌ی این روش این است که خوشه‌ها تا حد ممکن فشرده باشند.

- **k-clustering sum**: این روش مشابه روش

minimum k-clustering است با این تفاوت که به جای قطر از میانگین فاصله‌های بین گره‌ها استفاده می‌شود.

- **k-center**: برای هر خوشه i یک نقطه‌ی مرجع x_i با

نام مرکز^{۱۲} در نظر گرفته می‌شود و d_i ، بیشترین مقدار فاصله‌ی هر نقطه‌ی خوشه تا مرکز محاسبه می‌شود. خوشه‌ها و مرکزها تا زمانی که به کمترین مقدار d_i برسیم تغییر می‌کنند.

- **k-median**: مشابه k-center است با این تفاوت که

به جای بیشترین فاصله از مرکز، از فاصله‌ی میانگین تا مرکز استفاده می‌شود.

یکی از محبوبترین روش‌های خوشه‌بندی روش k-means [۹] است. در اینجا تابع هزینه، مجموع فاصله‌های گره‌های داخل هر خوشه است :

$$\sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - c_i\|^2 \quad (۱-۵)$$

که S_i بیانگر گره‌های داخل خوشه‌ی i ام و c_i مرکز آن خوشه است. مساله‌ی k-means می‌تواند توسط الگوریتم Lloyd حل شود. در این الگوریتم کار با چند مرکز که در فاصله‌ی مناسبی از هم قرار دارند شروع می‌شود. در تکرار اول هر گره به

نزدیکترین مرکز تخصیص داده می‌شود. سپس برای هر خوشه مرکز جدید تخمین زده می‌شود و مجموعه مراکز جدید بدست می‌آید و به همین ترتیب مراحل دوباره تکرار می‌شود. بعد از چند تکرار، مراکز تقریباً ثابت می‌مانند و خوشه‌ها بدست می‌آیند. جواب بدست آمده بهینه نیست و خوشه‌های حاصل به مراکز انتخاب شده‌ی اولیه بستگی دارند اما مزیت این روش همگرایی سریع آن است. برای دقیق‌تر کردن جواب می‌توان الگوریتم را چند بار و با مراکز مختلف تکرار کرد و بهترین جواب را انتخاب کرد.

روش‌های دیگری نیز برای استفاده از k-means ارائه شده‌اند که مشکل همه‌ی این روش‌ها این است که باید تعداد خوشه‌ها را ابتدا مشخص باشد.

۵-۲- الگوریتم‌های سلسله مراتبی

معمولاً اطلاعات کمی در مورد ساختار جوامع گراف در دسترس است و تعداد جوامعی که گراف به آن باید تقسیم شود مشخص نیست. در چنین شرایطی روش‌های خوشه‌بندی کارایی چندانی ندارند و نیاز است که دلیل مناسبی برای تعداد خوشه‌های بدست آمده ارائه کرد. از طرف دیگر گراف ممکن است ساختار سلسله مراتبی داشته باشد، برای مثال ممکن است چند سطح مختلف از گره‌ها وجود داشته باشد و خوشه‌های کوچکتر درون خوشه‌های بزرگتر قرار داشته باشند و یک رابطه‌ی سلسله مراتبی بین خوشه‌ها برقرار باشد. در چنین مواردی می‌توان از روش‌های خوشه‌بندی سلسله‌مراتبی استفاده کرد که سطوح مختلف گراف را مشخص می‌کنند.

گام اول خوشه‌بندی سلسله مراتبی تعریف یک معیار شباهت^{۱۳} است. بعد از انتخاب معیار شباهت، میزان شباهت بین هر دو گره صرف نظر از متصل بودن یا نبودن آن‌ها محاسبه می‌شود. بعد از این عمل یک ماتریس شامل مقایر شباهت بین گره‌ها بدست می‌آید.

الگوریتم‌های خوشه‌بندی سلسله مراتبی به دو دسته تقسیم می‌شوند :

۱. **الگوریتم‌های تراکمی**^{۱۴}: خوشه‌های با شباهت بالا

به طور تکراری ترکیب می‌شوند.

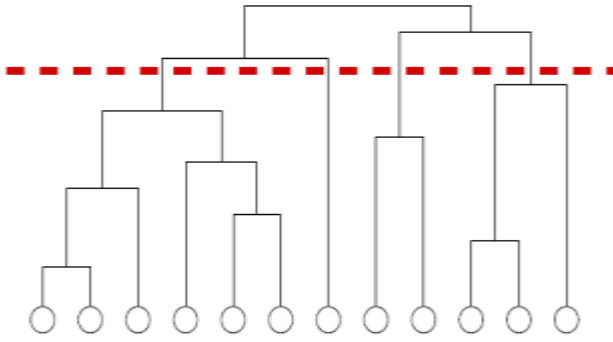
۲. **الگوریتم‌های تقسیم کننده**^{۱۵}: خوشه‌ها به طور

تکراری با حذف یال‌های بین گره‌های با شباهت کم به خوشه‌های کوچکتر تقسیم می‌شوند.

¹³ similarity

¹⁴ agglomerative

¹² centroid



شکل ۵-۲ نمایش سلسله‌مراتبی خوشه‌ها بوسیله‌ی دندروگرام

۵-۳- روش‌های طیفی

خوشه‌بندی طیفی شامل تمام روش‌هایی است که گراف را با استفاده از بردارهای ویژه‌ی ماتریس‌های مربوط به گراف یا ماتریس‌های مشتق شده خوشه‌بندی می‌کند [۱۰]. در خوشه‌بندی طیفی مجموعه‌ی ابتدایی از اشیا به مجموعه‌ای از نقاطی در فضا تبدیل می‌شوند که مختصات آنها همان عناصر بردار ویژه هستند. سپس مجموعه نقاط بدست آمده، با روش‌های عادی خوشه‌بندی مانند k-means خوشه‌بندی می‌شوند.

ماتریس لاپلاس^{۱۹} یکی از ماتریس‌های متداول برای خوشه‌بندی طیفی است. یک گراف را می‌توان با استفاده از مقادیر ویژه و بردارهای ویژه‌ی ماتریس لاپلاس آن گروه‌بندی کرد. کوچکترین مقدار ویژه‌ی یک ماتریس لاپلاس، صفر است و بردار ویژه‌ی متناظر با آن $[1, \dots, 1, 1]$ می‌باشد.

یک راه ساده برای یافتن دومین کوچکترین مقدار ویژه برای هر ماتریس متقارن محاسبه‌ی کمترین مقدار LX^T است که X یک بردار ستونی با n عضو است $([x_1, x_2, \dots, x_n])$ و شرایط زیر برقرار است:

۱. طول بردار X برابر با یک است.

۲. عمود بر بردار ویژه متناظر با کوچکترین مقدار ویژه است.

کمترین مقدار X برابر با دومین بردار ویژه است. وقتی L ماتریس لاپلاس یک گراف با n گره باشد، ما اطلاعات بیشتری داریم. بردار ویژه متناظر با کوچکترین مقدار ویژه برابر با یک است. بنابراین اگر X عمود بر یک باشد باید داشته باشیم:

$$X^T \mathbf{1} = \sum_{i=1}^n x_i = 0 \quad (۲-۵)$$

علاوه‌براین، برای ماتریس لاپلاس معادله‌ی زیر را داریم:

این دو کلاس به دو فرآیند متضاد اشاره دارند، الگوریتم‌های تراکمی پایین به بالا هستند چون در ابتدا هر گره را به عنوان یک خوشه‌ی مجزا در نظر می‌گیرند و در آخر تمام گره‌های گراف به عنوان یک خوشه در نظر گرفته می‌شوند. الگوریتم‌های تقسیم‌کننده رویکرد بالا به پایین دارند. این الگوریتم‌ها به ندرت استفاده می‌شوند بنابراین از بحث بیشتر روی آنها خودداری می‌کنیم.

از آنجایی که خوشه‌ها بر اساس شباهت مشترک‌شان ترکیب می‌شوند ضروری است که یک معیار برای تخمین میزان شباهت خوشه‌ها از روی ماتریس شباهت گره‌ها تعریف کنیم. در خوشه‌بندی تک پیوندی^{۱۶} شباهت بین دو خوشه کمترین مقدار x_{ij} است که گره‌ی i در یک خوشه و گره‌ی j در خوشه‌ی دیگر قرار دارد. در طرف مقابل، بیشترین مقدار x_{ij} برای خوشه‌بندی تمام پیوندی^{۱۷} استفاده می‌شود. در خوشه‌بندی میانگین پیوندی^{۱۸} از میانگین x_{ij} استفاده می‌شود.

مراحل کار می‌تواند توسط یک دندروگرام مانند شکل ۲-۷ نمایش داده شود. توقف خوشه‌بندی برای انتخاب یک خوشه‌بندی از گراف، بر اساس رسیدن به شرایط خاصی انجام می‌شود. برای مثال به تعداد خوشه‌های مورد نظر رسیده‌ایم یا یک معیار کیفیت را بهینه کرده‌ایم.

خوشه‌بندی سلسله‌مراتبی این مزیت را دارد که نیازی به اطلاعات خاصی در مورد گراف و تعداد خوشه‌ها نداریم. نتیجه‌ی این خوشه‌بندی به نوع معیار شباهت استفاده شده بستگی دارد. این روش همچنین یک ساختار سلسله‌مراتبی می‌سازد که در بسیاری از مواقع غیر واقعی است، زیرا ممکن است گراف یک ساختار سلسله‌مراتبی نداشته باشد. یکی دیگر از مشکلات این روش این است که یک گره‌ی با تنها یک همسایه، به عنوان یک خوشه‌ی مجزا در نظر گرفته می‌شود که غیر معقول است.

زمان اجرای الگوریتم‌های سلسله‌مراتبی برای حالت تک پیوندی $O(n^2)$ و برای حالات میانگین و کامل، $O(n^2 \log n)$ است.

¹⁵ divisive

¹⁶ Single linkage clustering

¹⁷ complete linkage clustering

¹⁸ average link clustering

¹⁹ laplacian

$$L = D - A$$

با استفاده از روابط بالا به این نتیجه می‌رسیم که $X^T L X$ برابر با مجموع $(x_i - x_j)^2$ به ازای تمام یال‌های (i, j) از گراف است. همچنین این شرط که X باید بر ۱ عمود باشد نیز باید برقرار باشد، بنابراین مجموع X_i ها نمی‌تواند صفر باشد. همچنین این شرط را که $\sum_{i=1}^n X_i^2 = 1$ نیز برقرار است، بنابراین تمام X_i ها نمی‌توانند صفر باشند. بنابراین X باید شامل هم مقادیر منفی و هم مقادیر مثبت باشد.

یک خوشه‌بندی از گراف را می‌توانیم با جدا کردن مجموعه گره‌های متناظر با i های منفی و مثبت از بردار X بدست بیاوریم. این انتخاب برابر بودن اندازه‌ی دو جامعه را تضمین نمی‌کند اما اندازه‌ها نزدیک به هم خواهند بود. همچنین مجموعه‌ی برشی شامل یال‌های کمی خواهد بود زیرا مقدار $(x_i - x_j)^2$ زمانی کمینه است که علامت X_i و X_j یکی باشد، یعنی هر دو گره در یک مجموعه قرار داشته باشند. در اینجا مقدار صفر به عنوان مقدار آستانه^{۲۰} برای جداسازی دو گروه در نظر گرفته شده است، اما این مقدار می‌تواند مقداری غیر از صفر نیز باشد.

۵-۳-۱- الگوریتم‌های تقسیم‌کننده^{۲۱}

یک راه ساده برای کشف اجتماعات در گراف، یافتن و حذف کردن یال‌هایی است که جوامع مختلف را به هم وصل می‌کند. با این کار جوامع از یکدیگر مجزا و مشخص خواهند شد. این ایده، فلسفه‌ی روش‌های تقسیم‌کننده است. نکته‌ی اصلی در اینجا یافتن ویژگی یال‌های بین جوامع است. یکی از معروف‌ترین روش‌های این دسته، روش ارائه شده توسط نیومن و گیروان است که در ادامه به آن می‌پردازیم.

۵-۴- Newman-Girvan روش

یکی از محبوب‌ترین الگوریتم‌های کشف اجتماعات، روش ارائه شده توسط نیومن و گیروان [۱۱] است. این روش از این جهت نیز مهم است که آغازگر یک دوره‌ی جدید در زمینه‌ی کشف اجتماعات در شبکه‌ها پیچیده بود. در این روش یال‌ها بر اساس

مقدار مرکزیت^{۲۲} آن برای حذف انتخاب می‌شود. مراحل این الگوریتم عبارتند از:

۱. محاسبه‌ی مرکزیت هر یال
۲. حذف یال با بیشترین مقدار مرکزیت
۳. محاسبه‌ی مجدد مرکزیت یال‌ها
۴. تکرار از مرحله‌ی ۲

نیومن و گیروان روی مفهوم بینابینی تمرکز کردند. بینابینی متغیری برای بیان میزان شرکت یک یال در یک فرآیند است. آنها سه فرآیند را در نظر گرفتند:

- بینابینی یال‌ها در کوتاه‌ترین مسیرها
- بینابینی یال‌ها در راه رفتن تصادفی^{۲۳}
- بینابینی یال‌ها در جریان^{۲۴}

در ادامه هر کدام از این سه مفهوم را مورد بررسی قرار خواهیم داد. بینابینی یال‌ها برابر با تعداد کوتاه‌ترین مسیرهای بین تمام جفت گره‌ها است که از یال مورد نظر عبور می‌کنند. محاسبه‌ی بینابینی تمام یال‌های یک گراف با استفاده از روش‌ها مبتنی بر جستجوی اول سطح در زمان $O(nm)$ یا $O(n^2)$ در صورت پراکنده بودن ماتریس، قابل محاسبه است. از دیدگاه انتشار اطلاعات این احتمال وجود دارد که اطلاعات به جای کوتاه‌ترین مسیر، از مسیرهای تصادفی حرکت کند. در این حالت بینابینی یک یال برابر با تعداد دفعاتی که یک راه رونده‌ی تصادفی از آن عبور کرده است. یک راه رونده‌ی تصادفی با احتمال مساوی به یال‌های متصل به یک گره حرکت می‌کند. یک جفت گره‌ی تصادفی s و t انتخاب می‌شوند و راه رونده از s شروع می‌کند و وقتی به t رسید متوقف می‌شود. در آخر احتمال عبور راه رونده از هر یال محاسبه می‌شود و این مراحل برای تمام جفت گره‌های ممکن تکرار و میانگین احتمال عبور هر یال محاسبه می‌شود. محاسبه‌ی بینابینی حرکت کننده‌ی تصادفی نیاز به معکوس کردن ماتریس مجاورت برای هر مرحله دارد که زمان آن $O(n^3)$ خواهد بود. برای هر جفت گره نیز میانگین تعداد عبور راه رونده را محاسبه کنیم که زمان آن $O(mn^2)$ خواهد بود. بنابراین مجموع زمان این الگوریتم $O(n^3)$ خواهد بود.

در بینابینی جریان گراف به عنوان یک شبکه از مقاومت‌ها در نظر گرفته می‌شود و یال‌ها دارای مقاومت واحد هستند. اگر یک اختلاف ولتاژ بین هر دو گره ایجاد شود در هر یال یک جریان

²² centrality

²³ random walk edge betweenness

²⁴ current flow edge betweenness

²⁰ threshold

²¹ divisive

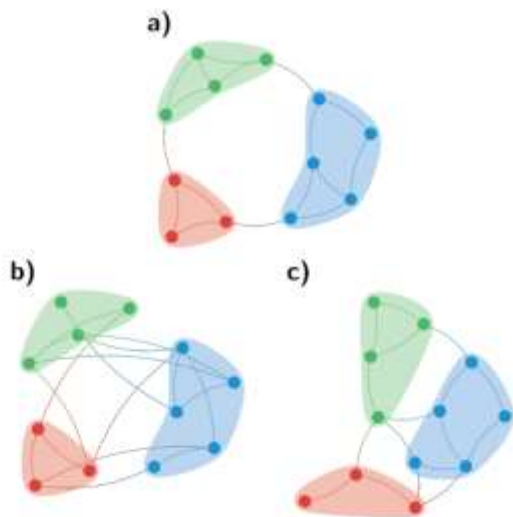
مشخص تمرکز دارد [۱۴] که به عنوان روش SBM^{۳۰} شناخته می‌شود.

اما این روش‌ها معمولاً به اهمیت واقعی یک خوشه‌بندی نمی‌پردازند و روشی اشتباه را در پیش می‌گیرند. ما دنبال یافتن احتمال یک خوشه‌بندی ثابت که شامل E یال داخلی می‌باشد نیستیم، بلکه در پی یافتن خوشه‌ای با حداقل E یال داخلی در یک گراف تصادفی هستیم.

در واقع روش‌های قبلی خوشه‌ها را ثابت در نظر می‌گرفتند و یال‌ها را به صورت تصادفی در این خوشه‌ها توزیع می‌کردند، در حالی که معیار اهمیت به دنبال یافتن خوشه‌ها در یک گراف تصادفی می‌باشد. به عبارت دیگر این روش‌ها این موضوع که یک جایگشت از گره‌ها نیز می‌تواند یک خوشه را تشکیل دهد نادیده گرفته‌اند در حالی که در معیار اهمیت، به این موضوع پرداخته می‌شود [۱۵].

تفاوت دو روش در شکل ۶-۱ بیان شده است.

خوشه‌بندی قسمت a را در نظر بگیرید. هدف تخمین زدن احتمال وجود خوشه در یک گراف تصادفی با در نظر گرفتن اهمیت خوشه است. در قسمت b و c نیز همان گراف نمایش داده شده است، با این تفاوت که در قسمت b گره‌ها در همان



شکل ۶-۱ تفاوت خوشه‌بندی با تمرکز بر بیشینه‌سازی اهمیت و سایر روش‌های مشابه [۱۵]

خوشه‌های قسمت a قرار گرفته‌اند اما در قسمت c خوشه‌ها یال‌های داخلی بیشتری دارند. در روش‌های موجود، خوشه‌ها ثابت در نظر گرفته می‌شوند و احتمال قرارگیری یال‌ها در هر خوشه محاسبه می‌شود، این مطلب در قسمت b نمایش داده

به وجود می‌آید که مقدار آن با قانون کیرشهف^{۲۵} قابل محاسبه است. این مراحل برای همگی جفت گره‌ها تکرار می‌شود. بینابینی جریان برای هر یال برابر با میانگین جریان عبوری از آن یال است. می‌توان نشان داد که این معیار برابر با همان معیار راه رونده تصادفی است، زیرا اختلاف ولتاژ و راه رونده به یک شکل روی یال‌ها حرکت می‌کنند. بنابراین پیچیدگی زمانی این روش نیز مانند راه رفتن تصادفی $O(n^3)$ است.

همان‌طور که مشاهده شد محاسبه بینابینی یال‌ها پیچیدگی محاسباتی کمتری نسبت به دو روش دیگر دارد. همچنین در کاربردهای عملی الگوریتم نیومن-گیروان^{۲۶} با استفاده از بینابینی یال‌ها نتایج بهتری نسبت به سایر معیارها می‌دهد. در گراف‌های دارای جوامع که گراف به سرعت به چند بخش تقسیم می‌شود، مرحله‌ی محاسبه‌ی مجدد مقادیر مرکزیت باید بین مولفه‌های همبندی از گراف انجام شود که توسط یالی که در این مرحله برای حذف انتخاب شده است انجام شود و مقدار مرکزیت سایر یال‌ها تغییر نمی‌کند. الگوریتم نیومن-گیروان بسیار کند است و تنها برای گراف پراکنده با کمتر از ۱۰۰۰۰ گره قابل اجراست.

۶- کشف اجتماعات در شبکه‌های اجتماعی یک لایه پویا با رویکرد بیشینه‌سازی اهمیت

۶-۱- معیار اهمیت

اگرچه معیار پیمانه‌ی تعداد یال‌های یک اجتماع را با یک گراف تصادفی مقایسه می‌کند، اما این معیار هیچ اطلاعاتی در مورد اهمیت یک خوشه‌بندی در اختیار قرار نمی‌دهد، زیرا یک گراف تصادفی و یک گراف پراکنده^{۲۷} بدون داشتن ساختار اجتماع نیز می‌توانند پیمانه‌ی بالایی داشته باشند. روش‌هایی برای اندازه‌گیری میزان اهمیت یک خوشه‌بندی ارائه شده‌اند. یکی از این روش‌ها، تحجیر^{۲۸} است که روی احتمال یافتن E یال در یک گراف تصادفی تمرکز دارد [۱۲]. یکی دیگر از معیارهای محلی نیز وجود دارد که درجات را ثابت نگه می‌دارد و به دنبال احتمال اتصال یال‌ها به یک اجتماع می‌گردد [۱۳] که در روش OSLOM [۱۱] استفاده شده است. روش دیگری نیز وجود دارد که روی راست‌نمایی^{۲۹} تولید یک گراف با خوشه‌بندی و توزیع درجه‌ی

²⁵ Kirchof's law

²⁶ Girvan-Newman

²⁷ Sparse

²⁸ Surprise

²⁹ Likelihood

³⁰ Stochastic Block Model

شده است. این روش‌ها احتمال وجود خوشه‌های با یال‌های داخلی بیشتر را در نظر نمی‌گیرند، در حالی که معیار اهمیت روی احتمال یافتن یک خوشه‌بندی متراکم در یک گراف تصادفی تمرکز دارد.

$$S(\sigma) = -\log \Pr(\sigma) = \sum_c \binom{n_c}{2} D(p_c \| p) \quad (4-6)$$

با این وجود، روش‌های موجود ممکن است نتایج خوبی داشته باشند، برای مثال محاسبه‌ی احتمال وجود E یال داخلی به نظر می‌رسد که به نتایج مناسبی منجر شود [15]. اگر احتمال یافتن یال‌های داخلی زیادی در یک خوشه‌بندی زیاد باشد، مطمئناً یافتن یک خوشه‌بندی با یال‌های زیاد نیز محتمل‌تر می‌باشد. از طرف دیگر، اگر احتمال یافتن یال‌های داخلی زیاد در یک خوشه‌بندی کم باشد، مطمئناً احتمال یافتن یک خوشه‌بندی با یال‌های زیاد نیز کم است. بین این دو مطلب فضای مبهمی وجود دارد که برای فهم دقیق آن نیاز به بررسی و تحلیل بیشتر می‌باشد.

در معیار اهمیت، روی احتمال یافتن خوشه‌هایی با تراکم بالا در یک گراف تصادفی تمرکز می‌شود. اندازه‌ی خوشه‌ها متناسب با اندازه‌ی گراف با n گره یعنی sn و چگالی q در نظر گرفته می‌شود که در آن $1 > s > 0$ می‌باشد.

$$D(q \| p) = H(q, p) - H(q) \quad (5-6)$$

که در آن $H(q)$ برابر است با:

$$H(q) = -q \log q - (1-q) \log(1-q) \quad (6-6)$$

$$H(q, p) = -q \log p - (1-q) \log(1-p) \quad (7-6)$$

۶-۲- تعمیم معیار اهمیت به شبکه‌های پویا

فرمول ارائه شده برای معیار اهمیت، در شبکه‌های ایستا کارایی دارد. برای استفاده از این معیار در شبکه‌های پویا در زمان نیاز است تغییراتی در آن اعمال شود. در این بخش به معرفی معیار اهمیت پویا، با استفاده از تعمیم معیار اهمیت ایستا می‌پردازیم. همینطور با استفاده از روش ابتکاری لووین^{۳۴} روشی جهت کشف اجتماعات در شبکه‌های اجتماعی یک‌لایه‌ی پویا ارائه خواهد شد.

۶-۲-۱- معیار اهمیت پویا

برای کشف اجتماعات در شبکه‌های پویا، نیاز است در معیار اهمیت تغییراتی ایجاد شود تا اضافه و یا کم شدن یال‌ها و گره‌ها نیز در نظر گرفته شود. با اضافه شدن یال به شبکه، ۴ عمل زیر قابل انجام است:

- ساختار اجتماعات را دست نخورده و بدون تغییر نگه داریم.
- دو اجتماع را به یک اجتماع ادغام کنیم.
- یک گره را به یک اجتماع موجود تخصیص دهیم.

این روش‌ها احتمال وجود خوشه‌های با یال‌های داخلی بیشتر را در نظر نمی‌گیرند، در حالی که معیار اهمیت روی احتمال یافتن یک خوشه‌بندی متراکم در یک گراف تصادفی تمرکز دارد.

با این وجود، روش‌های موجود ممکن است نتایج خوبی داشته باشند، برای مثال محاسبه‌ی احتمال وجود E یال داخلی به نظر می‌رسد که به نتایج مناسبی منجر شود [15].

اگر احتمال یافتن یال‌های داخلی زیادی در یک خوشه‌بندی زیاد باشد، مطمئناً یافتن یک خوشه‌بندی با یال‌های زیاد نیز محتمل‌تر می‌باشد. از طرف دیگر، اگر احتمال یافتن یال‌های داخلی زیاد در یک خوشه‌بندی کم باشد، مطمئناً احتمال یافتن یک خوشه‌بندی با یال‌های زیاد نیز کم است. بین این دو مطلب فضای مبهمی وجود دارد که برای فهم دقیق آن نیاز به بررسی و تحلیل بیشتر می‌باشد.

در معیار اهمیت، روی احتمال یافتن خوشه‌هایی با تراکم بالا در یک گراف تصادفی تمرکز می‌شود. اندازه‌ی خوشه‌ها متناسب با اندازه‌ی گراف با n گره یعنی sn و چگالی q در نظر گرفته می‌شود که در آن $1 > s > 0$ می‌باشد.

احتمال یافتن یک زیرگراف با اندازه‌ی n_c و چگالی q در یک گراف با اندازه‌ی n و چگالی p برابر است با:

$$\Pr(S(n_c, q) \subseteq G(n, p)) = e^{-\binom{n_c}{2} D(q \| p)} \quad (1-6)$$

که در این فرمول $D(q \| p)$ برابر با واگرایی کولبک-لیبلر^{۳۱} می‌باشد و از طریق فرمول (۵-۲) قابل محاسبه است.

$$D(q \| p) = q \log \frac{q}{p} + (1-q) \log \frac{1-q}{1-p} \quad (2-6)$$

برای مقادیر $p \neq q$ این احتمال یک توزیع گاوسی^{۳۲} با نرخ متناسب با فاصله‌ی بین p و q می‌باشد که با فرمول واگرایی کولبک-لیبلر بیان شده است. طبق این فرمول هرچه اندازه‌ی زیرگراف بیشتر باشد، احتمال یافتن زیرگراف با چگالی متفاوت با p کمتر می‌شود.

با ترکیب این احتمال‌ها، فرمول (۵-۳) برای محاسبه‌ی احتمال یافتن یک خوشه‌بندی در یک گراف تصادفی به دست می‌آید.

$$\Pr(\sigma) = \prod_c \exp\left(-\binom{n_c}{2} D(p_c \| p)\right) \quad (3-6)$$

³³ Entropy

³⁴ Louvain Heuristic

³¹ Kullback-Leibler divergence

³² Gaussian

- یک اجتماع جدید ایجاد کنیم و دو گره را به آن تخصیص دهیم.
- یال‌های اضافه شده به شبکه می‌توانند یکی از انواع زیر باشند:
- یال‌های داخل اجتماع: هر دو گرهی این یال از قبل در شبکه موجود هستند و هر دو در یک اجتماع مشترک قرار دارند.
- یال‌های بین اجتماع: دو گرهی این نوع یال از قبل در شبکه وجود دارند، اما در دو اجتماع مختلف قرار دارند.
- یال‌های نیمه جدید: یکی از گره‌های این نوع یال، به تازگی به شبکه اضافه شده است.
- یال‌های جدید: هر دو گرهی این یال به تازگی به شبکه اضافه شده‌اند.

در ادامه به بررسی هر نوع یال و اعمال قابل انجام در مورد هر حالت می‌پردازیم.

۶-۲-۱-۱- یال‌های داخل اجتماع

در این حالت، دو گرهی یال اضافه شده به لایه‌ی k' از قبل در شبکه موجود بوده و هر دو گره در یک اجتماع یکسان قرار دارند. در نتیجه تراکم یال‌های داخل اجتماع افزایش خواهد یافت و باعث افزایش معیار اهمیت شبکه خواهد شد. بنابراین در این حالت بهترین انتخاب این است که ساختار اجتماعات دست نخورده باقی بماند.

۶-۲-۱-۲- یال‌های بین دو اجتماع

در این حالت، دو گرهی یال اضافه شده به لایه‌ی k' از قبل در شبکه موجود بوده و دو گره متعلق به اجتماعات مختلفی هستند. این یال را به صورت (i, j, w_{ij}) در نظر می‌گیریم که گره‌های i و j به ترتیب متعلق به اجتماعات C_i و C_j می‌باشند و وزن این یال برابر با w_{ij} می‌باشد.

در این حالت دو عمل قابل انجام است:

۱. ساختار اجتماعات بدون تغییر باقی بماند.
۲. دو اجتماع با یکدیگر ادغام شوند.

در صورت انتخاب عمل ۱ اهمیت جدید برابر با مقدار زیر خواهد بود:

$$(۸-۶)$$

$$S_1(\sigma) = \sum_c \binom{n_c}{2} D(p_c \| p') = \sum_{c \neq c_j} \binom{n_c}{2} D(p_c \| p') + \binom{n_{c_j}}{2} D(p_{c_j} \| p') + \binom{n_{c_j}}{2} D(p_{c_j} \| p')$$

$$p' = \frac{m + w_{ij}}{\binom{n}{2}}$$

در صورت انتخاب عمل ۲ مقدار اهمیت جدید برابر با مقدار زیر خواهد بود:

$$(۹-۶)$$

$$S_2(\sigma) = \sum_{c \neq c_i, c_j} \binom{n_c}{2} D(p_c \| p') + \binom{n_{c_i} + n_{c_j}}{2} D(p_{c_{i,j}} \| p')$$

$$p' = \frac{m + w_{ij}}{\binom{n}{2}}, p_{c_{i,j}} = \frac{m_{c_{i,j}} + w_{ij}}{\binom{n_i + n_j}{2}}$$

تفاوت میزان اهمیت دو عمل بالا برابر است با:

$$(۱۰-۶)$$

$$\Delta S(\sigma) = \binom{n_{c_i} + n_{c_j}}{2} D(p_{c_{i,j}} \| p') - \binom{n_{c_i}}{2} D(p_{c_i} \| p') - \binom{n_{c_j}}{2} D(p_{c_j} \| p')$$

اگر $\Delta S(\sigma) > 0$ باشد عمل ۲ انتخاب می‌شود، در غیر این صورت عمل ۱ انتخاب خواهد شد.

۶-۲-۱-۳- یال‌های نیمه جدید

در این حالت، یکی از گره‌های متصل به یال جدید، از قبل در شبکه موجود بوده و یکی از گره‌ها جدید است. این یال را به صورت (i, j, w_{ij}) در نظر می‌گیریم که گرهی i گرهی جدید، گرهی j به ترتیب متعلق به اجتماع C_j و وزن این یال برابر با w_{ij} می‌باشد.

در این حالت دو عمل قابل انجام است:

۱. گرهی i به اجتماع مربوط به گرهی j اضافه شود.
۲. یک اجتماع جدید برای گرهی i ایجاد شود.

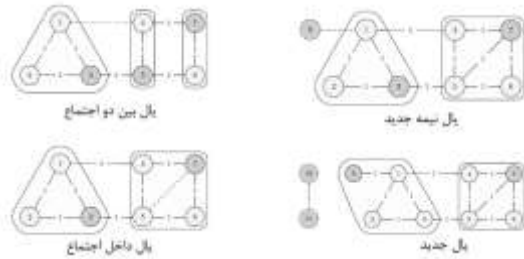
در صورت انتخاب عمل ۳ مقدار اهمیت جدید برابر با مقدار زیر خواهد بود:

$$(۱۱-۶)$$

$$S_3(\sigma) = \sum_{c \neq c_j} \binom{n_c}{2} D(p_c \| p') + \binom{n_{c_j} + 1}{2} D(p_{c_j} \| p')$$

$$p' = \frac{m + w_{ij}}{\binom{n+1}{2}}, p_{c_j} = \frac{m_{c_j} + w_{ij}}{\binom{n_j + 1}{2}}$$

در صورت انتخاب عمل ۴ مقدار اهمیت جدید برابر با مقدار زیر خواهد بود:



شکل ۳-۶ انواع یال‌های افزایشی به گراف

برای پشتیبانی کامل از شبکه‌های پویا، نیاز است حذف یال‌ها نیز در نظر گرفته شود. با توجه به نوع یال حذف شده، ۴ عمل زیر قابل انجام است:

- ۵- ساختار اجتماعات بدون تغییر باقی بماند.
 - ۶- یک اجتماع به دو یا بیشتر اجتماع دیگر شکسته شود.
 - ۷- یک یال لبه از یک اجتماع حذف شود.
 - ۸- اجتماعی که با دو گره یال حذف شده ایجاد شده است حذف شود.
- درواقع عمل‌های ۲ و ۸ تعداد اجتماعات را کاهش می‌دهند، در حالی که عمل‌های ۴ و ۶ تعداد اجتماعات را افزایش می‌دهند. عمل‌های ۳، ۵ و ۷ نیز تغییری در تعداد اجتماعات ایجاد نمی‌کنند.
- الگوریتم ۶-۱ شبه کد روش DySig را نمایش می‌دهد. ورودی الگوریتم گراف $G(V, E)$ و لیست یال‌های اضافه شده به گراف یعنی A و یا کم شده از گراف یعنی R می‌باشد. برای ذخیره‌ی ساختار اجتماعات، از دو گراف استفاده شده است، C_{II} که ساختار گراف اصلی را نگهداری می‌کند و C_{III} که ساختار گراف حاصل از ادغام اجتماعات را نگهداری می‌کند.

۶-۵- تشریح مراحل الگوریتم پیشنهادی

در ابتدا خوشه‌های گراف اولیه با استفاده از روش ابتکاری مبتنی بر اهمیت تعیین می‌شود. سپس یال‌های موردنظر برای حذف و

الگوریتم ۶-۱ شبه کد DySig

دو گام تا زمانی که بهبودی در پیمانی مشاهده نشود تکرار می‌شوند.

۶-۴- الگوریتم پیشنهادی برای کشف اجتماعات شبکه‌های یک‌لایه‌ی پویا با روش ابتکاری مبتنی بر معیار اهمیت

در این بخش روشی ابتکاری برای کشف اجتماعات در شبکه‌های یک‌لایه‌ی پویا با تمرکز بر بیشینه سازی معیار اهمیت ارائه خواهد شد. در ادامه‌ی این نوشته این روش تحت عنوان DySig شناخته خواهد شد.

روش ابتکاری ارائه شده مانند روش لووین از دو گام تکراری تشکیل شده است. برای پشتیبانی از شبکه‌های یک‌لایه‌ی پویا تغییراتی در معیار اهمیت داده شد تا بتوان پس از اضافه و حذف یال از شبکه ساختار اجتماعات را با هزینه‌ای اندک به روز رسانی کرد. در واقع میزان تغییر در معیار اهمیت پس از انتقال محلی گره‌ها بین اجتماعات محاسبه شده و اجتماع مناسب برای انتقال گره انتخاب می‌شود. انتقال مجموعه گره‌ی i با n_i گره و e_{ir} یال در اجتماع r و e_{is} یال در اجتماع s از اجتماع r به اجتماع s تفاوت اهمیتی مطابق با فرمول ۵-۱۷ ایجاد می‌کند.

بعد از اضافه کردن یال جدید به گراف، با توجه به نوع یال ۴ عمل زیر قابل انجام می‌باشد:

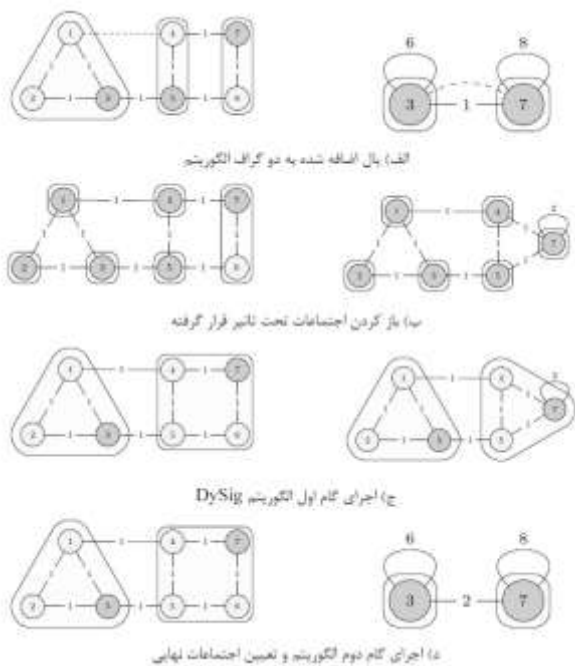
- ۱- ساختار اجتماعات بدون تغییر باقی بماند.
- ۲- دو اجتماع با یکدیگر ادغام شوند.
- ۳- گره‌های جدید به یک اجتماع موجود تخصیص داده شوند.
- ۴- یک اجتماع جدید ساخته شود و گره‌های جدید به آن اختصاص داده شوند.

شکل ۳-۶ چهار نوع یال ممکن را نمایش می‌دهد.

(۶-۱۷)

$$\Delta S(\sigma) = \binom{n_r}{2} D(q_r \| p) - \binom{n_r - n_i}{2} D(q_r' \| p) - \binom{n_s}{2} D(q_s \| p) + \binom{n_s + n_i}{2} D(q_s' \| p),$$

$$q_r' = \frac{m_r - e_{ir}}{\binom{n_r - n_i}{2}}, q_s' = \frac{m_s - e_{is}}{\binom{n_s + n_i}{2}}$$



شکل ۶-۴ مراحل اجرای الگوریتم DySig بعد از اضافه شدن یک یال بین اجتماع

برای بررسی میزان صحت تشخیص اجتماعات، از معیار NMI استفاده شده است.

معیار (Normalized Mutual Information) NMI یک معیاری است که برای اندازه‌گیری تطابق بین دو تقسیم بندی یا اجتماع استفاده می‌شود. این معیار بر پایه اطلاعات مشترک بین دو تقسیم بندی می‌تواند میزان همبستگی و تشابه بین آن‌ها را سنجیده و ارزیابی کند. NMI به دو عامل اصلی بر می‌گردد: آنتروپی و اطلاعات مشترک. آنتروپی یک مفهوم از حافظه‌ی تصادفی است و نشان می‌دهد که چقدر می‌توانیم انتظار داشته باشیم که یک رویداد خاص در یک تقسیم بندی رخ دهد. آنتروپی برابر با صفر است در صورتی که تقسیم بندی دقیقاً یک اجتماع را تشکیل دهد، و بیشترین مقدار خود را در صورتی بدست می‌آورد که تقسیم بندی هیچ الگویی نداشته باشد و به طور کامل تصادفی باشد.

اطلاعات مشترک میزان تطابق بین دو تقسیم بندی را نشان می‌دهد و بر اساس اندازه و جهش‌های مشترک بین دو تقسیم بندی محاسبه می‌شود. NMI با استفاده از این دو عامل، یک مقدار نرمال‌شده را بر می‌گرداند که بین ۰ و ۱ واقع می‌شود. این مقدار به ما می‌گوید که تقسیم بندی‌ها چقدر با یکدیگر همخوانی دارند، به این صورت که مقدار صفر به معنای هیچ تطابق و مقدار یک به معنای تطابق کامل است. بنابراین، با استفاده از معیار NMI می‌توانیم تشابه و همبستگی بین دو

```

1:  $V = \{n_1, n_2, \dots, n_n\}$ ,  $E = \{(i, j_1), (i, j_2), \dots, (i, j_m)\}$ 
2:  $A \leftarrow \text{array}(\{i_1, j_1\}, \dots, \{i_m, j_m\})$ 
3:  $R \leftarrow \text{array}(\{i_1, j_1\}, \dots, \{i_m, j_m\})$ 
4: procedure MAIN( $G = (V, E), A, R$ )
5:  $C_{II} \leftarrow \{C_1, C_2, \dots, C_n\}$ ,  $C_{ul} \leftarrow \{\}$ ,  $C_{max} \leftarrow C_{II}$ 
6: INITPARTITION( $C_{max}$ )
7:  $\text{sig} \leftarrow \text{Significance}(C_{max})$ ,  $\text{old\_sig} \leftarrow 0$ 
8:  $m \leftarrow 1$ ,  $n \leftarrow 1$ 
9: while ( $\text{sig} \geq \text{old\_sig} \vee m \leq |A| \vee n \leq |R|$ ) do
10:  $C_{max} \leftarrow \text{OSELEVEL}(C_{max})$ 
11:  $(n, c) \leftarrow \text{COMMUNITYCHANGEDNODES}(C_{II}, C_{max})$ 
12:  $C_{II} \leftarrow \text{UPDATECOMMUNITIES}(C_{II}, n, c)$ 
13:  $\text{old\_sig} \leftarrow \text{sig}$ ,  $\text{sig} \leftarrow \text{Significance}(C_{II})$ 
14:  $C_{ul} \leftarrow \text{PARTITIONTOGRAPH}(C_{II})$ 
15: if  $m \leq |A|$  then
16:    $(\text{src}, \text{dest}) \leftarrow A[m]$ 
17:    $n_{nodes} \leftarrow \text{AFFECTEDBYADDITION}(\text{src}, \text{dest}, C_{II})$ 
18:    $C_{II} \leftarrow \text{ADDEDGE}(\text{src}, \text{dest}, C_{II})$ 
19:    $C_{II} \leftarrow \text{DISBANDCOMMUNITIES}(C_{II}, n_{nodes})$ 
20:    $C_{ul} \leftarrow \text{SYNCCOMMUNITIES}(C_{II}, C_{ul}, n_{nodes})$ 
21: end if
22: if  $n \leq |R|$  then
23:    $(\text{src}, \text{dest}) \leftarrow R[n]$ 
24:    $n_{nodes} \leftarrow \text{AFFECTEDBYREMOVAL}(\text{src}, \text{dest}, C_{II})$ 
25:    $C_{II} \leftarrow \text{REMOVEEDGE}(\text{src}, \text{dest}, C_{II})$ 
26:    $C_{II} \leftarrow \text{DISBANDCOMMUNITIES}(C_{II}, n_{nodes})$ 
27:    $C_{ul} \leftarrow \text{SYNCCOMMUNITIES}(C_{II}, C_{ul}, n_{nodes})$ 
28: end if
29:  $C_{max} \leftarrow C_{ul}$ ,  $m \leftarrow m + 1$ ,  $n \leftarrow n + 1$ 
30: end while
31: end procedure

```

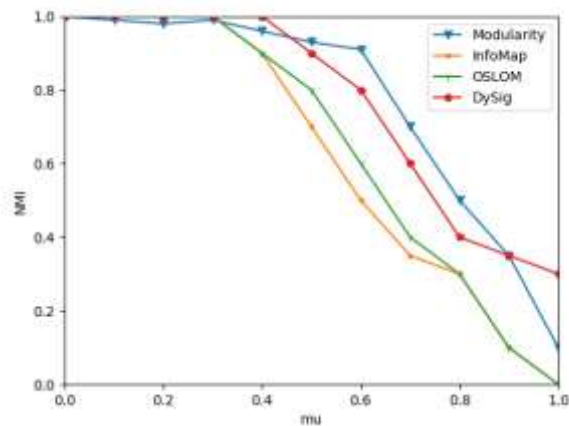
اضافه شدن به صورت متوالی در شبکه اعمال می‌شوند. با هر تغییر در مجموعه یال‌های گراف، اجتماعاتی که مربوط به یال‌های تغییر یافته هستند از گراف C_{II} استخراج می‌شوند و با استفاده از گراف C_{II} به گره‌های سازنده تجزیه می‌شوند. سپس بعد از حذف یا اضافه کردن یال مدنظر، الگوریتم DySig تنها روی گره‌های حاصل از تجزیه‌ی اجتماعات تغییر یافته اجرا می‌شود و ساختار جدید اجتماعات شکل می‌گیرد و در نهایت، گراف‌های C_{II} و C_{ul} به روز رسانی می‌شوند.

شکل ۶-۴ تمام مراحل الگوریتم DySig را بعد از اضافه شدن یک یال بین اجتماع به گراف نمایش می‌دهد. روش ابتکاری لووین نسبت خطی با تعداد یال‌ها دارد و از مرتبه‌ی $O(m)$ یعنی تعداد یال‌ها می‌باشد. در حالت پویا، تنها یک یا دو اجتماعی که تحت تاثیر حذف و یا اضافه کردن یال قرار گرفته‌اند مجدداً خوشه‌بندی می‌شوند که این عمل نیز در بدترین حالت از مرتبه‌ی $O(m)$ خواهد بود. البته در عمل اندازه‌ی یک اجتماع بسیار کوچکتر از اندازه‌ی کل گراف خواهد بود و بنابراین زمان اجرا کمتر از حالت حدی خواهد بود.

۶-۶-۶ ارزیابی الگوریتم

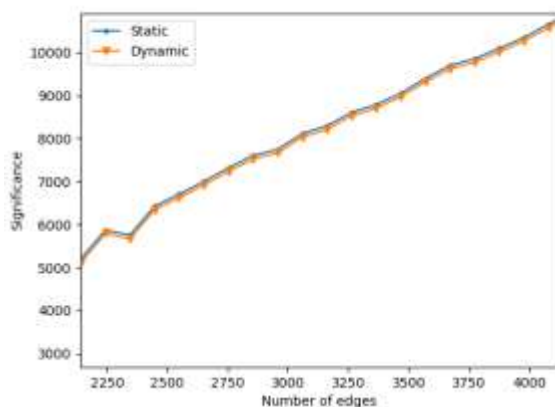
۶-۶-۶-۱ ارزیابی روی مجموعه داده‌های ساختگی

در این بخش با استفاده از مجموعه داده‌ی تولید شده توسط Benchmark LFR به بررسی نتایج حاصل از الگوریتم DySig می‌پردازیم.



شکل ۵-۶ مقایسه‌ی NMI الگوریتم‌های Modularity، Infomap، OSLOM و DySig روی LFR Benchmark

برای بررسی کیفیت پویایی الگوریتم DySig نیز از مجموعه داده‌ی تولید شده توسط LFR Benchmark استفاده شد. گراف تولید شده شامل ۵۰۰ گره و ۴۱۱۸ یال می‌باشد. در ابتدا نیمی از یال‌ها به عنوان گراف پایه خوشه‌بندی شد و سپس در ۲۰ بازه‌ی زمانی مختلف یال‌ها به صورت تصادفی به گراف اضافه شدند و الگوریتم پویا و ایستا روی آن اجرا شدند. میزان تفاوت در معیار اهمیت در دو حالت ایستا و پویا بیانگر کیفیت الگوریتم پویا می‌باشد. نتیجه‌ی این آزمایش در شکل ۶-۶ نمایش داده شده است. مطابق انتظار، الگوریتم پویا تفاوتی با الگوریتم ایستا ندارد و دو نمودار منطبق بر یکدیگر هستند.



شکل ۶-۶ مقایسه‌ی الگوریتم پویا و ایستای مبتنی بر معیار اهمیت

۶-۶-۲- ارزیابی روی مجموعه داده‌ی حقیقی

برای بررسی زمان اجرای روش DySig ابتدا روش را روی مجموعه داده‌ی شبکه استناد نظریه فیزیک انرژی‌های بالا^{۳۶} که یک شبکه‌ی بدون جهت با ۲۷۷۷۰ گره و ۳۵۲۸۰۷ یال می‌باشد

تقسیم بندی یا اجتماع را اندازه‌گیری کنیم و معیار NMI برای بررسی صحت تشخیص اجتماعات می‌تواند به ما اطلاعات مفیدی درباره تطابق بین دو تقسیم بندی یا اجتماعات مختلف در دست دهد. با محاسبه NMI بین دو تقسیم بندی، می‌توانیم بفهمیم که آیا اجتماعات در هر دو تقسیم بندی به طور مشابه شناسایی شده‌اند یا خیر.

اگر NMI برابر با ۱ باشد، این نشانگر تطابق کامل بین دو تقسیم بندی است، به این معنی که هر دو تقسیم بندی دقیقاً یکسان هستند. اگر مقدار NMI بسیار نزدیک به ۱ باشد، نشان‌دهنده تطابق قوی بین دو تقسیم بندی است. در مقابل، اگر NMI به سمت صفر حرکت کند، تطابق کم و به طور تصادفی بین دو تقسیم بندی نشان می‌دهد. به عنوان مثال، فرض کنید دو روش مختلف برای تقسیم بندی یک مجموعه داده را ارائه داده‌ایم. با محاسبه معیار NMI می‌توانیم بفهمیم که این دو تقسیم بندی چقدر به یکدیگر نزدیک هستند و در چه میزان اجتماعات مشابهی را شناسایی کرده‌اند. این معیار می‌تواند ما را در انتخاب بهترین تقسیم بندی برای مجموعه داده‌ها یا ارزیابی دقت الگوریتم‌ها کمک کند.

فرمول محاسبه معیار NMI بر اساس آنتروپی و اطلاعات مشترک به صورت زیر است:

$$NMI = \frac{2 \times I(X;Y)}{H(X)+H(Y)}$$

که در آن X و Y دو تقسیم بندی یا اجتماع مورد بررسی هستند. $I(X;Y)$ اطلاعات مشترک بین X و Y است که برابر است با آنتروپی X منهای آنتروپی شرطی X به شرط Y (یا بالعکس). $H(X)$ و $H(Y)$ به ترتیب آنتروپی تقسیم بندی X و تقسیم بندی Y هستند.

برای این منظور مجموعه داده‌های مختلفی با پارامترهای مختلف μ تولید شد و اجتماعات آن توسط الگوریتم‌های مختلفی استخراج شد. برای مقایسه‌ی عملکرد الگوریتم DySig از الگوریتم‌های Modularity، Infomap، OSLOM استفاده شده است. نتیجه‌ی اجرای این الگوریتم‌ها در شکل ۵-۶ قابل مشاهده است.

همان طور که مشاهده می‌شود برای پارامترهای وضوح مختلف عملکرد الگوریتم DySig بهتر از Infomap [۲۲] و OSLOM [۲۳] و قابل مقایسه با Modularity [۱۱] می‌باشد.

³⁶ High-energy physics theory citation network

شده در این مقاله از تمام حالات مختلف از جمله حذف یال پشتیبانی می‌کند.

همچنین اجتماعات کشف شده با این روش کاملاً منطبق بر اجتماعات بهینه هستند و معیار اهمیت در هر زمان، برابر با مقدار بهینه می‌باشد. روش ارائه شده در این مقاله نیز مانند دو روش قبلی تنها تغییرات شبکه را در نظر گرفته و ساختار اجتماعات بخش تغییر یافته را به روز رسانی می‌کند و در نتیجه از سرعت بالایی برای کشف اجتماعات برخوردار می‌باشد.

۷- منابع

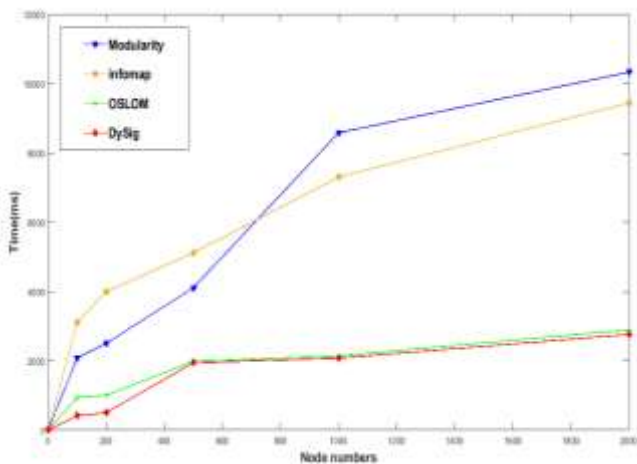
- [1] Kim, Jungeun, and Jae-Gil Lee. "Community detection in multi-layer graphs: A survey." *ACM SIGMOD Record* 44, no. 3 (2015): 37-48.
- [2] Berlingerio, Michele, Michele Coscia, and Fosca Giannotti. "Finding and characterizing communities in multidimensional networks." In *2011 international conference on advances in social networks analysis and mining*, pp. 490-494. IEEE, 2011.
- [3] Börner, Katy, Michael Conlon, Jon Corson-Rikert, and Ying Ding. "VIVO: A semantic approach to scholarly networking and discovery." *Synthesis lectures on the Semantic Web: theory and technology* 7, no. 1 (2012): 1-178.
- [4] Fortunato, Santo, and Darko Hric. "Community detection in networks: A user guide." *Physics reports* 659 (2016): 1-44.
- [5] Newman, Mark EJ. "Spectral methods for community detection and graph partitioning." *Physical Review E* 88, no. 4 (2013): 042822.
- [6] Loe, Chuan Wen, and Henrik Jeldtoft Jensen. "Comparison of communities detection algorithms for multiplex." *Physica A: Statistical Mechanics and its Applications* 431 (2015): 29-45.
- [7] Whang, Joyce Jiyoung, David F. Gleich, and Inderjit S. Dhillon. "Overlapping community detection using neighborhood-inflated seed expansion." *IEEE Transactions on Knowledge and Data Engineering* 28, no. 5 (2016): 1272-1284.
- [8] Leskovec, Jure, Kevin J. Lang, Anirban Dasgupta, and Michael W. Mahoney. "Statistical properties of community structure in large social and information networks." In *Proceedings of the 17th international conference on World Wide Web*, pp. 695-704. 2008.
- [9] MacQueen, J. "Classification and analysis of multivariate observations." In *5th Berkeley Symp. Math. Statist. Probability*, pp. 281-297. 1967.
- [10] Newman, Mark EJ. "Spectral methods for community detection and graph partitioning." *Physical Review E* 88, no. 4 (2013): 042822.
- [11] Newman, Mark EJ. "Modularity and community structure in networks." *Proceedings of the national academy of sciences* 103, no. 23 (2006): 8577-8582.
- [12] Aldecoa, Rodrigo, and Ignacio Marín. "Surprise maximization reveals the community structure of complex networks." *Scientific reports* 3, no. 1 (2013)

آزمایش می‌کنیم. نتایج بعد از اضافه کردن ۱، ۲، ۵، ۱۰ و ۲۰ دسته یال ۱۰۰ تایی برای هر افزایش، محاسبه شده اند.

جدول ۶-۱ نتایج حاصل از مقایسه‌ی روش DySig با روش‌های Modularity، Infomap و OSLOM را نمایش می‌دهد. در این آزمایش پس از اضافه شدن ۱۰۰، ۲۰۰، ۵۰۰، ۱۰۰۰ و ۲۰۰۰ گره و اجرای الگوریتم‌ها زمان اجرا بر حسب میلی ثانیه اندازه‌گیری شده است.

جدول ۶-۱ مقایسه‌ی زمان اجرای الگوریتم‌های مختلف

	۲۰۰۰	۱۰۰۰	۵۰۰	۲۰۰	۱۰۰	
Modularity	۱۰۳۲۴	۸۵۷۴	۴۱۰۳	۲۵۰۱	۲۰۸۶	
Infomap	۹۴۳۳	۷۳۰۱	۵۱۱۴	۳۹۹۴	۳۱۰۷	
OSLOM	۲۸۸۸	۲۱۴۰	۱۹۸۴	۱۰۰۳	۹۴۶	
DySig	۲۷۵۹	۲۰۷۴	۱۹۴۲	۵۰۴	۴۱۱	



شکل ۶-۷ مقایسه‌ی زمان اجرای الگوریتم‌های مختلف

۶-۷- جمع بندی و نتیجه گیری

در این مقاله سعی بر این بود تا رویکردی جدید مشابه معیارهای مشابهت اجتماع در شبکه‌های اجتماعی به منظور بهبود این روش‌ها ارائه گردد. برای حل مشکلاتی مانند محدودیت وضوح معیار پیمانگی، از معیار اهمیت استفاده شد. همچنین دو روش قبل تنها برای شبکه‌های افزایشی کاربرد داشتند، اما روش ارائه

- [13] Lancichinetti, Andrea, Filippo Radicchi, and José J. Ramasco. "Statistical significance of communities in networks." *Physical Review E* 81, no. 4 (2010): 046110.
- [14] Dubik, Mikael. "A comparative evaluation of state-of-the-art community detection algorithms for multiplex networks." (2017).
- [15] Traag, Vincent A., Gautier Krings, and Paul Van Dooren. "Significant scales in community structure." *Scientific reports* 3, no. 1 (2013): 1-10.
- [16] Dubik, Mikael. "A comparative evaluation of state-of-the-art community detection algorithms for multiplex networks." (2017).
- [17] Fernandes, Andreia, Patrícia CT Gonçalves, Pedro Campos, and Catarina Delgado. "Centrality and community detection: a co-marketing multilayer network." *Journal of Business & Industrial Marketing* (2019).
- [18] Bakhthemmat, Ali, and Mohammad Izadi. "Communities detection for advertising by futuristic greedy method with clustering approach." *Big Data* 9, no. 1 (2021): 22-40.
- [19] Su, Yan. "Accurate Marketing Algorithm of Network Video Based on User Big Data Analysis." *Mathematical Problems in Engineering* 2022 (2022).
- [20] Leiva, Fabiola Herrera, Romina Torres, Orietta Nicolis, and Rodrigo Salas. "Characterization of the chilean public procurement ecosystem using social network analysis." *IEEE Access* 8 (2020): 138846-138858.
- [21] Sharma, Prem Sagar, Divakar Yadav, and R. N. Thakur. "Web Page Ranking using Web Mining Techniques: A comprehensive survey." *Mobile Information Systems* 2022 (2022).
- [22] Yu-Liang, Lu, Tian Jie, Guo Hao, and Wang Yu. "Infomap based community detection in weibo following graph." In 2012 Second International Conference on Instrumentation, Measurement, Computer, Communication and Control, pp. 1220-1222. IEEE, (2012).
- [23] Lancichinetti, Andrea, Filippo Radicchi, José J. Ramasco, and Santo Fortunato. "Finding statistically significant communities in networks." *PloS one* 6, no. 4 (2011): e18961.