



دانشگاه آزاد اسلامی قزوین - واحد علوم و تحقیقات

دانشکده مهندسی کامپیوتر

سمینار دوره کارشناسی ارشد مهندسی کامپیوتر - گرایش نرم افزار

کشف انجمن در شبکه های اجتماعی

نگارش

رزا شفاف

استاد راهنما

جناب آقای دکتر محمد صنیعی

زمستان ۱۳۹۳

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

تقدیم به

روح پاک پدرم که عالمانه به من آموخت تا چگونه در عرصه
زندگی، ایستادگی را تجربه نمایم

و به مادرم، دریای بی کران فداکاری و عشق که وجودم برایش همه
رنج بود و وجودش برایم همه مهر

و

مردانه های لطیف برادری که پناهم داد

با عشق تقدیم به مادر، برادر و روح پاک پدرم

تشکر و قدردانی

از استادی که با روشنه های اندیشه اش چراغ راهم شد

با سپاس فراوان از استاد راهنمای فرهیخته ام جناب آقای دکتر محمد صنیعی که در طول مدت انجام این سمینار از رهنمود های علمی و اخلاقی ایشان بهره مند شدم و درگاه خداوند بزرگ را شاکرم که افتخار شاگردی ایشان را نصیبم نمود.

چکیده

درک پیدایش و تکمیل جوامع، یک موضوع بسیار طولانی مدت تحقیقاتی در جامعه‌شناسی به واسطه ارتباط بنیادین آن با مطالعه توسعه شهری، جرم‌شناسی، بازاریابی اجتماعی و سایر جنبه‌های مختلف است. با افزایش محبوبیت شبکه‌های خدمات اجتماعی آنلاین مانند فیس‌بوک، مطالعه ساختارها اهمیت بسیار بیشتری پیدا کرده است. شناخت و تشخیص جوامع نه تنها دارای اهمیت بسیاری است، بلکه همچنین کاربردهای مهمی نیز دارد. مثلاً، برای بازاریابی مؤثر و کارآمد به شکل آنلاین، مثل تبلیغات آنلاین، استفاده از استراتژی‌های بازاریابی ویروسی (مویرگی)، شناسایی جوامع در شبکه اجتماعی می‌تواند به هدف گذاری‌های بسیار دقیق‌تر و نتایج بازاریابی بهتر بیانجامد. در این بین، پروفایل‌های کاربری آنلاین و سایر اطلاعات ساختاری می‌تواند برای کشف سگمنت‌ها و گروه‌های کاربری اهمیت داشته باشد. به شکل کلی می‌توان گفت که در سال‌های اخیر شاهد آن بوده‌ایم که شبکه جهانی وب به یک رسانه اجتماعی شکوفا بدل شده است که به افراد امکان تقسیم و اشتراک نظرات، تجربیات و تخصص را حتی با فشردن یک دکمه می‌دهد. با استفاده گسترده از سیستم‌های پیام‌رسانی سریع و تغییرات بنیادین در زمینه سهولت انتشار محتویات و مطالب، محققین شبکه اجتماعی و محققین تئوری گراف‌ها هم‌اکنون به دنبال استنباط ساختارهای اجتماعی از طریق آنالیز الگوهای پیوندی میان افراد و صفحات وب می‌باشند. هرچند که بررسی ساختارهای اجتماعی انگیزه زیادی برای ایجاد الگوریتم‌های مختلف به وجود آورده است، اما بسیاری از آن‌ها به دلیل هزینه‌های محاسباتی، برای شبکه‌های اجتماعی بزرگ نامناسب هستند. همچنین، علاوه بر شناسایی ساختارهای اجتماعی احتمالی، چگونگی تعریف و توصیف جوامع کشف شده نیز در بسیاری از سناریوهای عملیاتی، اهمیت بسیار زیادی دارد.

واژگان کلیدی: جامعه، شبکه اجتماعی، الگوریتم، گراف

فهرست مطالب

عنوان	صفحه
چکیده	۱
فهرست مطالب	۲
فهرست شکل ها	۵
۱. فصل اول: کلیات	۶
۱.۱. مقدمه	۶
۱.۲. ادبیات تحقیق	۱۰
۱.۳. هدف	۱۲
۲. فصل دوم: تعریف جامعه و روش های کشف آن در شبکه های اجتماعی	۱۳
۲.۱. مقدمه	۱۳
۲.۲. تعریف جامعه	۱۳
۲.۳. آخرین پیشرفت ها در زمینه یابی های تشخیص جامعه	۱۵
۲.۳.۱. زمینه یابی اول، انجام شده توسط فورتوناتو	۱۵
۲.۳.۲. زمینه یابی دوم، انجام شده توسط پورتر	۱۶
۲.۳.۳. زمینه یابی سوم، انجام شده توسط یانگ	۱۷
۲.۳.۴. زمینه یابی چهارم، انجام شده توسط گولهایس و لمان	۱۷
۲.۳.۵. زمینه یابی پنجم، انجام شده توسط پنز	۱۸
۲.۳.۶. زمینه یابی ششم، انجام شده توسط پاپادوپولس	۱۸
۲.۳.۷. هفتمین و آخرین زمینه یابی توسط دنون	۱۹

۳.	فصل سوم: معرفی روش‌های تشخیص	۲۰
۳.۱.	مقدمه	۲۰
۳.۲.	روش‌های تشخیص جامعه	۲۲
۳.۲.۱.	تشخیص جامعه گره-محور	۲۲
۳.۲.۲.	تشخیص جامعه گروه-محور	۳۲
۳.۲.۳.	تشخیص جامعه شبکه-محور	۳۳
۳.۲.۴.	مدل فضای متاخر	۳۷
	ماتریس مدولاریته	۳۹
	ویژگی‌های مدولاریته	۴۰
۳.۳.	شبکه‌های اجتماعی صریح و ضمنی	۴۱
۳.۳.۱.	شبکه‌های اجتماعی صریح	۴۱
۳.۳.۲.	شبکه‌های اجتماعی ضمنی	۴۲
۳.۴.	ارزیابی	۴۲
۳.۴.۱.	بررسی شباهت‌های جفت به جفت	۴۲
۳.۴.۲.	CiteSeer	۴۳
۳.۵.	الگوریتم‌ها	۴۵
۳.۵.۱.	(BFS) breadth-first search	۴۵
	کاربردها	۴۷
۳.۵.۲.	(DFS) Depth-first search	۴۷
	ویژگی‌ها	۴۸
۳.۵.۳.	Random walks	۵۰

۵۱Louvain	۳.۵.۴
۵۱فصل چهارم : جمع بندی و نتیجه گیری	۴
۵۱نتیجه گیری	۴.۱
۵۳فهرست مراجع	۴.۲
۵۵واژگان انگلیسی به فارسی	۴.۳

فهرست شکل ها

صفحه

عنوان

- شکل ۱: نمایش نمونه گراف های باشگاه کاراته زاکاری ۱۲
- شکل ۲: درخت تشخیص جامعه در شبکه های اجتماعی ۲۲
- شکل ۳: گره های تشکیل دهنده یک کلیک ۲۴
- شکل ۴: نمونه دیگری از روش کلیک ۲۵
- شکل ۵: توصیف جوامع k-Clique در $k=4$ ۲۶
- شکل ۱: نمونه k-clique ۲۸
- شکل ۲: درجه در شبکه های جهت دار ۲۹
- شکل ۳: درجه در شبکه های غیر جهت دار ۳۰
- شکل ۴: طرح تجزیه k-core برای یک گراف کوچک ۳۱
- شکل ۵: خوشه بندی سلسله مراتبی ۳۲
- شکل ۶: هم ارزی ساختاری دو گره ۳۵
- شکل ۷: الگوریتم مبنای k-means ۳۶
- شکل ۸: شینگلینگ سریع دو سطحی ۳۷
- شکل ۹: حداقل سازی فاصله میان یک ماتریس فعل و انفعال و یک ساختار بلوکی ۳۸
- شکل ۱۰: حداکثر سازی مدولاریته ۴۱
- شکل ۱۶: توصیفی از چارچوب MapReduce ۴۴
- شکل ۱۱: ترتیب مورد انتظار برای گره ها در الگوریتم breadth-first search ۴۸
- شکل ۱۲: ترتیب مشاهده گره ها ۵۰
- شکل ۱۳: مثالی از درخت هم پوشانی از رئوس ۵۲

۱. فصل اول: کلیات

۱.۱. مقدمه

در ابتدای روند مطالعه شبکه‌های اجتماعی و حتی استخراج جوامع آن‌ها با کمتر از ۱۰۰ کاربر و ۲۰ اجتماع روبه‌رو بودیم. حال آنکه با توجه به انواع گروه‌ها و تعداد افراد مختلف عضو، از یک سو و تعداد برنامه‌های کاربردی وب سایت‌های اجتماعی از سوی دیگر، نه تنها مطالعه شبکه به صورت مشاهده پروفایل‌ها و گروه‌های مشترک در بین اعضا امکان‌پذیر نخواهد بود بلکه حتی شباهت‌های کاربران و برنامه‌ها نیز ناشناخته باقی خواهد ماند. لذا به صورت جدی نیاز به شناخت و تفکیک جوامع داریم.

معنای عمومی جامعه در شبکه‌های اجتماعی، گرد آمدن تعدادی گره^۱ (کاربران، مقالات و ...) می‌باشد به گونه‌ای که اعضای این جامعه بیشترین تعاملات را با یکدیگر داشته باشند و روابط خارج از جامعه آن‌ها کمترین مقدار را به خود اختصاص دهد. پس از تفکیک شبکه به جوامع مختلف، می‌توان با توجه به شباهت‌های گروه‌هایی که به وجود آمده‌اند، به مطالعه آن‌ها پرداخت؛ به بیان دیگر می‌توان از تکنیک تقسیم‌بندی و غلبه برای مطالعه شبکه استفاده نمود. از طرفی پس از شناخت جامعه و بر اساس ویژگی‌ها و دلایلی که افراد در آن‌ها گرد هم آمده‌اند، می‌توان سایر مطالعات و ارائه خدمات با کیفیت بهتری انجام داد.

شبکه‌های اجتماعی بدلیل دسترسی رایانه‌های شخصی، تلفن‌های همراه و دیگر نوآوری‌های سخت افزار اخیر مانند تبلت در سال‌های اخیر بسیار محبوب شده‌اند. این موضوع با محبوبیت رو به رشد بسیاری از شبکه‌های اجتماعی آنلاین مانند توییتر، فیس‌بوک و لینکدین مشهود است. چنین شبکه‌های اجتماعی به انفجار فوق‌العاده‌ای از داده‌های شبکه محور در طیف گسترده‌ای از حالات منجر شده است. شبکه‌های اجتماعی را می‌توان یا در زمینه سیستم‌هایی مانند فیس‌بوک که به صراحت برای تعاملات اجتماعی طراحی شده است، و یا از منظر سایت‌های دیگر مانند فلیکر که برای سرویس‌های مختلف مانند اشتراک‌گذاری محتوا طراحی شده اما به طور گسترده‌ای اجازه تعاملات اجتماعی را می‌دهد، تعریف کرد.

ظهور شبکه‌های اجتماعی آنلاین به یکی از هیجان‌انگیزترین وقایع در این دهه تبدیل شده است. بسیاری از شبکه‌های اجتماعی آنلاین مثل توییتر، لینکدین، و فیس‌بوک به طور فزاینده‌ای محبوب شده‌اند. علاوه بر این، تعدادی از شبکه‌های چند رسانه‌ای مانند فلیکر نیز افزایش سطح محبوبیت در سال‌های اخیر را بخود دیده است. بسیاری از این شبکه‌های اجتماعی از لحاظ محتوا بسیار غنی می‌باشند، و آنها به طور معمول شامل مقدار بسیار زیادی از محتوا و داده‌ی ارتباطی هستند که می‌تواند برای تجزیه و تحلیل قوی‌تر باشد. اطلاعات ارتباطی اساساً ساختار گراف از شبکه‌های اجتماعی و ارتباطات بین اشخاص می‌باشند. در حالی که داده‌های محتوا شامل متن، تصاویر و دیگر داده‌های چند رسانه‌ای در شبکه می‌باشند. غنای این شبکه فرصت بی‌سابقه‌ای برای تجزیه و تحلیل داده‌ها در زمینه شبکه‌های اجتماعی فراهم می‌کند.

مجموعه داده‌های نشات گرفته در بسیاری از حوزه‌های مختلف جهان واقعی را می‌توان در قالب شبکه‌های تعامل در یک حالت بسیار طبیعی، موجز و معنی‌دار نشان داد. این امر به ویژه در زمینه‌های اجتماعی واقعی، با توجه به پیشرفت‌های اخیر

¹ Node

در فناوری اینترنت و برنامه های کاربردی وب منجر به طیف متنوعی از شبکه های اجتماعی تکاملی شده است. تجزیه و تحلیل چنین شبکه هایی می تواند منجر به کشف الگوهای مهم شود و به طور بالقوه بر خواص مهم حاکم بر رشد چنین شبکه هایی سایه بيفکند.

نشان داده شده است که بسیاری از این شبکه ها طبیعت و یا ساختار جامعه قوی پیمانه ای را نشان میدهند. دستور کار مهم پژوهش این است که جوامع مورد علاقه را شناسایی و رفتار آنها را در طول زمان مطالعه می کنند. با توجه به اهمیت این مسئله فعالیت های قابل توجهی در این زمینه به ویژه در چند سال گذشته انجام شده است.

اخیرا مشاهده شده است در حالی که شبکه های اجتماعی در گروهی از عرصه های گوناگون بوجود می آیند اما آنها اغلب مفاهیم و یا تم های مهم مشترکی دارند. مطالعه ی چنین شبکه های رابطه ای پیچیده، به تازگی به عنوان علم شبکه مورد اشاره قرار گرفته است، و می تواند بینشی در ساختار، خواص و رفتار ضروری خود ارائه کند.

به طور کلی، یک شبکه اجتماعی به عنوان شبکه ای از تعاملات و یا روابط، که در آن گره ها از بازیگران (نقش آفرینان) تشکیل شده و لبه ها شامل روابط یا تعاملات بین این بازیگران می باشند، تعریف می شود. یک کلیت از ایده شبکه های اجتماعی، شبکه های اطلاعاتی است، که در آن گره می تواند شامل بازیگران و یا اشخاص باشد، و لبه روابط بین آنها را شامل می شود. واضح است که مفهوم شبکه های اجتماعی به صورت خاص از یک شبکه اجتماعی مبتنی بر اینترنت مانند فیس بوک محدود نمی شود؛ مسئله شبکه های اجتماعی اغلب در زمینه جامعه شناسی از نظر فعل و انفعالات عمومی بین هر گروه از بازیگران مورد مطالعه قرار گرفته است. این تعاملات ممکن است در هر شکل متعارف و یا غیر متعارف، چهره به چهره، تعاملات مخابراتی، پست الکترونیک و یا فعل و انفعالات متقابل نامه پستی باشد.

به طور کلی مطالعات چندانی در زمینه تحلیلی شبکه اجتماعی بر تعاملات آنلاین متمرکز نشده است، و به طور تاریخی از آغاز و محبوبیت کامپیوترها یا اینترنت پیشی گرفته اند. یک نمونه مثال کلاسیک در این زمینه، مطالعه میلگرام² در دهه ۶۰ است، کسی که احتمال اینکه هر جفت از بازیگران در سیاره حداکثر ۶ درجه از هم جدا می باشند، را فرض کرد. در حالی که چنین فرضی در طول دهه های اخیر به صورت حدس و گمان باقی مانده است، پیشرفت شبکه های اجتماعی آنلاین این امکان را فراهم کرده است که این چنین فرض هایی را حداقل در شبکه های آنلاین مورد آزمون قرار دهیم. این همچنین بعنوان پدیده دنیای کوچک اشاره می شود. این پدیده در زمینه داده های پیام MSN مورد آزمون قرار گرفته و نشان میدهد که میانگین طول مسیر بین دو کاربر MSN ۶.۶ می باشد. این امر می تواند تاییدی بر قانون "۶ درجه جدایی" در شبکه های اجتماعی باشد. چنین مثال هایی به هیچ عنوان منحصر به فرد نمی باشند، دامنه وسیعی از داده های آنلاین در دسترس می باشند که برای تایید صحت، دسته ای دیگر از فرضیات همانند تحلیل، روی قطر یا طول، یا پیوستگی ترجیحی مورد استفاده قرار گرفته اند. به طور کلی، "دسترسی به مقدار زیادی از داده ها در محیط های آنلاین یک انگیزه جدید در مورد مطالعه علمی و آماری شبکه های اجتماعی ایجاد کرده است".

این جنبش داده محور به ایجاد حجم قابل توجهی از تحقیقات منجر شده است، که اساساً مبتنی بر فعل و انفعالات انسانی است. در بسیاری از موارد، بینش های زیربنایی قابل اجرا در شبکه های اجتماعی معمولی نیز هستند. قبل از بحث در مورد موضوع

² Milgram

پژوهش به طور تفصیلی، ما به طور خلاصه شرایط مختلف برای تجزیه و تحلیل شبکه های اجتماعی را برمی شماریم، و به طور خاص بین حالات متعارف و غیر متعارف تمایز ایجاد می کنیم. به طور اخص، این محیط های متفاوت به صورت زیر می باشند :

- کلاسیک ترین تعریف از یک شبکه اجتماعی، تعریفی است که صرفاً بر تعاملات انسانی مبتنی است. این مطالعه کلاسیک از شبکه های اجتماعی در زمینه جامعه شناسی است. این مطالعات به طور سنتی با روش پر زحمت و دشوار برای اندازه گیری تعاملات بین اشخاص با جمع آوری داده های واقعی در مورد تعامل انسان به طور دستی انجام شده است. مثال این موضوع "آزمایش شش درجه جدایی میلگرام است، که برای شرکت کنندگان به منظور تست اینکه آیا دو بازیگر دلخواه را می توان با زنجیره ای از ۶ لبه با استفاده از ارسال ایمیل به طور محلی بهم مرتبط کرد یا خیر" می باشد. چنین آزمایشی برای انجام در یک راه کاملاً رضایت بخش سخت است، چرا که بازیگران در چنین آزمایشی ممکن است میزان پاسخی داشته باشند که نمی توان از نظر رفتار تعاملات اجتماعی مدل سازی کرد. به عنوان مثال آزمایش میلگرام، که در آن نتایج اغلب به علت میزان کم ارسال ها که هرگز به هدف نرسیده اند، مورد سوال و تردید باشد. علاوه بر این، چنین تجربه های اجتماعی اغلب به سمت اهداف وضعیت بالا به منظور حصول اطمینان از احتمال ارسال های منطقی مغرضانه می باشند. با این حال، این نتایج در نهایت حداقل از دیدگاه کیفی پذیرفته شده و مقبول است، حتی اگر قانون شش درجه، بسته به ماهیت شبکه ی تحت مطالعه، دقیقاً درست نباشد. با این وجود، قطعاً به نظر می رسد "پدیده جهان کوچک" درست است، زیرا قطر اغلب چنین شبکه هایی نسبتاً کوچک هستند.

تحلیل اجتماعی چنین شبکه هایی همچنین در زمینه علوم شناختی مدل سازی شده اند، جایی که جنبه های شناختی چنین تعاملاتی برای اهداف تحلیل مورد استفاده قرار می گیرند. بیشتر تحقیقات در زمینه سنتی شبکه های اجتماعی از این منظر اجرا شده اند. تعدادی از کتاب های منتشره^۳ و^۴ درکی از چنین دیدگاهی را فراهم می کنند. به هر حال، این کار موضوعات داده محور را که در شبکه های اجتماعی آنلاین و متصل به اینترنت رایج می باشد را بحث نمی کند.

- تعدادی از مقدر سازهای تکنولوژی همانند ارتباطات راه دور، ایمیل الکترونیک، و پیام دهنده های^۵ چت الکترونیک (همانند اسکایپ^۶، گوگل تاک، MSN)، می توانند بعنوان شکل غیر مستقیمی از شبکه های اجتماعی مورد توجه قرار گیرند، زیرا آنها ماهیتاً بعنوان ارتباطات بین کاربران مختلف طراحی شده اند. یک مزیت چنین برنامه هایی این است که آثاری از داده ارتباطات اغلب در دسترس می باشد. این داده می تواند برای تجزیه و تحلیل بسط یافته از چنین شبکه های اجتماعی استفاده شود. برخی مثال ها تجزیه و تحلیل بسیط در مجموعه داده های ایمیل ENRON، یا ارزیابی اخیر ۶ درجه جدایی در محتوای داده های MSN می باشند.

³ D. J. Watts. Six Degrees: The Science of a Connected Age, W. W. Norton and Company, 2004

⁴ S. Wasserman, K. Faust. *Social Network Analysis: Methods and Applications*

⁵ Messenger

⁶ Skype

• در سال های اخیر، تعدادی از سایت ها به طور صریح به منظور مدل سازی تعامل بین بازیگران مختلف بوجود آمده اند. برخی از نمونه های چنین شبکه های اجتماعی فیس بوک^۷، مای اسپیس^۸ و یا لینکداین^۹ می باشند. علاوه بر این، سایت هایی که برای به اشتراک گذاری محتوای رسانه های آنلاین، مانند فلیکر^{۱۰}، یوتیوب یا دلشس^{۱۱} استفاده می شود، همچنین می توانند اشکال غیر مستقیم از شبکه های اجتماعی در نظر گرفته شوند، چرا که آنها اجازه ی سطح گسترده ای از تعاملات با کاربر را می دهند. در این موارد، تعامل در حوالی یک سرویس خاص مانند به اشتراک گذاری محتوا متمرکز می باشد، که هنوز بسیاری از اصول اساسی شبکه های اجتماعی اعمال می شود. توجه داشته باشید که چنین شبکه های اجتماعی ای بسیار غنی می باشند، زیرا آنها حاوی مقدار بسیار زیادی از محتوا مانند متن، عکس، صوتی و یا ویدئویی می باشند. چنین محتوایی می تواند برای طیف گسترده ای از اهداف قوی تر باشد. به طور خاص، تعامل بین لینک ها و محتوا انگیزه ای برای طیف گسترده ای از برنامه های کاربردی معدن (کاوش) فراهم کرده است. علاوه بر این، رسانه های اجتماعی تعدادی از راه های منحصر به فرد برای کاربران برای تعامل با یک دیگر، مانند ارسال وبلاگ ها، یا برچسب زدن تصاویر به یکدیگر ارائه کرده است. که این نوع از تعامل غیر مستقیم می باشد، آنها دانش محتوا محور غنی فراهم می کنند که می تواند برای مقاصد کاوش مورد استفاده قرار گیرد. در سال های اخیر، حتی این امر ممکن شده است که محتوای مبتنی بر حسگر زمان واقعی را با شبکه های اجتماعی پویا ادغام کرد. این به دلیل توسعه حسگرها، شتاب سنج^{۱۲}، دستگاه های تلفن همراه و سایر دستگاه های GPS می باشد، که می تواند در یک محیط اجتماعی برای ارائه یک تجربه پویا و تعاملی استفاده شود.

• در نهایت، تعدادی از شبکه های اجتماعی نیز می توانند از انواع خاصی از تعاملات در جوامع مختلف ساخته شوند. یک مثال کلاسیک، جامعه علمی است که در آن شبکه های کتابشناختی می تواند از هر دو شرکت مؤلف یا داده استناد ساخته شده باشد. این می تواند در رابطه با محتوای نشریات به منظور استخراج روند و الگوهای جالب در مورد مقالات زمینه ای استفاده شود. توجه داشته باشید که بیشتر تجزیه و تحلیل برای مورد اول بکار می رود، گرچه بسیاری از اطلاعات و مطالب در دسترس است، به خاطر روشی که در آن چنین شبکه های اسناد بایگانی شده است. تعدادی مجموعه های سند و شبکه های کتابشناختی به صراحت آرشیو می شوند، و می توان به دلیل محتوایی که همراه با چنین شبکه هایی در دسترس می باشد آنها را در رابطه با تکنیک های داده محور اصولی استفاده کرد.

در حالی که نتایج حاصل از این تحقیق ممکن است برای همه انواع مختلف شبکه های اجتماعی بکار رود، تمرکز خاص بر مسائل داده محور است که در زمینه شبکه های اجتماعی آنلاین بوجود می آیند. همچنین درک این مطلب اهمیت زیادی دارد که

⁷ Facebook

⁸ My Space

⁹ LinkedIn

¹⁰ Flickr

¹¹ Delicious

¹² Accelerometer

شبکه اجتماعی آنلاین می تواند به طور کلی تری از یک سایت آنلاین مانند فیس بوک، توییتر^{۱۳} و یا لینکداین که به طور رسمی به عنوان سایت های شبکه های اجتماعی تبلیغ می شوند، تعریف شود. در واقع، هر وب سایت و یا نرم افزار که یک تجربه ی اجتماعی به صورت تعاملات کاربر فراهم می کند، می تواند به عنوان یک شکل از شبکه های اجتماعی در نظر گرفته شود. به عنوان مثال، سایت های رسانه های به اشتراک گذاری مانند فلیکر، یوتیوب^{۱۴}، و یا دلشس به طور رسمی شبکه های اجتماعی در نظر گرفته نمی شوند، در عین حال آنها امکان تعاملات اجتماعی در زمینه تبادل اطلاعات در مورد محتوای به اشتراک گذاشته شده را فراهم می آورند. به طور مشابه، بسیاری از برنامه های کاربردی چت، موبایل، وب و اینترنت نیز جنبه های اجتماعی تعبیه شده را دارند. علاوه بر این، بسیاری از برنامه های کاربردی موبایل مانند گوگل لیتیود^{۱۵} امکان تعبیه ضمنی اطلاعات سنسور یا GPS را می دهد، و این به منظور فعال کردن تعاملات کاربر استفاده می شود. این ها همه اشکال جدیدی از شبکه های اجتماعی می باشند، که هر کدام از آنها مجموعه ای از چالش های منحصر به فرد برای تجزیه و تحلیل به ارمغان می آورند. بنابراین، تعریف ما از شبکه های اجتماعی نسبتاً گسترده است، و بسیاری از مطالب، جنبه هایی که مربوط به این اشکال جایگزین شبکه های اجتماعی هستند، را مطالعه خواهند کرد.

۱.۲. ادبیات تحقیق

تجزیه و تحلیل جوامع اجتماعی تمرکز بسیاری از مطالعات در طول هشتاد سال گذشته بوده است. یکی از اولین مطالعات در این زمینه عبارتند از کارهای رایس^{۱۶} در "تجزیه و تحلیل جوامعی از افراد بر اساس تعصبات سیاسی و الگوهای رای گیری". یک مطالعه جدیدتر در امتداد خطوط مشابه اما با تمرکز بر ساختار شبکه وبلاگ سیاسی توسط آدامیک و گلانس^{۱۷} مورد بحث قرار گرفت. اومانز^{۱۸} اولین فردی بود که نشان داد گروه های اجتماعی می توانند با مرتب سازی مجدد ردیف ها و ستون های ماتریس توصیف کننده روابط اجتماعی، تا زمانی که آنها یک شکل تقریبی بلوک-قطر بگیرند، دوباره ارزیابی شوند. در واقع این ایده هنوز هم به عنوان ابزار اساسی برای تجسم ساختار اجتماعی و ساختار خوشه بندی کلی تر عمل می کند. ویس و یاکوبسون^{۱۹} کار گروه ها را در یک سازمان دولتی مورد بررسی قرار دادند. موضوع اصلی کار آنها شناسایی گره های رابط و استفاده از چنین گره هایی برای جدا کردن ساختار جامعه بود. در واقع این کار می تواند به عنوان یک نسخه اولیه از مفهوم مرکزیت بینابینی که توسط نیومن^{۲۰} معروف شد، تصور شود. مطالعه کلوب کاراته^{۲۱} که توسط زاکاری^{۲۲} مطالعه می شود، گراف

¹³ Twitter

¹⁴ Youtube

¹⁵ Google Latitude

¹⁶ Rice

¹⁷ Adamic & Glance

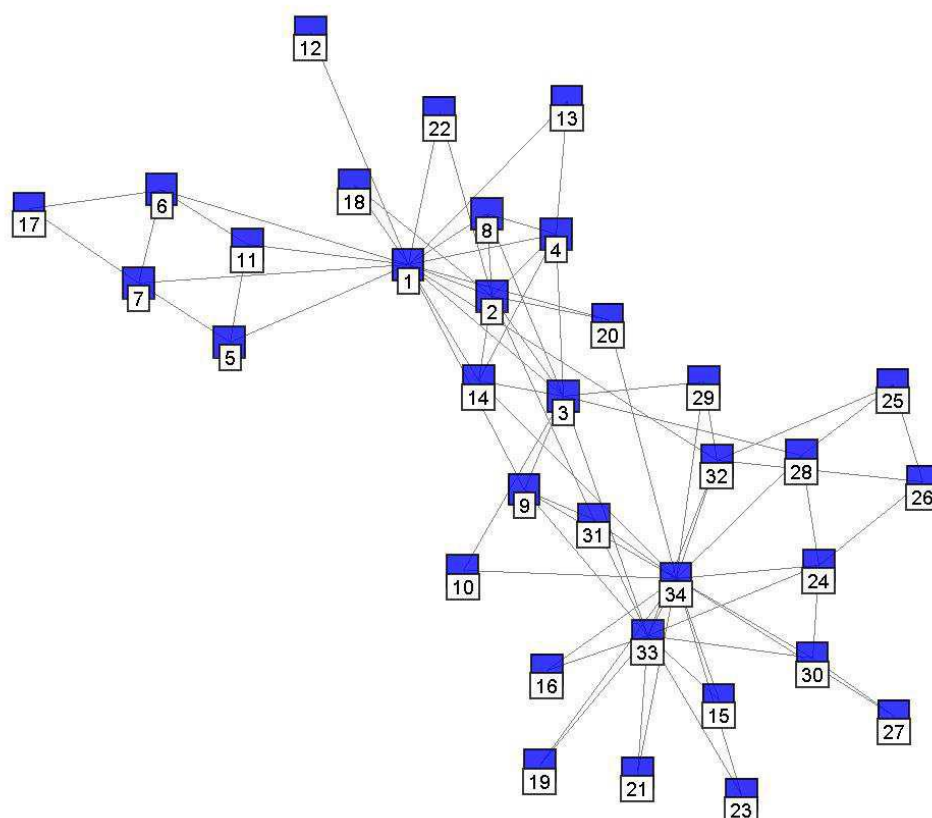
¹⁸ Homans

¹⁹ Weiss & Jacobson

²⁰ Newmann

²¹ Karate club

معروفی است که معمولاً بعنوان ارزیابی مقایسه‌ای برای آزمون الگوریتم‌های تشخیص جامعه استفاده می‌شود (شکل ۱). این شامل ۳۴ راس، اعضای باشگاه در آمریکا، که در طول ۳ سال مورد مشاهده قرار گرفتند می‌باشد و شامل یک نمونه انشقاقی جامعه و بنابراین موضوع بسیاری از مطالعات می‌باشد. مطالعه دیگری که توسط بک و آتالای^{۲۳} انجام گرفت، این بود که شبکه اجتماعی وام‌ها در میان موسسات مالی را تحلیل کردند تا درکی از چگونگی تعاملات بین جوامع چندگانه‌ی تأثیر گذار بر سیستم بعنوان کل بدست آورند. بیشتر این مطالعات بر درک ساده ساختار زیربنایی جامعه و تکامل آن می‌باشد.



شکل ۱: نمایش نمونه گراف‌های باشگاه کاراته زاکاری

²² Zachary

²³ Bech & Atalay

علاوه بر مطالعه شبکه های اجتماعی بشر، جانور شناسان و زیست شناسان همچنین مطالعه رفتار اجتماعی سایر حیوانات و مخلوقات دریایی را شروع کردند. لوسو^{۲۴} در یک مطالعه برجسته رفتار شصت و دو دلفین اقیانوس نیوزیلند را آزمایش کرد. این مطالعه به رفتار اجتماعی شصت و دو دلفین نگاهی انداخت و لبه ها بین حیوانات متصل بهم بیشتر دیده شد تا به صورت یک شانس تصادفی. لوسو اشاره می کند که ساختار واضح و اختصاصی گراف منتج پیشنهاد می کند که رفتار اجتماعی چنین پستاندارانی اغلب نسبتاً برجسته است. اخیراً یک تیم بین رشته ای شامل حیوانشناسان و دانشمندان کامپیوتری رفتار اجتماعی در گورخرها را مطالعه کرده اند. بینش ها اعم از شناسایی رهبران تاثیرگذار گله در جوامع، و رفتار تکاملی شبکه های اجتماعی ناشی از گورخرها به درک قابل توجه فزاینده ای از چگونگی ارتباط و اجتماعی شدن این حیوانات برای بقا، منجر شده است.

۱.۳. هدف

پیش از هر چیز باید بدانیم که هدف از تشخیص جامعه در شبکه های اجتماعی چه خواهد بود. در ابتدا، به منظور درک فعل و انفعالات اجتماعی، بسیار ضروری است که این ساختارهای اجتماعی یا «جوامع» را شناسایی کنیم. ساختار جوامع معمولاً نشان دهنده ویژگیهای جذاب مشترک در میان اعضا، مانند سرگرمی های مشترک، تفریحات، فعالیت های اجتماعی یا رتبه (جایگاه) می باشد. اهمیت ساختار جوامع از یک دیدگاه دقیق تر، فراتر از شبکه های اجتماعی خواهد بود: برای مثال، در اسناد هایپرلینک شده ای مانند شبکه جهانی وب، ویژگی برجسته «افراد» (که در اینجا صفحات وب هستند) شامل موضوعات مرتبط و دیدگاه های مشترک است. به این ترتیب، تحقیقات بسیاری برای شناسایی جوامع در گراف وب و شرایط مرتبط به انجام رسیده است.

شاید یکی از جالب توجه ترین ویژگی های ساختار جوامع، تکامل آن ها در طول زمان باشد: اعضا و فعل و انفعالات چگونه تغییر می کنند، و چگونه این تغییر با ویژگی های افراد و پیشآمدها در چارچوب زمانی همبسته می شود؟ آیا سرعت گردش گروه ها آهسته است یا سریع؟ آیا آن ها در طول یک زمان طولانی همبسته و تفکیک میشوند یا به شکل خودبه خودی؟ با پیشرفت تکنولوژی، جمع آوری داده های طولی به طور پیوسته امکان پذیر شده است و به ما امکان پاسخ گویی سیستماتیک به این پرسش ها را می دهد. با این حال، این وظیفه نیازمند الگوریتم هایی است که نه تنها جوامع را شناسایی کنند بلکه همچنین تغییرات آن ها را در طول زمان بسنجند.

^{۲۴} Lusseau

۲. فصل دوم: تعریف جامعه و روش های کشف آن در شبکه های اجتماعی

۲.۱. مقدمه

همان گونه که در بخش قبل اشاره شد، معنای عمومی جامعه در شبکه های اجتماعی گرد آمدن تعدادی از گره ها می باشد به گونه ای که اعضای این جامعه بیشترین تعاملات ها را با یکدیگر داشته باشند و روابط خارج از جامعه آن ها، حداقل مقادیر را به خود اختصاص دهند. در مقالات، مروری بر دسته بندی های روش های استخراج جامعه ارائه شده است.

آنچه که به عنوان مسئله اصلی در انتخاب یا ایجاد یک روش مطرح می باشد، مشخص نمودن پارامترها و ویژگی های شبکه و جوامع آن است. یک محقق و یا توسعه دهنده باید به خوبی از شبکه (هدف از استخراج و مطالعه شبکه و در نظر گرفتن کاربران و روابط مربوط به آن ها) آگاه باشد و بهترین مدل را انتخاب نماید. مثلاً چنانچه نوع رابطه دانش آموزان یک کلاس مدنظر باشد، کافی است از تعدادی از دانش آموزان به استخراج کلاس و روابط نزدیک هم شاگردی ها بپردازیم؛ در نتیجه یک جستجوی کلی می تواند ایده آل ترین مدل به حساب آید، حال تمام شبکه های اجتماعی آنلاینی را در نظر بگیرید که بیش از ده ها هزار کاربر دارند و این مسئله باعث می شود که زمان و پردازش های کنونی استفاده از تکنیک های انتخاب تصادفی و نمونه برداری غیرقابل اجتناب شوند. موارد قابل تفکیک در انتخاب یا ایجاد روش های استخراج و اکتشاف جامعه در شبکه های اجتماعی باید بر مبنای پارامترهای زیر صورت پذیرد: جهت دار بودن یا دوطرفه بودن روابط؛ وزن دار بودن روابط؛ درزگیری انواع مختلف کاربران و روابط یا تک نوع بودن هر یک، شبکه های مقیاس پذیر و یا تعداد محدودی کاربر و انتخاب یکی از مدل ها بر مبنای خوشه بندی و یا ساختارهای لینکی در میان انواع روش ها.

۲.۲. تعریف جامعه

تعریف جامعه یک وظیفه نسبتاً چالش برانگیز و دشوار می باشد. تعاریف بسته به محققین و بسته به الگوریتم ها متفاوت می باشند. رایج ترین تعریف مورد استفاده، تعریف یانگ می باشد "یک جامعه بعنوان گروهی از گره های شبکه که در آن لینک های متصل کننده گره ها متراکم هستند اما بین آنها پراکندگی وجود دارد" این تعریف برای گراف ها قابل کاربرد است و می تواند به گراف های دو قسمتی بسط یابد.

فورتوناتو^{۲۵} سه سطح برای تعریف یک جامعه را شناسایی می کند: تعاریف محلی، تعاریف جهانی، و تعریف مبتنی بر شباهت راس. در گروه تعریف محلی، تعریف از جوامع متشکل است از: "بخش هایی از گراف با چند ارتباط کوچک با بقیه سیستم." در این تقسیم بندی، جوامع از منظر ساختار داخلی شان مستقل از مابقی بخش های نمودار مورد مطالعه قرار می گیرند. در گروه تعریف عمومی، یک معیار عمومی در ارتباط با نمودار برای محاسبه جوامع استفاده می شود. این معیار عمومی وابسته به

²⁵ Fortunato

الگوریتمی است که برای قراردادی جوامع بکار می رود. چه معیار خوشه بندی و چه یک معیار مبتنی بر فاصله معرفی شود ؛ در اغلب موارد، معیاری که به طور رایج استفاده می شود، نشان می دهد که گراف شامل یک ساختار جامعه متفاوت از یک گراف تصادفی است. در گروه تعریف جامعه مبتنی بر شباهت رأس، جوامع به عنوان گروه هایی از رؤس مشابه بهم در نظر گرفته می شوند.

فورتوناتو بیشتر جوامع را طوری تعریف می کند که، در آنها خوشه^{۲۶} و یا ماژول^{۲۷} ها، به عنوان " گروهی از رؤس که احتمالاً ویژگی های مشترکی را به اشتراک می گذارند و یا نقش مشابهی در گراف بازی می کنند" می باشند. وی تعریف را برحسب الگوریتم بکار گرفته شده تعیین می کند، و منجر به شناسایی حداقل ۸ تعریف متفاوت می شود :

- ❖ گروه^{۲۸}: زیرگروه هایی که اعضایش با یکدیگر دوست می باشند.
- ❖ N- گروه^{۲۹} با دو متغیر: زیرگراف بیشینه به طوری که فاصله ی هر جفت از راسش بیشتر از n نمی باشد.
- ❖ K-plex: زیرگراف بیشینه ای که هر راس به تمام رؤس زیرگراف به جز k امین آنها نزدیک است.
- ❖ LS-Set: زیرگرافی که درجه داخلی بیشتر از درجات خارجی می باشد.
- ❖ Lambda set: زیرگرافی که هر جفت از رؤس دارای یک اتصال لبه بزرگتر از هر جفتی که با یک راس از زیرگراف و یک راس خارج از زیرگراف شکل گرفته است، می باشد.
- ❖ جوامعی که مبتنی بر معیار برازندگی و یا معیار کیفیت می باشند.
- ❖ جوامعی که توسط میانگین الگوریتم های مبتنی بر پیمانه ای بودن مشخص می شوند.
- ❖ خوشه ها: جوامعی که با استفاده از روش های خوشه بندی معروف استخراج می شوند.

پورتر^{۳۰} ریشه های جامعه ی مورد مطالعه در زمینه های جامعه شناسی و انسان شناسی را یادآوری می کند. وی جوامع را بعنوان " گروه های منسجمی از گره ها که به یکدیگر متراکم تر از گره های جوامع دیگر متصل هستند" تعریف می کند . تفاوت در روش های برجسته در ارزیابی او متکی بر تعریف عبارت " متراکم تر " است، که با پنج نوع از الگوریتم ها، یعنی : تکنیکهای خوشه بندی ، الگوریتم تابع کیفیت، الگوریتم تشخیص جامعه مبتنی بر محوریت و سایر موارد مشابه، الگوریتم های دسته نفوذ و در نهایت الگوریتم های بهینه سازی پیمانه شناسایی می شوند.

²⁶ Cluster

²⁷ Module

²⁸ Clique

²⁹ n-clique

³⁰ Porter

گولباس و لمان^{۳۱} جامعه را بعنوان " زیرمجموعه متراکمی از گره ها که تنها به طور پراکنده به مابقی شبکه متصل هستند" تعریف می کنند.

پاپادوپولس^{۳۲} جوامع را بعنوان " گروه هایی از رئوس که به یکدیگر نسبت به مابقی شبکه متراکم تر متصل هستند" تعریف می کند.

می توان مشاهده کرد که تمامی تعاریف کاملاً مشابه هستند با این حال هنوز هم ممکن است در تعریف ریاضی رسمی مرتبط با آنها متفاوت باشند. جوامع نیز ممکن است از دیدگاه های مختلف در نظر گرفته شوند. رویکرد اولیه گراف زیربنایی را تقسیم می کند، یعنی با تقسیم گراف یا ساختار شبکه به جوامع متمایز با استفاده از الگوریتم های بهینه. سپس روش دوم جوامع همپوش را تشخیص می دهد و به دنبال بهترین آرایش جامعه می باشد.

روت^{۳۳} یک نوع جدید از جامعه را تعریف می کند: " جوامع بینشی یا شناختی"^{۳۴}، که "جوامع یا گروهی از نهادهای بینشی هستند که نظرات دانشی مشترک را به اشتراک می گذارند"، بعنوان مثال گروهی از محققان که یک موضوع دقیق تکی را بررسی می کنند. این نوع جدید از مفهوم جامعه نیاز به انواع جدید ساختارها برای ادامه دادن با توصیفاتشان دارند. روت استفاده از شبکه گالیوس^{۳۵} را انتخاب کرد.

۲.۳. آخرین پیشرفت ها در زمینه یابی های تشخیص جامعه

بیشتر زمینه یابی ها، مقالات تحقیقی و روش ها را بر حسب الگوریتم تشخیص جامعه طبقه بندی می کنند.

۲.۳.۱. زمینه یابی اول، انجام شده توسط فورتوناتو

اولین مورد از این هفت مطالعه اصلی توسط فورتوناتو^{۳۶} نسبت به بسیاری از روش های تشخیص جامعه، جامع می باشد و مبتنی بر یک نمایش گرافیکی می باشد. این زمینه یابی مروری موثر از این زمینه را فراهم می کند و به توصیف پایه روش تشخیص جامعه، اتخاذ یک دیدگاه فیزیک آماری و به طور خاص بر تکنیک های طراحی شده توسط فیزیکدانان آماری تمرکز می کند. بحث او همچنین شامل مسائل مهمی مانند : اهمیت خوشه، روش هایی که باید با آنها تست و مقایسه شود و کاربرد در شبکه واقعی است. روش ها به هشت خانواده طبقه بندی می شوند ، یعنی :

³¹ N. Gulbahce and S. Lehmann

³² Papadopoulos

³³ Roth

³⁴ Epistemic communities

³⁵ Galios

³⁶ S. Fortunato

- روش های سنتی مبتنی بر خوشه بندی مانند k-means و سایر کاربردها
- الگوریتم های تقسیم کننده^{۳۷} عمدتاً مبتنی بر خوشه بندی سلسله مراتبی
- الگوریتم های مبتنی بر پیمانه ای بودن^{۳۸} و سایر الگوریتم های مشابه
- الگوریتم های طیفی^{۳۹}
- الگوریتم های پویا^{۴۰} و سایر الگوریتم های مشابه
- روش های آماری مبتنی بر استنباط
- روش های چند وضوحی^{۴۱}
- روش هایی برای یافتن جوامع همپوش^{۴۲} و سایر روش های متفرقه

۲.۳.۲. زمینه یابی دوم، انجام شده توسط پورتر

زمینه یابی دوم، که توسط پورتر انجام شده فقط شامل رویکردهای تقسیم بندی گراف می باشد و بینشی به تکنیک های گرافیکی از طریق اشاره به اولین زمینه یابی ارائه میدهد. مجموعه بسیطی از تکنیک ها برجسته می شوند، به طوری که برخی از موضوعات حل نشده ی مهم باقی می مانند. نمونه های کاربردی در برخی شبکه های اجتماعی بزرگتر بعلاوه ی گروه بندی تشخیص جامعه به ۵ تکنیک اصلی تفکیک می شوند :

³⁷ Divisive

³⁸ Modularity

³⁹ Spectral

⁴⁰ Dynamic

⁴¹ Multi-Resolution

⁴² Overlapping

- تکنیک های مبتنی بر محوریت پیرامون الگوریتم نیومن
- روش های محلی پیرامون روش تراوش k-clique
- روش های بهینه سازی پیمانه ای پیرامون الگوریتم نیومن
- روش های تقسیم بندی طیفی پیرامون الگوریتم سیمون^{۴۳}
- روش های مبتنی بر فیزیک که توسط قانون پاتس^{۴۴} برانگیخته شد

۲.۳.۳. زمینه یابی سوم، انجام شده توسط یانگ

زمینه یابی سوم توسط یانگ، نسبت به تمام تکنیک های مبتنی بر نمایش گرافیکی جامع است و یک مرور خوبی از طریق طبقه بندی تمام تکنیک ها در ساختار درختی، بر حسب ۳ طبقه ایجاد می کند :

- الگوریتم های مبتنی بر بهینه سازی
- الگوریتم های اکتشافی
- الگوریتم های مبتنی بر شباهت و روش های پیوندی

۲.۳.۴. زمینه یابی چهارم، انجام شده توسط گولهایس و لمان

چهارمین زمینه یابی از گولهایس و لمان یک زمینه یابی جزئی است که روش های تشخیص جامعه سلسله مراتبی را تجزیه و تحلیل می کند و تعدادی راهنمایی برای رویکردهای آتی فراهم می کند.

⁴³ Simon

⁴⁴ Potts

۲.۳.۵. زمینه یابی پنجم، انجام شده توسط پنز

زمینه یابی پنجم توسط پنز^{۴۵} روش های متعدد تشخیصی را ترکیب می کند و آنها را به ۵ دسته طبقه بندی می کند :

- رویکردهای کلاسیک شامل تقسیم بندی کلاسیک گراف بعنوان مثال، دوبخشی طیفی از کرنیگان و لین^{۴۶}، خوشه بندی و خوشه بندی سلسله مراتبی
- رویکردهای مجزایی که تلاش در تقسیم بندی گراف به جوامع مختلف با حذف لبه های متصل کننده جوامع مجزا دارند. در این گروه، پنز الگوریتم معروف گیروان-نیومن^{۴۷} و سایر رویکردهای مختلف را قرار میدهد.
- رویکردهای تراکمی نسبتا مشابه به همتای سلسله مراتبی و شامل یک روش مبتنی بر پیمانه ای بودن بهینه توسط نیومن و سایر الگوریتم ها.
- الگوریتم های نوع رندوم واک^{۴۸} و سایر روش های مبتنی بر زمان متوسط که نیاز به دستیابی به راس دارند.
- و نهایتا، گروه وسیعی از رویکردهای متفرقه

زمینه یابی پنز تنها رویکردهای تقسیم بندی گراف را آزمایش کرده است.

۲.۳.۶. زمینه یابی ششم، انجام شده توسط پاپادوپولس

زمینه یابی ششم توسط پاپادوپولس تکنیک های تشخیص جامعه را به ۵ دسته طبقه بندی می کند :

⁴⁵ Pons

⁴⁶ Kernighan & Lin

⁴⁷ Girvan-Newman

⁴⁸ Random walk

- کشف زیرگراف منسجم
- خوشه بندی راس
- بهینه سازی کیفیت جامعه
- تقسیم کننده
- روش های مبتنی بر مدل

۲.۳.۷. هفتمین و آخرین زمینه یابی توسط دنون

هفتمین و آخرین زمینه یابی توسط دنون^{۴۹} انجام شده است و عمدتاً بر عملکرد هر نوع از الگوریتم ها متمرکز می باشد. در میان این ۷ زمینه یابی، تنها یکی به جوامع همپوش اشاره کرده است در حالی که هیچ کدام استفاده از ساختارهای گراف ترکیبی را اشاره نکرده اند. یک زمینه دیگر که توسط این زمینه یابی ها برجسته شده است، ساختارهای گالیوس می باشد. همانطور که در ذیل بحث می شود، گالیوس ساختارهای پیچیده تری از گراف های معمول می باشد با این حال آنها معناهای بیشتری برای ساختارهای شبکه فراهم می کنند.

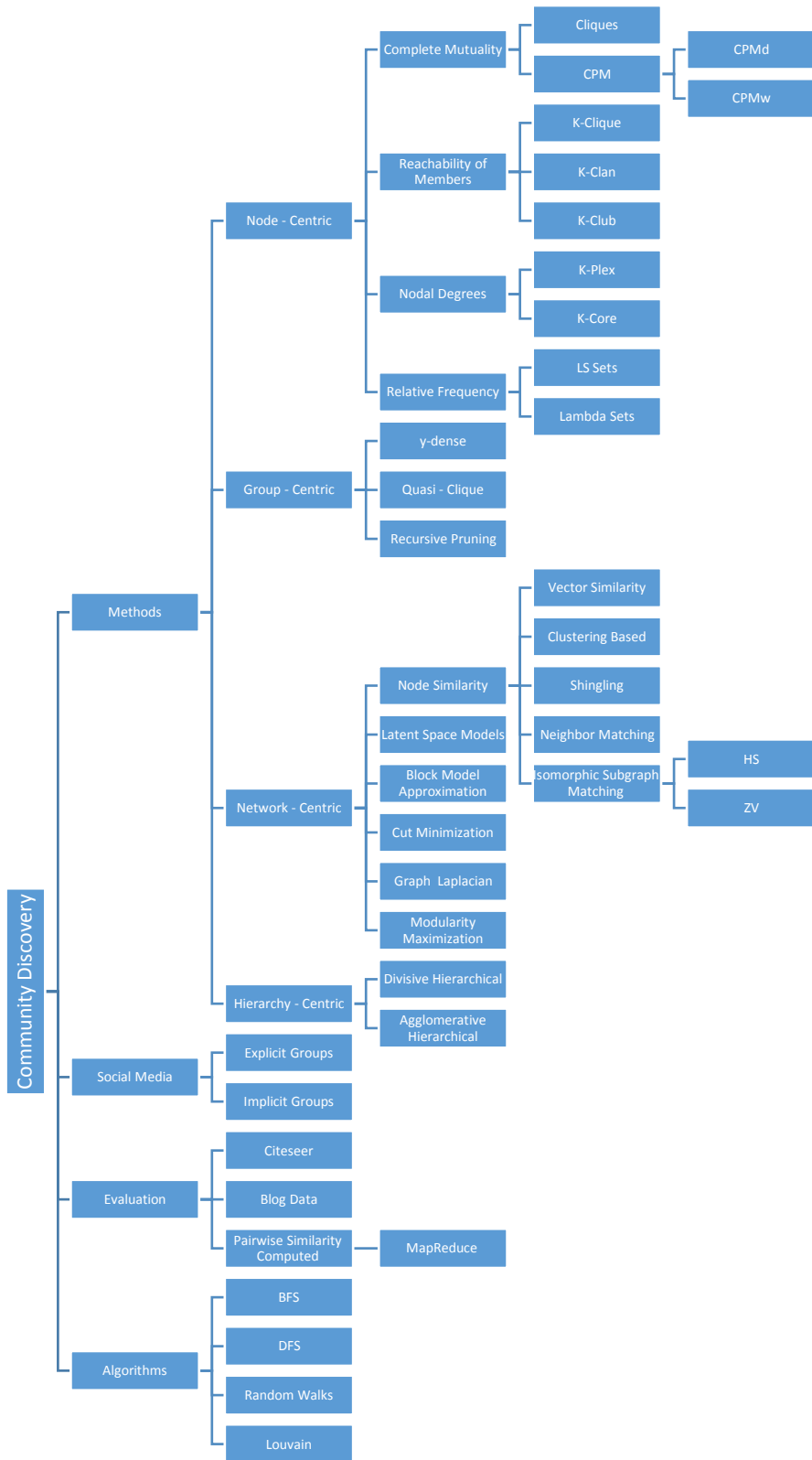
⁴⁹ Danon

۳. فصل سوم: معرفی روش‌های تشخیص

۳.۱. مقدمه

همان‌طور که گفتیم، جامعه از طریق افراد تشکیل می‌شود، بگونه‌ای که آنهایی که با هم در یک گروه، بیش از افرادی که در خارج از گروه قرار دارند، فعل و انفعال می‌کنند، جامعه نامیده می‌شوند. در حقیقت می‌توان گفت که یک جامعه، مجموعه‌ای از گره‌ها است که فعل و انفعال آن‌ها به شکل متناوب انجام می‌شود. با این حال مجموعه افرادی که در یک مکان هستند اما فعل و انفعال نمی‌کنند را نمی‌توان جامعه نامید (افرادی که در یک ایستگاه اتوبوس در انتظار هستند و با یکدیگر صحبت نمی‌کنند). دو نوع گروه در یک رسانه اجتماعی وجود دارند: گروه صریح و گروه ضمنی. گروه صریح از طریق عضو شدن افراد شکل می‌گیرد، یعنی بر اساس خواست خودشان، افراد در شبکه ثبت‌نام می‌کنند و عضو آن می‌شوند. گروه ضمنی از طریق فعل و انفعالات اجتماعی و بدون نیت قبلی شکل می‌گیرد.

همچنین گفتیم که تشخیص جامعه به معنای کشف گروه‌هایی است که در آن گروه‌ها عضویت افراد به شکل صریح، مشخص نیست. این کار به واسطه تدوین گروه‌های اجتماعی توانمند بر اساس ویژگی‌های شبکه اجتماعی انجام خواهد شد. ورودی فرآیند تشخیص جامعه، یک شبکه اجتماعی است و خروجی آن عضویت عاملان در جامعه می‌باشد. اجتماعی بودن انسان، سادگی استفاده از رسانه‌های اجتماعی برای گسترش زندگی اجتماعی به شکل‌هایی که پیشتر ممکن نبودند، و مشکلات ملاقات با افراد دیگری مثل دوستان و خویشاوندان در محیط فیزیکی و سهولت استفاده از رسانه‌های آنلاین برای انجام این کار، باعث شده است که ما به دنبال کشف جوامع در رسانه‌های اجتماعی باشیم. از این فرآیند می‌توان برای درک بهتر روابط میان افراد، مشاهده و پیمایش شبکه‌های بزرگ، و ایجاد مبنایی برای وظایف دیگر مانند داده کاوی استفاده کرد. از کاربرد های دیگر آن می‌توان به بازاریابی ویروسی (مثلاً یافتن مجموعه‌ای از کاربران برای دادن بن تخفیف و انجام تبلیغات) و تشخیص شیوع بیماری (کنترل مجموعه‌ای از گره‌ها که می‌توانند به تشخیص شیوع یا متوقف شدن یک بیماری واگیردار مانند آنفلوآنزا H1N1 کمک کنند) اشاره کرد؛ که هدف از این کار کنترل مسائل پیش گفته و حداکثر سازی بازدهی با توجه به بودجه محدودی است که در اختیار داریم. در این فصل به بررسی روش‌های تشخیص جامعه می‌پردازیم و برای انجام این کار، جامعه، ویژگی‌های آن و روش‌های تشخیص آن را بر مبنای درختی که تهیه کرده‌ایم معرفی می‌کنیم (شکل ۲).



شکل ۲: درخت تشخیص جامعه در شبکه‌های اجتماعی

معیارهای جامعه بر اساس وظایف مختلف، متفاوت هستند. به‌طور تقریبی، روش‌های تشخیص جامعه را می‌توان به چهار گروه تقسیم کرد که البته این چندان انحصاری نیست و ممکن است قابل تغییر باشد.

به‌طور کلی این چهار گروه عبارتند از: جامعه گره-محور که در آن هر گره در یک گروه، برخی ویژگی‌های خاص را دارا می‌باشد. جامعه گره-محور، که در آن روابط درون گروه را به صورت کلی بررسی می‌کنیم. این گروه می‌تواند ویژگی‌های مشخص و معینی را برقرار سازد بدون آنکه روی سطح گره متمرکز شود. جامعه شبکه-محور که کلیت شبکه موجود را در نهایت به چندین دسته مجزا و منفک از یکدیگر، تقسیم می‌کند. شبکه مبتنی بر سلسله مراتب که یک ساختار سلسله مراتبی از جوامع را به وجود می‌آورد.

۳.۲. روش‌های تشخیص جامعه

۳.۲.۱. تشخیص جامعه گره-محور

ویژگی‌های مختلفی که در اینجا برقرار می‌شوند شامل موارد زیر هستند:

- مشارکت متقابل کامل: Cliques
- دسترسی پذیری اعضا: k-clique, k-clan, k-club
- درجه گره: k-plex, k-core
- تناوب نسبی توازی‌های داخلی-خارجی: مجموعه‌های LS ، مجموعه‌های Λ

۳.۲.۱.۱. مشارکت متقابل کامل: Clique

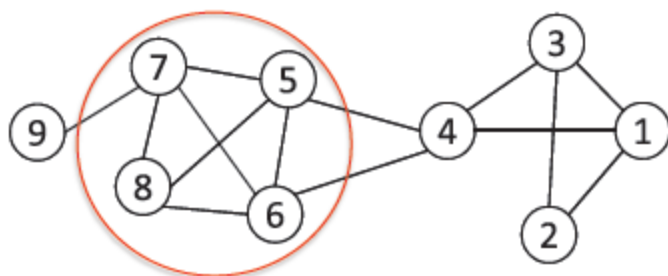
کلمه **Clique** بصورتی که در تئوری گراف‌ها مورد استفاده قرار می‌گیرد، از تحقیقات انجام شده توسط لوس و پری^{۵۰} (۱۹۴۹) به وجود آمد که از زیر-گراف‌های کامل برای مدلسازی کلیک‌ها (گروه افرادی که یکدیگر را می‌شناسند)، در شبکه‌های اجتماعی استفاده کردند. در ریاضیات و در زمینه نظریه گراف، یک کلیک^{۵۱} در یک گراف بدون جهت، یک زیر مجموعه

⁵⁰ Luce & Perry

⁵¹ Clique

از رئوس است که هر دو رأس این زیر مجموعه به وسیله یک لبه^{۵۲} (یال) به هم متصل شده‌اند. کلیک‌ها یکی از مفاهیم بنیادین نظریه گراف هستند و در بسیاری از مسائل دیگر ریاضی نیز مورد استفاده قرار می‌گیرند. در علوم کامپیوتر، مسئله کلیک یک مسئله محاسباتی برای یافتن یک کلیک ماکزیمم و یا همه کلیک‌ها در یک گراف مشخص است. آن را یک مسئله NP – کامل می‌دانیم که یکی از ۲۱ مسئله NP – کامل کارپ^{۵۳} است. به طور خلاصه آن را یک زیر گراف کامل ماکزیمم می‌دانیم که در آن همه گره‌ها در نزدیکی یکدیگر قرار دارند.

نمونه‌ای از آن را می‌توان در شکل ۳ مشاهده کرد:



شکل ۳: گره‌های ۵، ۶ و ۸ یک کلیک را تشکیل می‌دهند

یافتن کلیک ماکزیمم به این صورت است:

- در یک کلیک با اندازه k ، هر گره دارای درجه $k-1 \leq$ می‌باشد.
- گره‌های دارای درجه $k-1 <$ در کلیک ماکزیمم گنجانده نمی‌شوند.
- روش کار به این صورت است که شیوه هرس کردن^{۵۴} به شکل بازگشتی اعمال می‌شود:
- زیر مجموعه‌ای از شبکه (زیر- شبکه^{۵۵}) را از یک شبکه مشخص انتخاب می‌کنیم و یک کلیک را در آن، مثلاً از طریق روش حریص^{۵۶} می‌یابیم.

⁵² Edge

⁵³ Karp

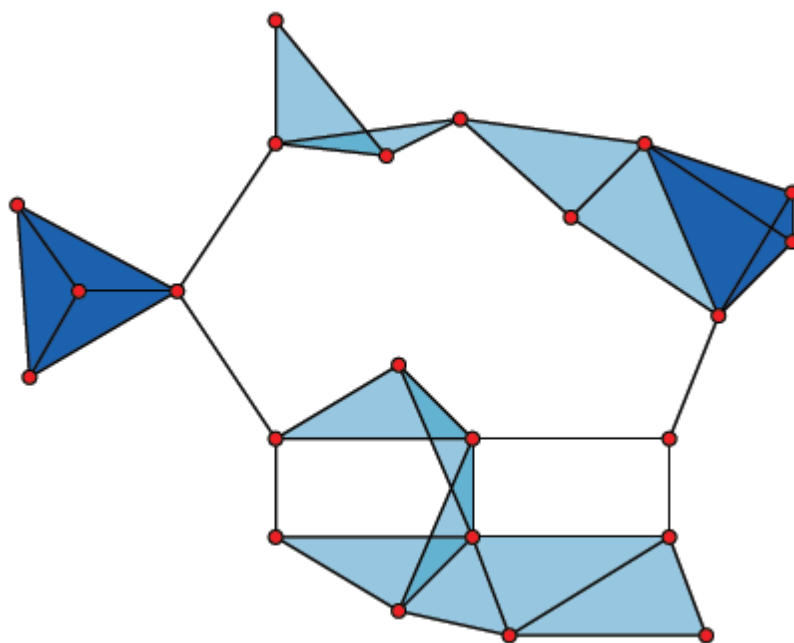
⁵⁴ Pruning

⁵⁵ Sub – network

- فرض کنیم که کلیک بالا دارای اندازه k باشد؛ به منظور یافتن یک کلیک بزرگتر، همه گره های دارای درجه $\leq k-1$ باید حذف شوند.

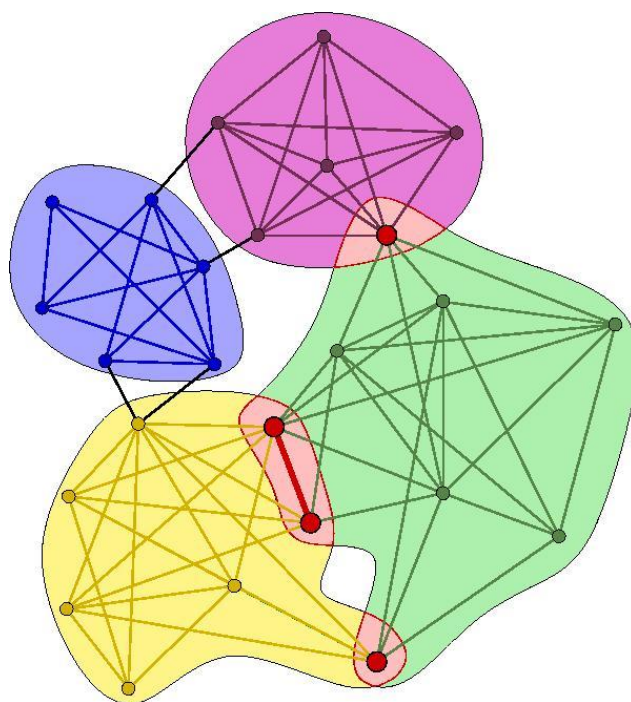
- این کار تکرار می شود تا شبکه به حد کافی کوچک گردد.
- چندین گره هرس می شوند و شبکه اجتماعی از توزیع قانون توان برای درجات گره ها استفاده می کند.

نمونه دیگری از روش کلیک را برای توصیف بیشتر در شکل ۴ نشان می دهیم:



شکل ۱۴: یک گراف کلیک دارای ۲۳ رأس ۱، ۴۲ کلیک رأس ۲ (لبه های آن)، ۱۹ کلیک رأس ۳ (مثلث های آبی کم رنگ و پر رنگ)، و ۲ کلیک رأس ۴ (نواحی آبی پر رنگ)؛ ۱۱ مثلث آبی روشن تشکیل دهنده کلیک های ماکزیمم هستند. دو، ۴-کلیک آبی تیره ماکزیمم و ماکزیمال هستند و تعداد کلیک گراف ۴ است.

CPM روشی رایج است برای آنالیز ساختار هم پوشانی جامعه در شبکه های اجتماعی. که می تواند جوامع را از K-Clique ها به وجود آورد که متناظر با زیر-گراف های k -گره هستند. تعریف آن امکان هم پوشانی بین جوامع را به شکل طبیعی می دهد، همان طور که در شکل ۴ مشاهده می کنیم. در این شکل، تعداد جوامع K-Clique برابر ۴ ($k=4$) است. جوامع با رنگ کدگذاری شده اند و هم پوشانی میان آن ها با رنگ قرمز نشان داده شده است. روش پیش گفته محلی است: اگر یک زیر گراف مشخص، شاخص های مورد نظر را به عنوان یک جامعه برقرار سازد، سپس به طور مستقل از اتفاقی که برای جامعه دیگر در دوردست می افتد، می توانیم آن را یک جامعه در نظر بگیریم. بالعکس در هنگام جستجو به دنبال جوامع به واسطه بهینه سازی از نظر کمیت عمومی، تغییر دوردستی در شبکه می تواند جوامع را در نواحی غیر آشفته، دوباره شکل دهی نماید. علاوه بر این نشان داده شده است که روش های عمومی ممکن است تحت مشکل محدودیت رزولوشن قرار بگیرند.



شکل ۵: توصیف جوامع K-Clique در $k=4$

این روش را می توان برای تشخیص جوامع از مطالعات متا آنالیز سرطان در شبکه های مختلف اجتماعی برای خوشه بندی اسناد و شبکه های اقتصادی مورد استفاده قرار داد.

روش پالایش کلیک جهت‌دار (CPMd)

در شبکه‌ای با پیوندهای جهت‌دار، یک k -Clique جهت‌دار، یک زیر-گراف کامل با k گره است که شرط زیر را برقرار می‌کند:

k گره می‌توانند به گونه‌ای منظم شوند که بین یک زوج اختیاری از آن‌ها، یک لینک جهت‌دار وجود داشته باشد که از رتبه بالاتری در جهت گره دارای رتبه پایین‌تر منتج شود. روش پالایش کلیک جهت‌دار می‌تواند جوامع شبکه را به صورت خوشه‌های پالایشی k -Clique جهت‌دار تعریف کند.

روش پالایش کلیک وزنی (CPMw)

در یک شبکه دارای لینک‌های وزنی، یک k -Clique وزنی در حقیقت عبارت است از یک زیر گراف کامل با k گره به گونه‌ای که میانگین هندسی $k(k-1)/2$ وزن پیوند درون k -Clique بزرگ‌تر از مقدار آستانه انتخابی α است. روش پالایش کلیک وزنی، جوامع شبکه وزنی را به صورت خوشه‌های پالایشی k -Clique وزنی تعریف می‌کند. توجه داشته باشید که میانگین هندسی وزن‌های پیوند در زیر گراف را شدت زیر گراف می‌نامیم.

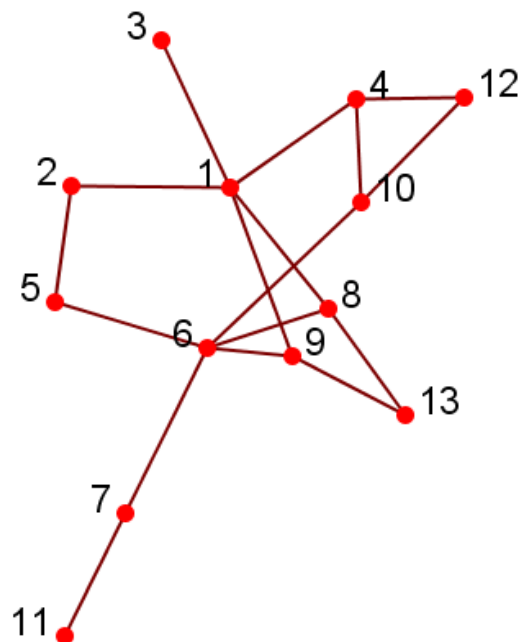
۳.۲.۱.۲. دسترسی پذیری اعضا : k -club, k -clique

هر گره در یک گروه باید حداکثر در تعداد k جهش قابل دسترسی باشد.

k -clique

یک زیرگراف ماکزیمال که در آن بزرگ‌ترین فاصله ژئودزیک بین گره‌ها برابر با $k \leq$ می‌باشد. یک k -clique می‌تواند دارای قطر بزرگ‌تر از k در زیرگراف باشد؛ مثلاً $\{12, 4, 10, 1, 6\}$ 2-clique. در زیرگراف: $d(1, 6) = 3$ (آنها در شکل ۶ می‌توانید مشاهده کنید).

Meta Network



powered by ORA, CASOS Center @ CMU

شکل ۱۵: نمونه k -clique

k-club

یک زیر ساختار با قطر k

مثال : $\{1, 2, 5, 6, 8, 9\}$, $\{12, 4, 10, 1\}$ بصورت ۲-کلاب هستند.

۳.۲.۱.۳. درجات گره: k-core, k-plex

درجه یک گره عبارت است از تعداد لبه‌های متصل به گره. بر حسب ماتریس مجاورت A ، درجه یک گره دارای اندیس i در یک شبکه غیر جهت دار بصورت زیر است:

$$k_i = \sum_j a_{ij},$$

که مجموع آن عبارت است از همه گره‌ها در شبکه.

در یک شبکه جهت دار، هر گره دارای دو درجه می‌باشد. درجه خارجی، تعداد لبه‌های خارجی انشعابی از هر گره است:

$$k_i^{\text{out}} = \sum_j a_{ji},$$

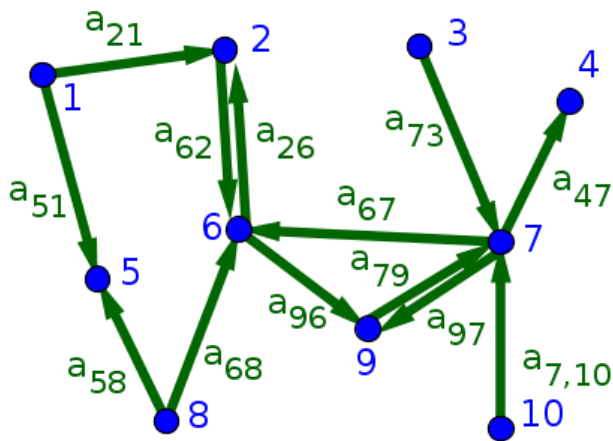
و درجه داخلی تعداد لبه‌های ورودی به داخل یک گره است:

$$k_i^{\text{in}} = \sum_j a_{ij}.$$

درجه کلی گره، مجموع درجات ورودی و خروجی است:

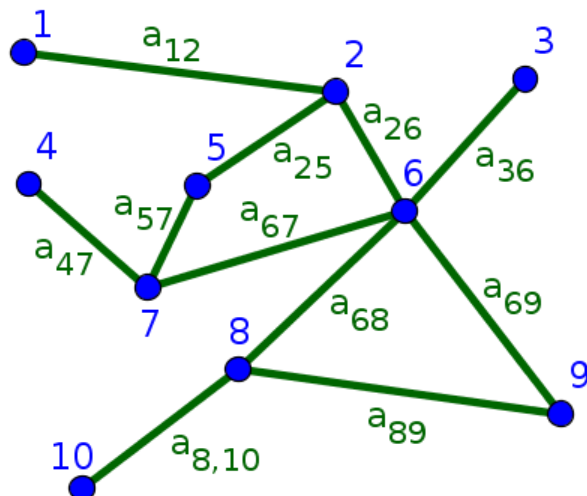
$$k_i^{\text{tot}} = k_i^{\text{in}} + k_i^{\text{out}}.$$

نمونه آنرا می‌توانید در شکل های ۷ و ۸ مشاهده کنید:



شکل ۱۶: برای این شبکه جهت دار، درجه ها عبارتند از:

$kin1=0, kout1=2, ktot1=2$, $kin2=2, kout2=1, ktot2=3$, $kin3=0, kout3=1, ktot3=1$, $kin4=1, kout4=0, ktot4=1$,
 $kin5=2, kout5=0, ktot5=2$, $kin6=3, kout6=2, ktot6=5$, $kin7=3, kout7=3, ktot7=6$, $kin8=0, kout8=2, ktot8=2$,
 $kin9=2, kout9=1, ktot9=3$, $kin10=0, kout10=1, ktot10=1$.

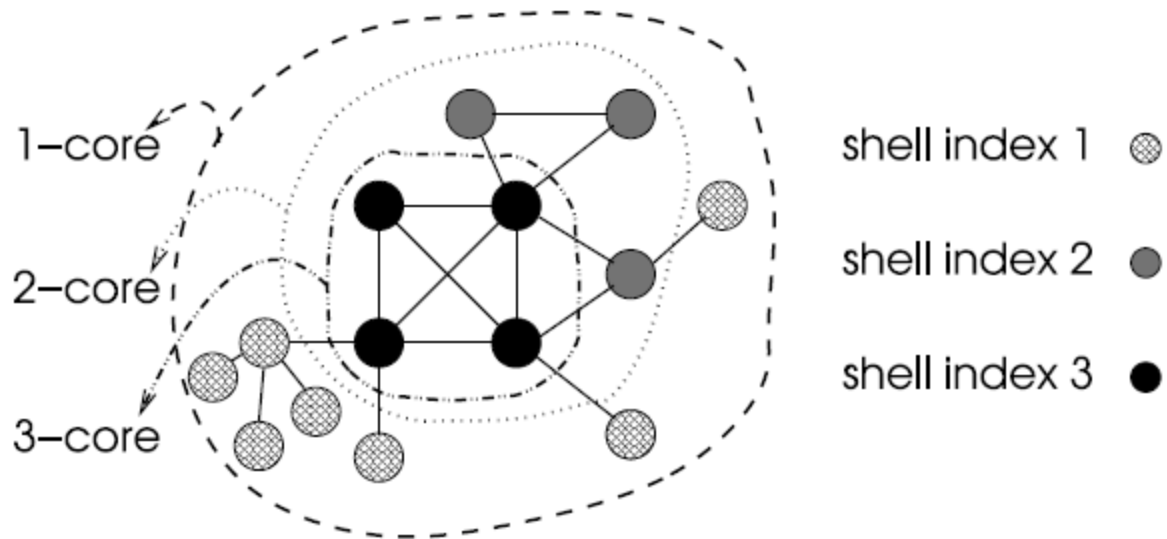


شکل ۱۷: برای این شبکه غیرجهت دار، درجه ها عبارتند از:

$k_1=1, k_2=3, k_3=1, k_4=1, k_5=2, k_6=5, k_7=3, k_8=3, k_9=2, \text{ and } k_{10}=1.$

هر گره باید دارای تعداد مشخصی از ارتباطات با گره های درون یک گروه باشد.

k-core : یک زیر ساخت که هر گره بواسطه آن به حداقل k عضو در گروه متصل می شود. در نتیجه، تجزیه **k-core** هسته های داخلی را به شکل پیوسته شناسایی می کند و شبکه ها را لایه به لایه تغذیه می نماید و ساختار k -پوسته را از خارجی ترین پوسته به داخلی ترین آن ها نمایش می دهد (شکل ۹).



شکل ۱۸: طرح تجزیه k -core برای یک گراف کوچک. هر خط بسته نشانگر مجموعه ای از رئوس متعلق به یک k -core مشخص است که دارای انواع مختلف روز متناظر با k پوسته می باشد.

k -plex: برای یک گروه دارای گره، هر گره باید مجاور حداقل $ns-k$ در گروه باشد.

۳.۲.۱.۴. مجموعه های LS و مجموعه های Lambda

مجموعه های LS

مقوله مجموعه های LS ابتدا توسط لوچیو و سامی^{۵۷} (۱۹۶۹) معرفی شد که عبارت "گروه های مینیمال" را اعلام کردند. لاولر^{۵۸} (۱۹۷۳) آنرا به مجموعه های LS تغییر نام داد (که نام آن از حرف اول نام محققین پدید آورنده بوجود آمد).

ماهیت این ایده آنست که یک مجموعه LS را می توان در حقیقت اتحاد زیر مجموعه های آن دانست و این اتحاد "بهتر" از هر زیر مجموعه ای است زیرا ارتباطات کمتری با خارج دارد. در شرایطی که مجموعه های LS معرفی شدند، گراف ها برای نمایش مدارات مؤلفه های الکترونیکی روی تراشه های سیلیکون مورد استفاده قرار گرفتند. هدف طراحی، گروه بندی مولفه ها در تراشه های فیزیکی بود به گونه ای که برای حداقل سازی تعداد ارتباطات در میان تراشه ها به کار گرفته شود.

⁵⁷ Luccio & Sami

⁵⁸ Lawler

مجموعه های Lambda

یک مجموعه Lambda عبارت است از یک زیر مجموعه ماکزیمم که دارای مسیرهای مستقل از لبه بیشتری در ارتباط با یکدیگر نسبت به مولفه های خارجی می باشند. از آن جا که می توانیم یک مجموعه S، Lambda، را با ایجاد ارتباط ماکزیمم در میان اجزای آن بررسی کنیم، ما همچنین می توانیم مجموعه های Lambda را بصورت مجموعه های λ -k نشان دهیم که در آن ها داریم $k = \lambda(S)$. در دندوگرام^{۵۹} شکل ۱۰، هر خوشه یک مجموعه Lambda است و ارتباطات میان مجموعه ها، با یک سطح خوشه بندی نشان داده می شود.

	0	0	0	0	0	0	0	0	0	1	1	1
λ	1	2	3	4	5	6	7	8	9	0	1	2
4
3	XXXXXXXX	XXXXXXXX
2	XXXXXXXXXXXXXXXXXX	XXXXXXXXXX
1	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX											

شکل ۱۹: خوشه بندی سلسله مراتبی ارتباطات لبه که هر خوشه یک مجموعه لامبدا با ارتباط λ است

محدودیت ها

جوامع گره-محور بیش از حد محدود هستند اما می توانند به عنوان هسته یک جامعه به کار گرفته شوند. آن ها مقیاس پذیر نیستند و معمولاً در آنالیز شبکه های کوچک مورد استفاده قرار می گیرند. گاهی اوقات آن ها با ویژگی های شبکه های مقیاس بزرگ همخوانی ندارند، مثلاً درجه های گرهی برای شبکه های مقیاس آزاد.

⁵⁹ Dendogram

۳.۲.۲.۱. شبه-کلیک ها (Quasi Clique)

شبه-کلیک ها در حقیقت زیر گراف هایی با چگالی لبه مشخص هستند. دو مسئله بهینه‌سازی مرتبط با شبه کلیک به طور طبیعی بوجود می آیند.

در اولی، ما چگالی لبه را تنظیم می کنیم و به دنبال شبه-کلیک حالت کاردینال ماکزیمم، حداقل با چگالی لبه تعیین شده هستیم. در دیگری ما یک لبه کاردینال تثبیت شده را مشخص می کنیم و سپس به دنبال شبه-کلیک چگالی لبه ماکزیمم خواهیم بود.

ارتباطات با گروه را به صورت کلی در نظر بگیرید. برخی از گره‌ها ممکن است ارتباط ضعیفی داشته باشند. یک زیر گراف با V_S گره و E_S لبه، در صورتی شبه-کلیک γ -چگال^{۶۰} است که:

$$\frac{E_S}{V_S(V_S - 1)/2} \geq \gamma$$

۳.۲.۲.۲. پیرایش (هرس کردن) بازگشتی^{۶۱}

هدف از پیرایش بازگشتی، تعیین مقادیر کانتوره ها در عمیق ترین سطح درخت است (یعنی موردی که به گذر^{۶۲} کنونی از داده‌ها اضافه شده باشد)، و می‌توان بازگشت را به گره‌هایی از درخت محدود کرد که دارای نسلی در عمیق ترین سطح کنونی باشند. مشاهده گره‌های دیگر ضروری نیست، زیرا تغییری در گره ها و یا نسل‌ها و آن‌ها پدید نمی‌آید، تنها گره‌های عمیق ترین سطح کنونی درخت تغییر می‌کنند.

روش کار به این صورت است که یک زیر گراف را نمونه برداری می‌کنیم، یک شبه کلیک γ -چگال را می‌یابیم. اندازه برابری خواهد بود با k . گره‌های دارای ویژگی‌های زیر را حذف می‌کنیم:

- که درجه آن‌ها $k\gamma <$ است.
- که همه همسایه‌های آن‌ها دارای درجه $k\gamma <$ هستند.

⁶⁰ γ – dense Quasi Clique

⁶¹ Recursive pruning

⁶² Pass

۳.۲.۳. تشخیص جامعه شبکه محور

در یک تعریف عمومی می توان گفت که کشف شبکه- محور مربوط است به مشارکت به عنوان یک جامعه پیچیده دائماً در حال تکامل از افراد، دستگاه ها، اطلاعات و سرویس های مرتبط شده به یکدیگر برای بهینه سازی مدیریت منابع و ارائه اطلاعات سطح بالاتر در مورد پیشامد ها و شرایط مورد نیاز در جهت افزایش قابلیت های تصمیم گیری. بسیاری عقیده دارند که اصطلاحاتی مانند "اطلاعات - محور" و "دانش - محور" در این زمینه مناسبتر هستند زیرا هدف، یافتن و بهره گیری از اطلاعات است.

برای تشکیل یک گروه، نیاز به بررسی ارتباطات عمومی میان گره های آن داریم. هدف از این روش تشخیص، افراز شبکه به مجموعه های منفصل است. در این روش، گروه بندی بر اساس اصول زیر انجام می شود:

- شباهت گرهی
- مدل فضای متأخر
- تقریب مدل بلوکی
- حداکثر سازی مدولاریته

۳.۲.۳.۱. روش شباهت گرهی

ویژگی های زیر در روش شباهت گرهی برقرار خواهند بود:

- اگر گراف با خودش مقایسه شود، هر گره باید مشابه خودش باشد. این یک ویژگی طبیعی برای همه معیارهای شباهت است.
- نمرات شباهت باید دارای دامنه ثابت باشند، شباهت یک گره به خودش همیشه دارای مقدار ماکزیمم است. این مشخصه برای معیارهای شباهت به صورت کلی کاملاً اختیاری است که در یک دامنه ثابت قرار داشته باشند.
- نمره شباهت باید خود معنادار باشد. به واسطه نرمال سازی ماتریس شباهت، نمره شباهت X_{ij} تنها در صورتی می تواند تغییر کند که نمرات دیگر شباهت نیز به همان شکل تغییر کنند.
- اگر دو گره دارای لبه های ورودی یا خروجی نباشند، آن ها را باید مشابه در نظر گرفت. تا جایی که ما می دانیم، این مشخصه تنها در روش هیمانز و سینگ^{۶۳} وجود دارد.

۳.۲.۳.۱.۱. روش انطباق همسایگی

از نرمال سازی ماتریس اجتناب می شود و ارائه نمونه هایی از گراف ها به گونه ای که همه مقادیر شباهت صفر باشند و یا همه مقادیر شباهت ۱ باشند آسان خواهد بود. شباهت گره ها به واسطه عدم وجود همسایه های درونی یا بیرونی از آنجا تشخیص داده

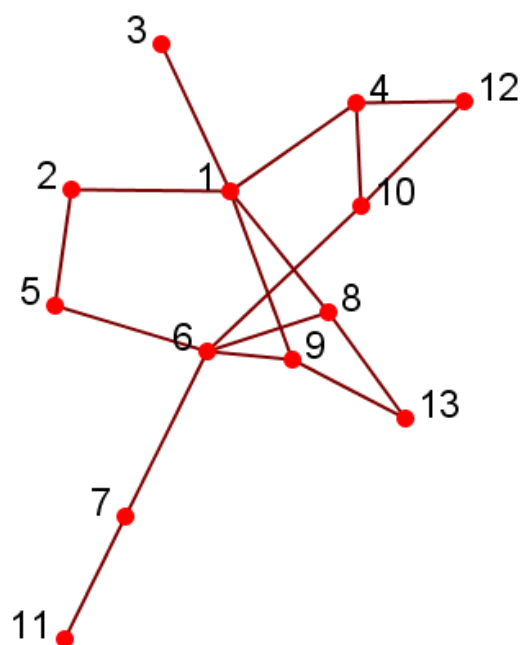
⁶³ Heymans and Singh

خواهد شد که در این حالت، شباهت درونی یا بیرونی برابر با ۱ باشند. در نتیجه مشاهده می کنیم که این روش همه شرایط پیش گفته را برقرار می کند.

همچنین روش های دیگری که با انطباق همسایگی مقایسه می شوند، روش هیمانز و سینگ (HS) و زاگر و ورگیس (ZV) هستند. هر دو این روش ها کاملاً تحت تأثیر پارامتر چگالی p در زمینه سرعت و بازدهی انطباقی می باشند.

در شکل ۱۱ می توانید به راحتی هم ارزی ساختاری دو گره را مشاهده کنید.

Meta Network



powered by ORA, CASOS Center @ CMU

شکل ۲۰: گره های ۸ و ۹ به لحاظ ساختاری هم ارز هستند

محدودیت ها

این روش بیش از حد محدود است و به ندرت در مقیاس بزرگ کاربردی دارد. محاسبه گروه هم ارز آزاد در آن دشوار است.

۳.۲.۳.۱.۲. شباهت برداری

در عمل بدلیل این محدودیت ها از شباهت برداری استفاده می شود، که نمونه های آن را می توان شباهت کسینوسی و شباهت ژاکارد^{۶۴} دانست.

مراحل کار برای خوشه بندی بر مبنای شباهت برداری عبارت است از :

- ارتباطات را بعنوان ویژگی ها در نظر می گیریم.
- از شباهت کسینوسی یا شباهت ژاکارد برای محاسبه شباهت رؤس استفاده می کنیم.

از الگوریتم خوشه بندی کلاسیک k-means بهره می گیریم :

- هر خوشه مرتبط با یک مرکز (نقطه مرکزی) است.
- هر گره مرتبط با یک خوشه با نزدیک ترین مرکز است.

الگوریتم مبنای k-means را در زیر مشاهده می کنید :

Algorithm 1 Basic K-means Algorithm.

- 1: Select K points as the initial centroids.
 - 2: **repeat**
 - 3: Form K clusters by assigning all points to the closest centroid.
 - 4: Recompute the centroid of each cluster.
 - 5: **until** The centroids don't change
-

شکل ۲۱: الگوریتم مبنای k-means

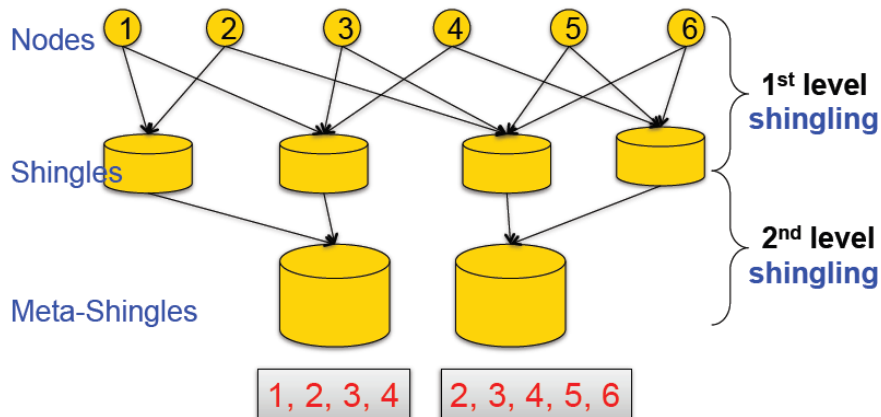
^{۶۴} Jaccard

۳.۲.۳.۱.۳. شینگلینگ^{۶۵}

محاسبه زوجی شباهت با استفاده از میلیون‌ها گره می‌تواند زمان بر باشد. شینگلینگ را می‌توان بر مبنای اصول زیر به کار گرفت:

- نگاشت هر بردار در داخل شینگلینگ‌های مختلف به گونه‌ای که شباهت ژاکارد بین دو بردار بتوانند از طریق مقایسه شینگل‌ها محاسبه شوند.
- با استفاده از تابع سریع هش
- بردارهای مشابه دارای شینگل‌های مشترک بیشتری پس از تبدیل هستند.

گره‌هایی که دارای بیش از یک شینگل هستند را می‌توان متعلق به یک جامعه دانست. در حقیقت ما یک شینگلینگ دو لایه را اعمال می‌کنیم (آن را در شکل ۱۳ مشاهده می‌کنید)



شکل ۲۲: شینگلینگ سریع دو سطحی

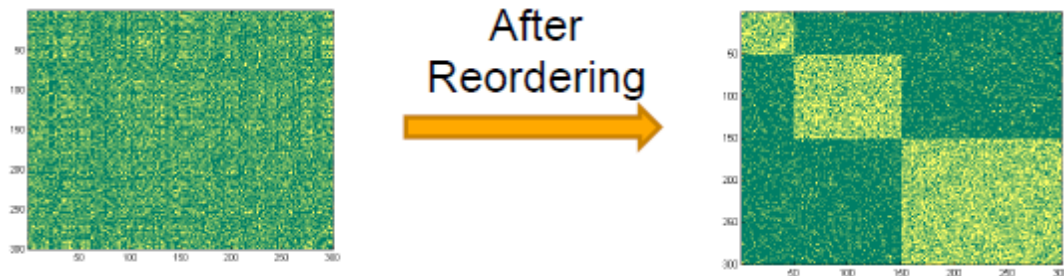
^{۶۵} Shingling

۳.۲.۴. مدل فضای متاخر

مورد دیگری در روش های تشخیص جامعه در شبکه های اجتماعی عبارت است از مدل های فضای متاخر؛ که گره های موجود در یک شبکه را به یک فضای دارای بعد کمتر تبدیل می کنند به گونه ای که فاصله یا شباهت میان گره ها، در فضای اقلیدسی حفظ شود.

۳.۲.۵. تقریب مدل بلوکی

هدف از این روش، حداقل سازی فاصله میان یک ماتریس فعل و انفعال و یک ساختار بلوکی میباشد. آن را به صورت زیر نشان می دهیم.



شکل ۲۳: شکل سمت چپ نشان دهنده ماتریس فعل و انفعال شبکه می باشد و شکل سمت راست ساختار بلوکی را نمایش می دهد.

حداقل سازی برش

در این روش، فعل و انفعالات میان گروهی باید نامتناوب باشد. برش در میان تعداد لبه های میان دو مجموعه از گره ها انجام

$$cut(C_1, C_2, \dots, C_k) = \sum_{i=1}^k cut(C_i, \overline{C_i})$$

می شود. هدف، حداقل سازی برش است.

دو گونه رایج آن عبارتند از برش نسبت:

$$C_1, C_2, \dots, C_k = \sum_{i=1}^k \frac{cut(C_i, \overline{C_i})}{|V_i|}$$

و برش نرمال سازی شده

$$(C_1, C_2, \dots, C_k) = \sum_{i=1}^k \frac{cut(C_i, \overline{C_i})}{vol(V_i)}$$

۳.۲.۶. لاپلاسیان گراف^{۶۶}

هدف اساسی برش را می توان به مسئله min-trace زیر ساده سازی کرد

$$\min_{S \in R^{n \times k}} Tr(S^T L S) \quad s.t. \quad S^T S = I$$

که در آن L لاپلاسیان گراف (نرمال ساز شده) می باشد

$$\begin{aligned} L &= D - A \\ \text{normalized-}L &= I - D^{-1/2} A D^{-1/2} \end{aligned} \quad D = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix}$$

۳.۲.۷. حداکثر سازی مدولاریته

مدولاریته به اندازه گیری فعل و انفعالات گروهی در مقایسه با ارتباطات رندوم مورد انتظار در گروه می پردازد. در یک شبکه دارای m لبه (بال)، برای دو گره دارای درجه های d_i و d_j ، ارتباطات مورد انتظار بین آن ها به صورت زیر هستند:

$$d_i d_j / 2m$$

سودمندی فعل و انفعال در یک گروه:

$$\sum_{i \in C, j \in C} A_{ij} - d_i d_j / 2m$$

⁶⁶ Graph Laplacian

برای افراز گروه به چندین گروه، عبارت زیر را به حداکثر می‌رسانیم :

$$\frac{1}{2m} \sum_C \sum_{i \in C, j \in C} A_{ij} - d_i d_j / 2m$$

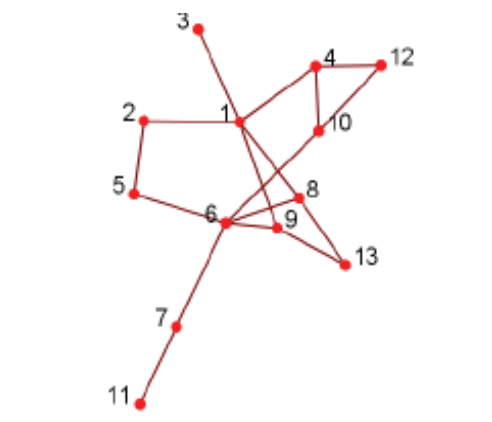
ماتریس مدولاریته

حداکثر سازی مدولاریته را می‌توان با فرمول ماتریس به شکل زیر بیان کرد :

$$Q = \frac{1}{2m} Tr(S^T B S)$$

B ماتریس مدولاریته است.

$$B_{ij} = A_{ij} - d_i d_j / 2m$$



شکل ۲۴: تعداد مورد انتظار لبه‌های بین ۶ و ۹ به صورت زیر است : $5 \times 3 / (2 \times 17) = 15/34$

ویژگی‌های مدولاریته

از ویژگی‌های مدولاریته می‌توان به این موضوع اشاره کرد که بین $(1, -1)$ قرار دارد و اینکه در صورتی که همه گره‌ها در یک گروه، خوشه بندی شوند، مدولاریته $= 0$ است. مدولاریته می‌تواند به شکل اتوماتیک تعداد بهینه خوشه‌ها را تعیین کند. در پایان می‌توان از محدودیت‌های روش تشخیص جامعه شبکه-محور، به نیاز به تعیین تعداد جوامع پیش از هر گونه عملیاتی توسط کاربر در این روش اشاره کرد.

۳.۲.۸. تشخیص جامعه بر مبنای سلسله مراتب

هدف از تشخیص جامعه بر مبنای سلسله مراتب عبارت خواهد بود از ایجاد یک ساختار سلسله مراتبی بر اساس توپولوژی شبکه که می‌تواند در نهایت آنالیز را در سطوح مختلف دقت (رزولوشن‌های گوناگون) تسهیل نماید.

روش‌های معرف اینگونه تشخیص عبارتند از:

- روش خوشه بندی سلسله‌مراتبی مقسم
- روش خوشه بندی سلسله‌مراتبی تجمعی

۳.۲.۸.۱. خوشه بندی سلسله‌مراتبی مقسم

این روش، گره‌ها را در چندین مجموعه قرار می‌دهد. در آن، هر مجموعه به مجموعه‌های کوچک‌تر تقسیم می‌شود. روش‌های مبتنی بر شبکه را می‌توان برای افراز مجموعه به کار گرفت. یک نمونه بخصوص از آن مبتنی است بر میانه لبه‌ها^{۶۷}؛ که در آن کوتاه‌ترین مسیرهای بین هر زوج از گره‌ها از لبه عبور می‌کند. لبه‌های (بال‌های) میان-گروهی دارای میانه لبه‌های بزرگ‌تری هستند. خوشه بندی مقسم بر مبنای میانه لبه‌ها، لبه‌های دارای بالاترین حالت میانی را حذف می‌کند.

⁶⁷ Edge-betweenness

۳.۲.۸.۲. خوشه بندی سلسله مراتب تجمعی

در ابتدا هر گره را به عنوان یک جامعه مقدار دهی می کند. دو جامعه که تأمین کننده شاخص های زیر باشند را انتخاب می کند و آن ها را در یکدیگر ادغام کرده و به صورت جوامع بزرگ تری درمی آورد.

- حداکثر افزایش مدولاریته
- حداکثر شباهت گرهی

اغلب الگوریتم های خوشه بندی سلسله مراتبی، یک درخت باینری را خروجی می دهند. هر گره دارای دو گره فرزند خواهد بود و ممکن است نامتعادل باشد.

خوشه بندی تجمعی ممکن است نسبت به مرحله پردازش گره ها و شاخص های یکپارچه سازی (ادغام) بسیار حساس باشد.

خوشه بندی مقسم پایدارتر است اما معمولاً محاسبات پرهزینه تری دارد.

۳.۳. شبکه های اجتماعی صریح و ضمنی

در این بخش، از جنبه عمومی به تعریف کلی شبکه های اجتماعی صریح و ضمنی می پردازیم. در سال های اخیر، پلتفرم های آنلاین مختلفی برای شبکه های اجتماعی به وجود آمده اند که تا حد زیادی از جنبه های اقتصادی و اجتماعی مورد توجه عموم قرار گرفتند. با این حال جنبه های مختلفی برای این مسئله وجود دارد. به عنوان مثال، فیسبوک مجموعه بسیار بزرگ از کاربران را به دست آورده است اما نتوانسته بطور متناسب از نظر سودآوری موفقیت آمیز باشد.

شبکه های اجتماعی آنلاین می توانند به طور گسترده به شبکه های اجتماعی صریح و ضمنی تقسیم بندی شوند:

۳.۳.۱. شبکه های اجتماعی صریح

شبکه های اجتماعی صریح (مانند فیسبوک، لینکداین، توئیتر، و مای اسپیس) جایی هستند که در آن کاربران، شبکه را به واسطه ارتباط صریح با سایر کاربران، احتمالاً، اما نه ضرورتاً، بر مبنای علایق مشترک به وجود می آورند.

۳.۳.۲. شبکه‌های اجتماعی ضمنی

شبکه‌های اجتماعی ضمنی (مانند لاست-اف^{۶۸}، اوت-برین^{۶۹}، و کالر^{۷۰}) شبکه‌هایی هستند که در آن‌ها کاربر به واسطه علائق و توجهاتی که دارد تعریف می‌شود و ارتباطات (ضمنی) بین کاربران، توسط خودشان، صریحاً به وجود نمی‌آید، بلکه صرفاً بر مبنای علائق و رفتار آنلاین آن‌ها تشکیل می‌شود. یک شبکه اجتماعی ضمنی می‌تواند موقتی باشد و تنها تا زمانی که لازم است باقی بماند، دقیقاً برخلاف اغلب شبکه‌های صریح. برای مثال، کالر می‌تواند موقعیت کاربران را هم‌زمان تعیین کند و گراف اجتماعی ضمنی آن‌ها را تشکیل دهد، و سپس از آن برای معرفی آیتم‌ها در میان کاربرانی که ممکن است یکدیگر را حتی نشناسند، استفاده می‌کند. این شبکه اجتماعی به کاربرانی که عکس‌های خود را از یک پیشامد گرفته‌اند و فرستاده‌اند (مثلاً از عروسی، بازی، مراسم موسیقی و غیره) اجازه می‌دهد که عکس‌های سایرین را بر اساس ماتریس‌های مبتنی بر موقعیت مکانی مشاهده کنند. برخلاف شبکه‌های صریح، شبکه‌های ضمنی محدود به کاربرانی که با هم دوست یا آشنا هستند نخواهند بود.

۳.۴. ارزیابی

۳.۴.۱. بررسی شباهت‌های جفت به جفت

محاسبه شباهت‌های جفت جفت در مجموعه‌های بزرگ اسناد وظیفه‌ای است که در انواع مسائل مختلف مانند خوشه بندی و تفکیک مرجع‌های مشترک میان اسناد اهمیت دارد. به عنوان مثال، در موتور جستجوی PubMed که اطلاعاتی را در مورد مقالات پزشکی به دست می‌دهد، یک ویژگی مرورگری “more like this” به صورت یک جستجوی ساده در میان نمرات تشابه اسناد، که به صورت آفلاین محاسبه می‌شوند، پیاده‌سازی شده است.

چارچوب MapReduce یکی از الگوریتم‌هایی است که امکان تفکیک حاصلضرب‌های داخلی دخیل در محاسبه شباهت اسناد را برای تفکیک مراحل ضرب و جمع به طریقی که به خوبی با الگوهای دسترسی به دیسک در چندین دستگاه منطبق باشند، را فراهم می‌آورد.

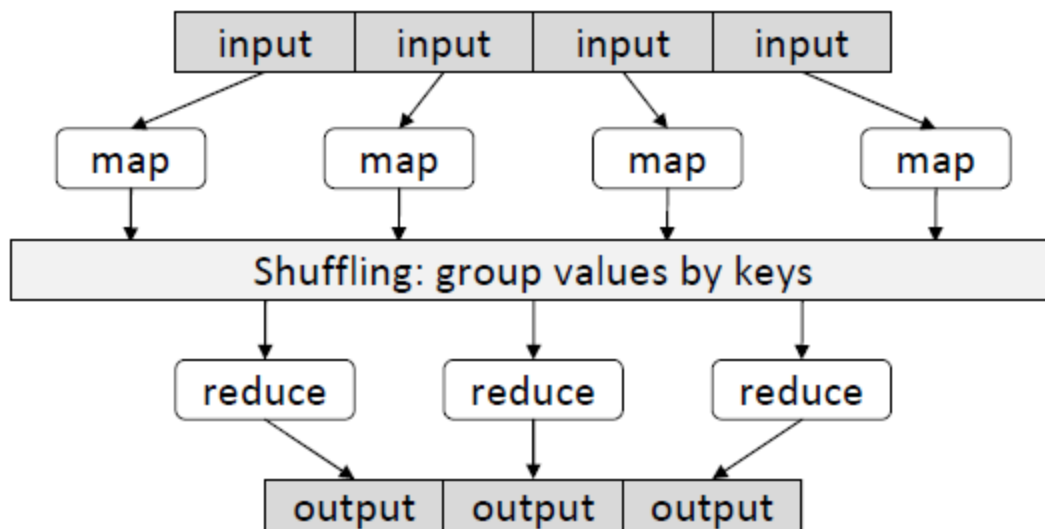
MapReduce مبتنی بر این مشاهده می‌باشد که چندین وظیفه دارای ساختار مشابهی هستند: محاسبه در تعداد زیادی از بایگانی‌ها اعمال می‌شود تا نتایج جزیی را حاصل آورد، که می‌توان آن‌ها را به طرق مختلف با هم ادغام کرد. به طور طبیعی، محاسبات و گردآوری در هر رکورد یا بایگانی براساس وظیفه مورد نظر متفاوت است، اما شاهکار مبنای آن‌ها یکسان باقی می‌ماند. MapReduce از طریق الهام گرفتن از توابع دارای مرتبه بالاتر در برنامه‌نویسی عاملی، می‌تواند انتزاعی را به دست دهد که شامل تعریف یک «نقشه بردار» و یک «کاهنده» توسط برنامه نویس است.

⁶⁸ Last.FM

⁶⁹ Outbrain

⁷⁰ Color

شکل ۱۶ نشان‌دهنده توصیفی از چارچوب MapReduce می‌باشد.



شکل ۲۵: توصیف چارچوب MapReduce : نقشه بردار برای همه بایگانی‌های ورودی اعمال می‌شود و نتایج را به وجود می‌آورد که توسط «کاهنده» گردآوری می‌شوند.

CiteSeer.۳.۴.۲

CiteSeer نیز یکی از سیستم‌های دسته‌بندی مراجع برای مقالات آکادمیک در فرمت الکترونیکی است. CiteSeer می‌داند که چگونه باید مراجع را افراز کند، آن‌ها را در یک مقاله مشابه با فرمت‌های مختلف بشناسد، و حوزه‌های ارجاعات را در بدنه مقالات شناسایی کند. این الگوریتم اغلب مزایای اندیس‌های ارجاعات معمول سنتی را در اختیار دارد که شامل موارد زیر می‌باشد: بازیابی مقاله از طریق پیوندهای (لینک‌های) ارجاعی (مثلاً ارائه فهرستی از مقالات که به یک مقاله مشخص ارجاع می‌کنند)، ارزیابی و رتبه‌بندی مقالات، محققین، ژورنال‌ها و غیره بر اساس تعداد ارجاعات، و شناسایی روند‌های تحقیق. CiteSeer مزایای زیادی نسبت به فهرست‌های ارجاعی سنتی دارد که شامل قابلیت ایجاد پایگاه‌های داده بروز رسانی شده می‌باشد که محدود به یک مجموعه از پیش تعیین شده از ژورنال‌ها نخواهند بود، عملکرد کاملاً مستقل با کاهش متناظر هزینه‌ها را در بر می‌گیرد، و مرورگری فعل و انفعالی توانمند را در میان مقالات با استفاده از حوزه ارجاعات انجام می‌دهد. با در اختیار داشتن یک مقاله به خصوص، CiteSeer می‌تواند چگونگی ارجاع به مقاله را در تحقیقات متعاقب نشان دهد. این حوزه می‌تواند

دربگیرنده خلاصه کوتاهی از مقاله، پاسخ و واکنش محققین دیگر به مقاله، و فعالیت‌های متعاقب مبتنی بر مقاله اصلی باشد. CiteSeer امکان یافتن مقالات را با استفاده از جستجوی کلمات کلیدی و یا با پیوندهای (لینک‌های) مربوط به ارجاعات، فراهم می‌آورد. موقعیت مربوط به یک مقاله به خصوص را می‌توان با استفاده از اطلاعات و یا شباهت برداری کلمات تعیین کرد.

یک فهرست از ارجاعات، پیوندهای میان مقالات را که محققین در هنگام ارجاع به یک مقاله دیگر به کار می‌گیرند، ارائه می‌کند. فهرست‌های ارجاع می‌توانند برای اهداف مختلف مفید باشند، مانند جستجوی مقالات، ارزیابی آن‌ها، آنالیز مقالات آکادمیک و غیره.

CiteSeer مقالات موجود روی شبکه جهانی وب را دانلود می‌کند، آن‌ها را به متن (text) تبدیل می‌کند، آن‌ها را برای استخراج ارجاعات و دامنه‌های ارجاعات در پیکره مقاله افراز می‌کند و و اطلاعات را در یک پایگاه داده‌ای ذخیره می‌کند. CiteSeer شامل الگوریتم‌هایی برای شناسایی و گروه بندی اشکال مختلف ارجاعات یک مقاله می‌باشد و عملکرد فهرست بندی تمام-متنی را در مقالات و ارجاعات انجام می‌دهد و همچنین پشتیبانی لازم را برای مرورگری از طریق پیوندهای ارجاعات به انجام می‌رساند. مقالات مرتبط به یک مقاله به خصوص را می‌توان با استفاده از اطلاعات لازم و یا شباهت برداری کلماتی یافت.

معایب CiteSeer در مقایسه با سایر فهرست‌های ارجاعی تجاری کنونی عبارتند از:

- CiteSeer پوشش کاملی را از همه ژورنال‌های مهم انجام می‌دهد. ما انتظار داریم که این موضوع با در دسترس قرار گرفتن مقالات بیشتر در طول زمان در حوزه آنالیز، مشکلات و معایب بیشتری را به وجود بیاورد (برای فرستنده ژورنال‌های بیشتر، نیاز به موافقت‌نامه و همکاری با انتشاراتی‌های مختلف وجود دارد).
- CiteSeer نمی‌تواند زیر شاخه‌های مستقل را دقیقاً شناسایی کند، مثلاً نمی‌تواند میان دو محقق که دارای نام مشابهی هستند تفکیک قائل شود. ما انتظار داریم که CiteSeer این موضوع را در طول زمان بهبود بخشد؛ که این کار از طریق جمع‌آوری پایگاه‌های داده‌ای گسترده و بهسازی آنالیزهای مورد استفاده در CiteSeer امکان‌پذیر خواهد بود.

مزایای CiteSeer در مقایسه با فهرست‌های ارجاعی تجاری کنونی عبارتند از:

- از آنجا که CiteSeer می‌تواند مقالات را به محض آن که در شبکه جهانی وب قرار می‌گیرند فهرست کند، باید کاربرد بیشتری برای محققین در جهت یافتن مقالات اخیر و برای به روز ماندن آن‌ها داشته باشد.
- CiteSeer مستقل است و نیازمند فعالیتی از طرف فرد در هنگام فهرست بندی نیست.

- CiteSeer می‌تواند برای انجام ارزیابی‌های دقیق‌تر از اثرات یک مقاله خاص مورد استفاده قرار بگیرد.

به‌طور خلاصه می‌توان گفت که در حال حاضر به دلیل سازماندهی شبکه جهانی وب، CiteSeer هم‌اکنون نمی‌تواند همانند سیستم‌های سنتی، فهرست بندی جامعی را انجام دهد، زیرا بسیاری از مقالات منتشر شده در حال حاضر به شکل آنلاین در دسترس نیستند. با این حال تا حد زیادی در حال تغییر است و انتظار می‌رود که این نقیصه در طول زمان برطرف شود. همچنین به واسطه مشکلات بالقوه، مثلاً در شناسایی اتوماتیک و ابهام زدایی مربوط به محققین، فهرست CiteSeer ممکن است همانند فهرست‌های سنتی دستی دقیق نباشد. با این حال، CiteSeer ممکن است به یک انسان اجازه بدهد که دقت نتایج را به واسطه مرورگری اطلاعات دقیق و شرایط ارجاع در پایگاه داده‌ای، به شکل ساده‌تری نسبت به سیستم‌های فهرست بندی ارجاعات سنتی ارزیابی کنیم.

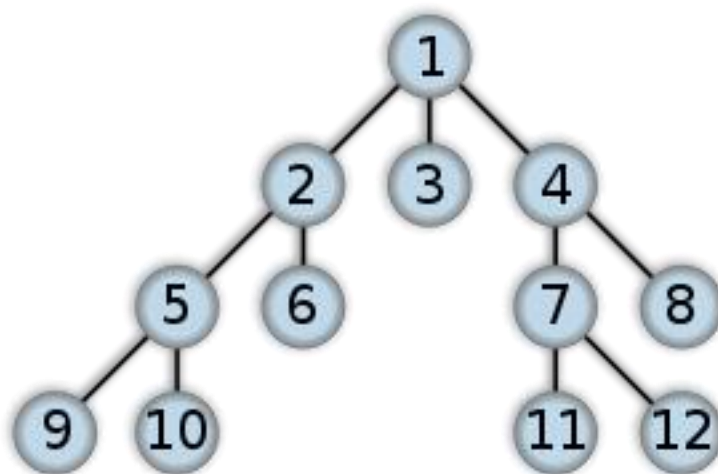
برخی از قابلیت‌های مهم CiteSeer عبارتند از: a. قابلیت افراز ارجاعات از مقالات و شناسایی ارجاعات در مقالات مشابهی که ممکن است ساختار متفاوتی داشته باشند، این مسئله امکان ایجاد آمارهایی را در ارتباط با ارجاعات به همراه گروه بندی مقالاتی که به یک مقاله مشخص ارجاع داده‌اند، فراهم می‌آورد؛ b. توانایی استخراج و نشان دادن دامنه ارجاعات به یک مقاله به خصوص، و ایجاد این امکان برای محقق‌هایی که چیزی را که محققین دیگر در خصوص یک مقاله خاص گفته‌اند، مشاهده کند؛ c. توانایی یافتن مقالات مرتبط بر مبنای ارجاعات مشترک و همچنین بر مبنای شباهت فاصله رشته‌ای و یا بردار کلمات.

۳.۵. الگوریتم‌ها

۳.۵.۱. breadth-first search (BFS)

در تئوری گراف، **breadth-first search** یا **BFS** یک استراتژی برای جستجو در یک گراف می‌باشد و مربوط به زمانی است که جستجو اساساً محدود به دو عملکرد باشد: a. بازدید و بازبینی یک گره از یک گراف، b. دسترسی پیدا کردن برای مشاهده گره‌هایی که در مجاورت گرهی هستند که به تازگی بازدید شده است. BFS در یک گره ریشه آغاز می‌شود و همه گره‌های مجاور را بازدید می‌کند. سپس برای هر یک از این گره‌های مجاور به ترتیب، گره‌های مجاور آن‌ها را که مشاهده نشده بودند بررسی می‌کند و همین‌طور الی آخر.

BFS ابتدا در اواخر دهه ۵۰ توسط ای. اف. مور^{۷۱} اختراع شد که برای یافتن کوتاه‌ترین مسیر در میان یک مارپیچ^{۷۲} به کار گرفته شد و به طور مستقل نیز توسط سی. وای. لی^{۷۳} به عنوان یک الگوریتم مسیر یابی مورد استفاده قرار گرفت (۱۹۶۱).



شکل ۲۶: ترتیب مورد انتظار برای گره‌ها در الگوریتم *breadth-first search*

در هنگام آنالیز الگوریتم‌ها، ورودی الگوریتم جستجوی **breadth-first search** می‌تواند یک گراف متناهی باشد که توسط یک فهرست مجاورت یا مورد مشابهی نمایش داده می‌شود. با این حال در کاربرد روش‌های پیمایش گراف در هوش مصنوعی، این ورودی می‌تواند نمایشی ضمنی از یک گراف نامتناهی باشد. در این حوزه، یک روش جستجو در صورتی کامل تلقی می‌گردد که به شکل تضمین شده، یک حالت هدف را در صورت وجود بیابد. الگوریتم **breadth-first search** کامل است اما الگوریتم **depth-first search** این طور نیست: هنگامی که به گراف‌های نامتناهی نمایش داده شده به شکل ضمنی اعمال می‌شود، در بخش‌هایی از گراف که حالت هدف ندارند و هرگز باز نمی‌گردند، گم می‌شود.

یک روش جستجو در صورتی بهینه می‌باشد که یافتن بهترین راه حل موجود، تضمین شده باشد. به عبارت دیگر، این روش، مسیری را به سمت حالت هدف خواهد یافت که دربرگیرنده حداقل تعداد مراحل باشد.

⁷¹ E.F. Moore

⁷² Maze

⁷³ C. Y. Lee

می‌توان از الگوریتم **breadth-first search** برای حل بسیاری از مسائل در تئوری گراف، استفاده کرد؛ مثلاً:

- گرد آوری کپی با الگوریتم چنی^{۷۴}
- یافتن کوتاه‌ترین مسیر بین دو گره u و v که در آن طول مسیر از طریق تعداد گوشه‌ها اندازه گرفته می‌شود.
- آزمایش یک گراف از نظر حالت دو بخشی^{۷۵}
- شماره گذاری (معکوس) مش کوتیل - مک کی^{۷۶}
- روش فورد-فولکرسون^{۷۷} برای محاسبه جریان ماکزیمم در یک شبکه جریان
- سریالی-سازی/سریال-زدایی یک درخت باینری در مقابل سریالی-سازی با ترتیب منظم، که امکان بازسازی درخت را به شکل پربازده فراهم می‌کند.
- ساخت تابع شکست تطبیق دهنده الگوی آهو-کوراسیک^{۷۸}

۳.۵.۲. (DFS) Depth-first search

(DFS) Depth-first search یک الگوریتم برای پیمایش جستجوی ساختارهای داده ای گراف یا درخت است که در

ریشه آغاز می‌شود (از طریق گزینش یک گره اختیاری به عنوان ریشه در خصوص یک گراف) و در امتداد هر شاخه پیش از ردگیری معکوس^{۷۹} اکتشافی را انجام می‌دهد.

نسخه‌ای از جستجوی **Depth-first search** در قرن نوزدهم توسط ریاضی‌دان فرانسوی ، چارلز پیر ترومو به عنوان یک استراتژی برای حل مارپیچ‌ها مورد بررسی قرار گرفت.

⁷⁴ Cheney

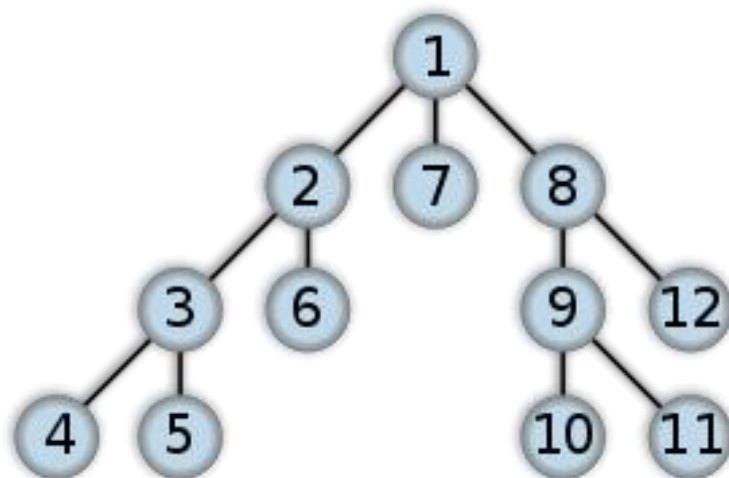
⁷⁵ Bipartiteness

⁷⁶ (Reverse) Cuthill-McKee mesh numbering

⁷⁷ Ford - Fulkerson

⁷⁸ Aho-Corasick

⁷⁹ Backtracking



شکل ۲۷: ترتیب مشاهده گره‌ها

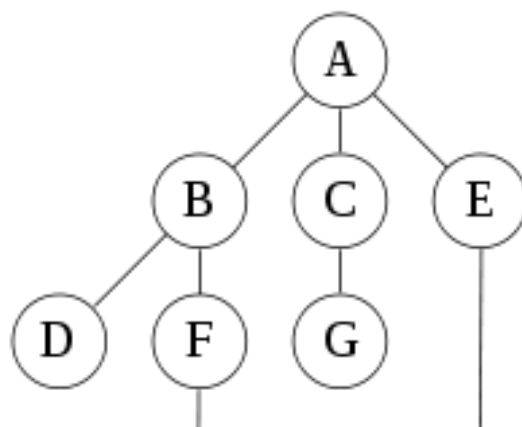
ویژگی‌ها

آنالیز زمانی و فضای DFS بر اساس حوزه کاربردها متفاوت است. در علوم کامپیوتر DFS برای پیمایش کل گراف به کار گرفته می‌شود و در مدت زمان $\Theta(|V| + |E|)$ این کار را انجام می‌دهد که در اندازه گراف به صورت خطی می‌باشد. در این کارکردها، الگوریتم DFS همچنین از فضای $O(|V|)$ در بدترین حالت برای ذخیره سازی پشته بردارها در مسیر جستجوی کنونی و نیز مجموعه "بردارهای بیشتر مشاهده شده" استفاده می‌کند. به این ترتیب، در این شرایط، محدودیت‌های زمانی و فضایی همانند breadth-first search هستند و گزینش الگوریتم مورد استفاده کمتر بستگی به پیچیدگی آن‌ها دارد و بیشتر مبتنی بر ویژگی‌های مختلف ترتیب رئوس ایجاد شده توسط دو الگوریتم می‌باشد.

DFS را می‌توان برای جمع‌آوری نمونه‌ای از گره‌های گراف به کار گرفت. با این حال، DFS ناقص، همانند BFS ناقص، در جهت گره‌های دارای درجه بالاتر، انحراف دارد.

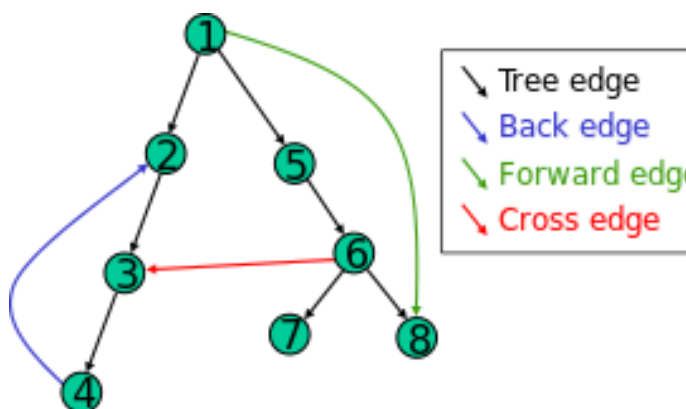
مثال

برای گراف زیر:



یک الگوریتم **depth-first search** که از A آغاز می‌شود، با فرض اینکه گوشه‌های سمت چپ در گراف نشان داده شده پیش از گوشه‌های سمت راست انتخاب شوند و با فرض اینکه جستجو، گره‌های بیشتر مشاهده شده را به یاد دارد، و آن‌ها را تکرار نمی‌کند، گره‌ها را با ترتیب زیر مشاهده خواهد کرد: A, B, D, F, E, C, G. گوشه‌های پیمایش شده در این جستجو یک درخت ترومو را تشکیل می‌دهند که ساختاری با کاربردهای مهم در تئوری گراف است.

یک نمایش ساده از یک جستجوی **depth first** در یک گراف، برحسب درخت هم‌پوشانی از رئوس است که در هنگام جستجو به وجود می‌آید. بر مبنای این درخت هم‌پوشانی، گوشه‌های گراف اصلی را می‌توان به سه گروه تقسیم کرد: گوشه‌های رو به جلو، که از گره یک درخت به سمت یکی از ولدهای آن حرکت می‌کند؛ گوشه‌های پشتی، که از یک گره به سمت یکی از اسلاف آن حرکت می‌کند، و گوشه‌های عرضی که هیچ یک از آن کارها را انجام نمی‌دهد. اگر گراف اصلی غیر جهت‌دار باشد، سپس همه گوشه‌های آن، گوشه‌های درخت یا گوشه‌های پشتی هستند.



شکل ۲۸: چهار نوع از گوشه‌ها که توسط یک درخت هم‌پوشانی تعریف شده است

الگوریتم هایی که از جستجوی depth-first search به عنوان ساختار تشکیل دهنده استفاده می کنند شامل موارد زیر

هستند:

- یافتن مؤلفه های مرتبط^{۸۰}
- رده بندی توپولوژیکی
- یافتن مؤلفه های مرتبط ۲- گوشه ای یا ۲- رأسی
- یافتن مؤلفه های مرتبط ۳- گوشه ای یا ۳- رأسی
- یافتن پل های یک گراف
- زایش کلمات به منظور ترسیم یک مجموعه حدی از یک گروه
- یافتن مولفه های قویاً مرتبط^{۸۱}
- آزمون حالت صفحه ای^{۸۲}
- حل معماهایی تنها با یک راه حل مانند مارپیچ ها. (DFS را می توان برای یافتن همه راه حل های یک مارپیچ از طریق در بر گرفتن گره ها در مسیر کنونی در مجموعه مشاهده شده، منطبق کرد)
- ایجاد مارپیچ هایی که می توانند از یک جستجوی رندوم depth-first استفاده کنند.
- یافتن ارتباط دوطرفه^{۸۳} در گراف ها

۳.۵.۳. Random walks

یک random walks عبارتست از فرمول ریاضی یک مسیر که متشکل از مراحل انجام است. برای مثال، مسیر دنبال شده توسط یک مولکول هنگام حرکت ویژگی گاز یا مایع، مسیر جستجوی یک حیوان به دنبال آذوقه، قیمت نوسانی سهام و وضعیت مالی یک قمارباز که می توانند به عنوان «پیاده روی رندوم» مدل سازی شوند، هرچند که ممکن است حقیقتاً رندوم نباشند. این عبارت ابتدا در سال ۱۹۰۵ توسط کار پیرسون معرفی شد و در حوزه های مختلف کاربرد دارد. انواع مختلفی از random walks وجود دارد و معمولاً آن ها را زنجیره های مارکوف یا فرایندهای مارکوف در نظر می گیرند. برخی از آن ها روی گراف ها، برخی روی خط، در صفحه، در ابعاد بالاتر و یا حتی در سطوح منحنی بوده و برخی از آن ها در گروه ها هستند. معمولاً آن ها نسبت به پارامتر زمان تغییر می کنند.

⁸⁰ Finding connected components

⁸¹ Finding strongly connected components

⁸² Planarity testing

⁸³ biconnectivity

این روش مبتنی است بر مفهوم آگاهانه مدولاریته شبکه. این روش امکان محاسبه پربازده رتبه‌بندی یک گوشه را در شبکه‌های بزرگ در زمان تقریباً خطی فراهم می‌آورد. پس از محاسبه رتبه‌بندی مرکزیت^{۸۴}، این الگوریتم به محاسبه مجاورت جفت به جفت بین گره‌های شبکه می‌پردازد؛ و در نهایت، ساختار جامعه را با اتخاذ یک استراتژی مدرن و شناخته شده روش Louvain با حداکثر سازی صحیح مدولاسیون شبکه کشف می‌کند. آزمایشات انجام شده نشان می‌دهند که این الگوریتم عملکرد بهتری نسبت به روش‌های دیگر دارد.

روش‌های پربازده مختلفی بر اساس اطلاعات محلی برای کشف ساختار جامعه در شبکه‌های بزرگ پیشنهاد شده‌اند و به واسطه هزینه‌های محاسباتی پایین، برای آنالیز شبکه‌های بزرگ مناسب هستند. نقص اصلی روش‌های موجود آن است که اطلاعات عمومی در مورد توپولوژی شبکه را مدنظر قرار نمی‌دهند. این استراتژی نوین معرفی شده دو مزیت دارد: اول اینکه از اطلاعات عمومی و محلی به شکل دوم آن استفاده می‌کند، و دوم اینکه به واسطه بهینه‌سازی، نتایج مناسبی را عرضه می‌کند.

۴. فصل چهارم : جمع بندی و نتیجه گیری

۴.۱. نتیجه گیری

مطالعه جوامع شبکه‌ای یا به عبارت دیگر مطالعه جامعه در شبکه‌های اجتماعی، از برخی جنبه‌ها بسیار قدیمی است و ریشه آن را می‌توان در جامعه‌شناسی، علوم کامپیوتر، آمار و رشته‌های دیگر جستجو کرد. با این حال، دامنه رو به گسترش شبکه‌های اجتماعی باعث افزایش توجه به این موضوع شده است. در این مقاله ما به بررسی روش‌های مدرن پرداختیم. این پیشرفت‌ها مشخص می‌کنند که چگونه می‌توان دانش و اطلاعات را در میان کاربران شبکه‌های اجتماعی تقسیم کرد. تحقیقات بیشتری در این حوزه با توجه به توان سازماندهی و علاقه تجاری به این دانش مورد نیاز است. روش‌های بیشتری باید ایجاد شوند و انتظار می‌رود که ابزارهای نرم افزاری مرتبط با آن‌ها بوجود بیایند.

⁸⁴ Centrality ranking

به طور کلی می‌توانیم این گونه جمع‌بندی کنیم که در این مقاله یک بازبینی نسبتاً جامع از الگوریتم‌های کشف جامعه به انجام رسید. هدف از این مقاله نیز همین بود: ارائه راهنمایی در خصوص روش‌های کشف جامعه در شبکه‌های اجتماعی. برای انجام این کار، ابتدا به تعریف جامعه پرداختیم و آنطور که از تنوع تعاریف برمی‌آید، هنوز یک تعریف کلی پذیرفته‌شده در این مورد وجود ندارد. همان طور که فورچوناتو^{۸۵} عنوان می‌کند، این عدم وجود یک چارچوب تئوری دارای پیامدهای مهمی نه تنها در خود وظیفه کشف جامعه است (به این گونه که اگر ما روی مفهوم جامعه توافق نظر نداشته باشیم، چگونه می‌توانیم آن را از یک شبکه استخراج کنیم؟)، بلکه همچنین روی جنبه‌های دیگر نیز اثرگذار است. ما روش‌های کشف جامعه را تحت دسته‌بندی‌های عمومی بررسی کردیم.

یکی از مسائلی که هنوز مورد بررسی قرار نگرفته است، مطالعه تداخل میان تعاریفی است که از جامعه ارائه شده است. ارتباطات بسیاری در میان تعاریف مختلف و الگوریتم‌های گوناگون وجود دارد. خوب است که یک نمایش دقیق گراف-مانند از این هم‌پوشانی‌ها ایجاد شود که گره‌های آن الگوریتم‌هایی باشند که دارای تعریف مشترک از نظر جامعه، ویژگی‌های ورودی/خروجی، برخی جوامع کیفی و یا یک روش اکتشاف فضای جستجو هستند. این شبکه پیچیده چند بعدی می‌تواند به منظور ایجاد یک درک روشن و دیدگاه دقیق در مورد مسئله کشف جامعه مورد استفاده قرار بگیرد.

یکی از مسائلی که باید در تحقیقات آینده مدنظر قرار بگیرد، کشف مهم‌ترین ویژگی‌های الگوریتم‌های گوناگون است. توجه زیادی به وجود حالت چند بعدی شده است که نه تنها به عنوان یک ورودی که باید پردازش شود، بلکه همچنین به عنوان یک راه‌حل، و همچنین برای استخراج جوامع چند بعدی حقیقی مدنظر قرار می‌گیرد. یک ویژگی جالب توجه دیگر نیز می‌تواند وجود سازماندهی‌های سلسله‌مراتبی و هم‌پوشانی ساختار جامعه به شکل هم‌زمان باشد زیرا این دو ویژگی، دیگر به عنوان ویژگی‌های منحصر به فرد شناخته نمی‌شوند.

⁸⁵ Fortunato

1. A Classification for Community Discovery Methods in Complex Networks, Michele Coscia, Fosca Giannottib, Dino Pedreschi, *Barabasi Lab, Northeastern University, Boston, USA*,
2. AN INTRODUCTION TO SOCIAL NETWORK DATA ANALYTICS, Charu C. Aggarwal, *IBM T. J. Watson Research Center*, 2011
3. COMMUNITY DISCOVERY IN SOCIAL NETWORKS: APPLICATIONS, METHODS AND EMERGING TRENDS, S. Parthasarathy, Y. Ruan, V. Satuluri, *The Ohio State University*, 2011
4. Survey on Social Community Detection , Michel Planté and Michel Crampes, Laboratoire de Genie Informatique et d'Ingenierie de Production (LGI2P), 25 Mar 2013
5. Community Detection and Mining in Social Media, Lei Tang and Huan Liu, Morgan & Claypool, September 2010
6. Measuring Similarity of Graph Nodes by Neighbor Matching , Mladen Nikolic, University of Belgrade
7. Pairwise Document Similarity in Large Collections with MapReduce, Tamer Elsayed, Jimmy Lin, and Douglas W. Oard, Human Language Technology Center of Excellence and UMIACS Laboratory for Computational Linguistics and Information Processing University of Maryland, June 2008
8. CiteSeer: An Automatic Citation Indexing System, C. Lee Giles, Kurt D. Bollacker, Steve Lawrence , NEC Research Institute, 4 Independence Way, Princeton, 1998
9. K-CORE DECOMPOSITION OF INTERNET GRAPHS: HIERARCHIES, SELF-SIMILARITY AND MEASUREMENT BIASES, Jos e Ignacio Alvarez-Hamelin, Luca Dall'Asta, Alain Barrat, Alessandro Vespignani, NETWORKS AND HETEROGENEOUS MEDIA, June 2008
10. Generalized Louvain method for community detection in large networks, Pasquale De Meo, Emilio Ferrarax, Giacomo Fiumara, Alessandro Proveti, Oxford-Man Institute, University of Oxford, UK
11. Stochastic blockmodel approximation of a graphon: Theory and consistent estimation , Edoardo M. Airoldi, Thiago B. Costa, Stanley H. Chan, Dept. Statistics, Harvard University, 8 Nov 2008
12. Recursion Pruning for the Apriori Algorithm, Christian Borgelt, School of Computer Science, Otto-von-Guericke-University of Magdeburg, Magdeburg, Germany
13. LS SETS, LAMBDA SETS AND OTHER COHESIVE SUBSETS , Stephen P. BORGATTI, Martin G. EVERETT, Paul R. SHIREY, Elsevier Science Publishers, Social Networks 12 (1990)

14. Dynamic Community Identification, Tanya Berger-Wolf, Chayant Tantipathananandh, and David Kempe, Springer Science+Business Media, 2010
15. Wikipedia contributors. "Net-centric." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 24 Nov. 2014. Web. 10 Feb. 2015.
16. Wikipedia contributors. "Breadth-first search." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 28 Jan. 2015. Web. 10 Feb. 2015.
17. Wikipedia contributors. "Depth-first search." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 5 Feb. 2015. Web. 10 Feb. 2015.
18. Wikipedia contributors. "Clique (graph theory)." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 5 Feb. 2015. Web. 10 Feb. 2015.

۱۹. شبکه های اجتماعی و روش های استخراج جوامع در آن ها، مهر آفرین آدمی دهکردی، محمد حسین ندیمی، رضا قائمی،

همایش ملی علوم و مهندسی کامپیوتر، ۱ و ۲ اسفند ۱۳۹۱

۴.۳. واژگان انگلیسی به فارسی

Module Epistemic communities	جوامع بینشی
Edge	لبه - گوشه
Pruning	هرس کردن
Sub – network	زیر - شبکه
Greedy	حریص
Relative frequency	تناوب نسبی
Dendogram	دندوگرام
Recursive	بازگشتی
Pass	گذر
Node similarity	شباهت گرهی
Bipartiteness	دو بخشی
Graph Laplacian	لاپلاسیان گراف
Edge-betweenness	حالت میانه-گوشه ها
Bipartiteness	دو بخشی
Backtracking	ردگیری معکوس
Centrality ranking	رتبه بندی مرکزی
Biconnectivity	دو - اتصالی

Abstract

Understanding the formation and evolution of communities is a long-standing research topic in sociology in part because of its fundamental connections with the studies of urban development,

Criminology, social marketing, and several other areas. With increasing popularity of online social network services like Facebook, the study of community structures assumes more significance. Identifying and detecting communities are not only of particular importance but have immediate applications. For instance, for effective online marketing, such as placing online ads or deploying viral marketing strategies [10], identifying communities in social network could often lead to more accurate targeting and better marketing results. Albeit online user profiles or other semantic information is helpful to discover user segments. On the whole we can say that recent years have seen that WWW is becoming a flourishing social media which enables individuals to easily share opinions, experiences and expertise at the push of a single button. With the pervasive usage of instant messaging systems and the fundamental shift in the ease of publishing content, social network researchers and graph theory researchers are now concerned with inferring community structures by analyzing the linkage patterns among individuals and web pages. Although the investigation of community structures has motivated many diverse algorithms, most of them are unsuitable for large-scale social networks because of the computational cost. Moreover, in addition to identify the possible community structures, how to define and explain the discovered communities is also significant in many practical scenarios.



Islamic Azad University

ISLAMIC AZAD UNIVERSITY

QAZVIN Branch

Survey of Community Discovery

By

Roza Shaffaf

Supervisor

Ph.D. Mohammad Saniee

Winter 2015