دومین همایش ملی مهندسی کامپیوتر، داده های حجیم و الگوریتم ها

مروری برالگوریتمهای انجمن یابی در تحلیل شبکه های اجتماعی

بهروزساعی[1*]،علیرضا نوروزی[2]

۱-دانشجوی کارشناسی ارشد فناوری اطلاعات ،گروه کامپیوتر،دانشگاه آزاد اسلامی شهر مجلسی

Email: behrozsaei1@yahoo.com

۲-استادیار،گروه کامپیوتر،دانشگاه آزاد اسلامی واحد شهرمجلسی

Email: Norouzi.arz@gmail.com

چکیده:

داده کاوی یکی از پشرفتهای اخیر در حوزه کامپیوتر برای اکتشاف عمقی داده هاست.داده کاوی اطلاعات حیاتی و مهمی را که برای برنامه ریزی های استراتژیک ،مورد نیاز است،آشکار میکند.برخی شبکه های اجتماعی (و اغلب شبکه های اجتماعی برخط) که خاصیت انجمنی (یعنی دارای انجمنهایی هستند) راازخود بروز میدهند،دارای ساختار سلسله مراتبی هستند. بدین معنا که انجمن های بزرگتر از یکمجموعه انجمن های کوچکتر تشکیل شده و این انجمن های کوچکتر خود از یک دسته انجمن های کوچکتر دیگرتشکیل شده است. شبکه های اجتماعی اغلب این خاصیت را از خود بروز می دهند**این مقاله** سعی دارد تعدادی از الگوریتمهای مهم در حوزه انجمن کاوی را که در کاوش شبکه های اجتماعی،بکار می روند،معرفی کند.الگوریتمهای تقسیم،ماجولاریتی،شعاع طیفی مقادیر ویژه،پویا،استنباط آماری و سایر روشها،در این مقاله به صورت کلی مرور و معرفی میشوند.

کلمات کلیدی:داده کاوی،شبکه های اجتماعی،الگوریتم های، تقسیم،ماجولاریتی،شعاع طیفی،مقادیر ویژه،پویا،استنباط آماری

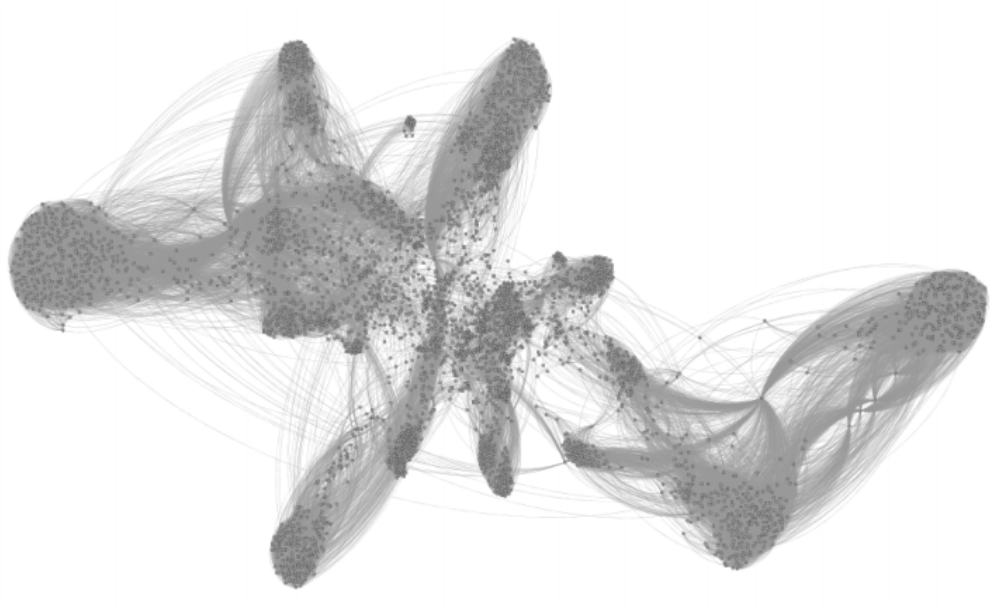
۱–مقدمه

یکی از سخت ترین کارهای مربوط به انجمن یابی در شبکه اجتماعی در همان گام نخست روی می دهد؛ سوال اساسی و پایه ایمربوط به تعریف انجمن هاستشکل ۱تصویر یک زیرگراف از گراف فیس بوک است که از ۴۰۳۹راس ۸۲۳۴یال شکیل شده می باشد. هرچقدر به لحاظ بصری یافتن انجمن های این گراف راحت به نظر می رسداما به لحاظ علم ریاضی و نظریه گراف ارائه تعریف و یافتن این انجمن ها پیچیده می نماید. در گراف شکل ۱توده های مجتمع راس ها، انجمن ها را تشکیل می دهند. شکل ۲یک گراف تصادفی است با تعداد راس ها ویال های.

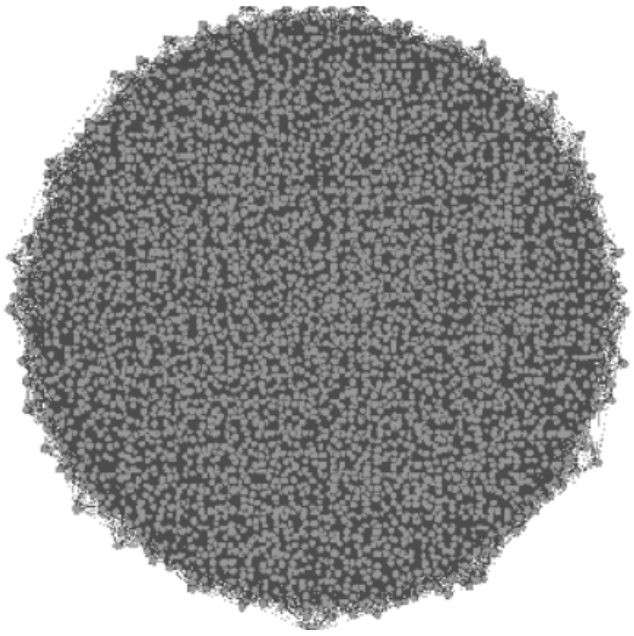
مساوی با گراف فیس بوک که برای هر یال دو راس را به طور تصادفی انتخاب کرده و به هم وصل می کندهمانطور که مشخص است. به لحاظ شهودی و تصویری گراف های مربوط به شبکه اجتماعی از گروه های متراکمزیادی تشکیل شده است اما گراف مربوط به گراف تصادفی تنها از یک توده تشکیل شده است. این دو گرافتفاوت های دیگری نیز دارند که در ادامه به آن خواهیم پرداخت، از جمله این تفاوت ها وجود چند راس (قطب[1])است که نقش ارتباطی را بازی می کنند.به صورت شهودی توزیع یال ها در گراف های واقعی همانند گراف فیس بوک [2]نه به

صورت سراسری که به صورتمحلی است. بدین معنی که تعداد یال های توزیع شده میان گروهی از راس ها بسیار بیشتر از تعداد یال های توزیعشده بین این گروه از راس ها با بقیه راس های گراف است. این ویژگی که در گراف های مربوط به داده های واقعیمی توان مشاهده کرد، انجمن خوانده می شوددر برخی از منابع انجمن، خوشه[۳] یا ماجول[۴] نیز خوانده می شود. بهعبارت دیگر انجمن ها مجموعه ای از راس هاست که با احتمال بیشتری نسبت به بقیه گراف ویژگی های مشترکیرا به اشتراک می گذارند. انجمن ها و انجمن یابی چندین سال است که به صورت گسترده مورد مطالعه قرارمی گیرد[۵،۶،۷،۸].در این پژوهش به گروه بندی که بین راس های گراف انجام می شود، تقسم بندی ،به هریک از این گروه ها در طول فرایند انجمن یابی، خوشه و پس از اتمام فرایند، انجمن می گویند. از کاربردهایانجمن یابی می توان به تبلیغات و بازاریابی اشاره کرد. از آنجایی که افراد حاضر در انجمن های تشکیل شده دریک شبکه اجتماعی به احتمال زیاد علایق مشترکی دارند، می توان با یافتن علایق آن ها از این اطلاعات به منظورتبلیغ محصولات خاص استفاده کرد. کاربردهای زیاد دیگری نیز می توان برای انجمن ها نام برد. همین کاربردفراوان در زمینه هایی همچون زیست شناسی، مهندسی کامپیوتر، اقتصاد این شاخه از علم شبکه های اجتماعی وتئوری گراف را به زمینه ای محبوب برای پژوهشگران جهت تحقیق تبدیل کرده است. برای مثال فرض کنید انجمنشامل دانش آموزان مربوط به یک مدرسه باشداین گروه از اعضای یک شبکه دارای ارتباطات بیشتری نسبت بهبقیه اعضای حاضر در یک شبکه اجتماعی هستند، حال در این مدرسه دانش آموزان مربوط به رده های تحصیلیمختلف ارتباطات بیشتری دارند. برای مثال دانش آموزان مربوط به کلاس های اول ارتباطات بیشتری باهم دارندو انجمن کوچکتری را تشکیل می دهنددر همین انجمن کوچکتر دانش آموزانی که در یک کلاس خاص (مثلااول یک) هستند نیز ارتباطات بیشتری با هم دارند و تشکیل انجمن کوچکتری را می دهند که انجمن بزرگتراجتماعی از این مجموعه هاست. برخی از الگوریتم های پیشنهاد شده برای انجمن یابی دقیقا به کاوش در اینساختار سلسله مراتبی می پردازند. هرچه گام های بیشتری اجرا شوند انجمن های کوچکتر را کشف می کنندبه طور خلاصه می توان گفت که هدف الگوریتم های انجمن یابی کشف ساختار انجمنی، و در صورت وجود یافتنویژگی سلسله مراتبی گراف ها تنها با اتکا به ساختار توپولوژیکی و اطلاعات ارائه شده توسط گراف است[۴]. ریشه های انجمن یابی را می توان در [۹]پی گرفت که در آن ویس [۵]و جاکوبسن [۶]برای یافتن کارگروه هاییک شرکت نخستین بار این مفهوم را به کار [۱]گرفتند. اهمیت این مقاله به این خاطر است که به معرفی برخی از مهمترین الگوریتمهای کشف انجمن در گراف شبکه اجتماعی می پردازد.

[1] Hub



شکل ۱:زیرگرافی ازگراف کامل فیسبوک با ۴۰۳۹ راس و۸۸۲۳۴ یال



شکل ۲:گراف تصادفی ۴۰۳۹ راس و۸۸۲۳۴ یال

به هر ساختار اجتماعی از افراد که بر اساس یک رابطه اجتماعی ایجاد می شود، یک شبکه اجتماعی می گوییم.
بنابراین هر شبکه اجتماعی شامل مجموعه ای از انسان ها و روابط اجتماعی بین آن هاست. لذا هر شبکه اجتماعیاز دو عنصر تشکیل شده است: موجودیت های شرکت کننده در ارتباط و ارتباط بین این موجودیت ها. درعلوم اجتماعی به موجودیت های شرکت کننده در ارتباط بازیگر و به ارتباطات بین این موجودیت ها رابطه گفتهمی شود .شبکه های اجتماعی به دو نوع برخط و برون خط تقسیم می شود. از شبکه های برون خط می توانبه شبکه دوستان، شبکه همکاران و شبکه همکلاسی ها اشاره کرد. از شبکه های برخط می توان به شبکه هایاجتماعی نظیر فیس بوک[۱]،توییتر[۲]و گوگل پلاس[۳] اشاره کرد.شبکه های اجتماعی از قرن نوزدهم مورد توجه قرار گرفت. پژوهش ها در این حوزه از دهه چهل به بعد باتعریف ابزارهایی چون گراف اجتماعی [۲]شتاب بیشتری گرفت. در سال 1994واسرمن [۳]با چاپ کتاب تحلیلشبکه های اجتماعی [۱]این زمینه از علم را وارد دوره جدیدی کرد، و پس از آن شبکه های اجتماعی به صورتجدی در زیرمجموعه های علوم اجتماعی و ریاضی مورد بررسی قرار گرفت.بحث های جسته و گریخته ای از سال 1960در مورد شبکه های اجتماعی برخط به راه افتاد. نخستین شبکهاجتماعی در سال 1997با نام سیکس دیگریز [۴]راه اندازی شد. اما انقلاب عظیم در هزاره دوم میلادی به وقوعپیوست جایی که از سال 2002به بعد شبکه هایی نظیر فرنداستر [۵]اورکات [۶]و لینکداین [۷]روی وب قرار گرفتند.پدیده بزرگ دیگر شبکه اجتماعی فیس بوک بود که در سال 2004توسط مارک زاکربرگ [۸]به دنیا معرفی شد.[۲]

۲–تجزیه و تحلیل شبکه های اجتماعی

تجزیه و تحلیل شبکه اجتماعی برای بررسی ساختار روابط اجتماعی یک گروه با هدف کشف ویژگی ها و روابط گروه یا افراد می باشد. همچنین اینطور هم می توان تعریف کرد ، تجزیه و تحلیل شبکه های اجتماعی به درک رابطه بین "بازیگران "می باشد که بازیگر(گره) می تواند یک فرد، یک سازمان، یک رویداد یا یک شی باشد. امروزه تجزیه و تحلیل شبکه های اجتماعی مورد مطالعه محققان رشته هایی مانند : جامعه شناسی، ارتباطات، علوم کامپیوتر، آموزش و پرورش ،اقتصاد ، جرم شناسی، علم مدیریت، پزشکی،علوم سیاسی و سایر رشته ها قرار گرفته است

۳–تعریف انجمن

همانطور که درمقدمه اشاره شد ، یکی از دشواری های مربوط به انجمن ها ارائه تعریفی برای آناست. تعریف انجمن ها نیازمند مقداری چشم پوشی و اختیار است، تا از دشواری های کار کاسته شود. از طرفیاصولا تعریف دقیق انجمن ها برای کاربردهای مختلف می تواند متفاوت باشد. اگر بین هر دو راس تابعی به عنوان تابع فاصله وجود داشته باشد، می توانیم انجمن ها را به عنوان مجموعه ای ازراس ها تعریف کنیم که به فاصله کمی از هم قرار دارند. این تعریف بیشتر در خوشه بندی داده ها استفاده می شود.
اما همچنان که پیش تر عنوان شد انجمن ها مجموعه ای از راس هاست که تعداد یال های داخل آن مجموعه (یعنییال هایی که هر دو راس انتهایی آن داخل انجمن باشند) از یال های آن مجموعه، با دیگر راس های گراف بهمراتب بیشتر باشد. این تعریف به عنوان پایه ای برای اغلب تعاریف دیگر به حساب می آید.

1 www.twitter.com
2 www.plus.google.com
3 Wasserman
4 SixDegree.com
5 Orkut
6 Linkdin
7 MarkZuckerberg

۴– الگوریتم های انجمن یابی

الگوریتم های ارائه شده برای انجمن یابی را می توان به شش گروه تقسیم کرد:

1.الگوریتم های تقسیم

2.الگوریتم های براساس ماجولاریتی

3.الگوریتم های براساس شعاع طیفی مقادیر ویژه

4.الگوریتم های پویا

5.الگوریتم های براساس استنباط های آماری[۳]

6.سایر روش ها[۴]

هر کدام از الگوریتم های ارائه شده برای انجمن یابی در یکی از این دسته ها قرار می گیرد. در ادامه این بخش بهتعریف این گروه ها خواهیم پرداخت و برای آن ها معروف ترین الگوریتم را شرح خواهیم داد

۴–۱ الگوریتم های تقسیم

فلسفه کلی این دسته از الگوریتم ها یافتن یال های بین انجمنی و حذف آن ها است. اگر تمام یال های بینانجمن ها را حذف کنیم، آن انجمن ها مولفه های جدا از هم را تشکیل می دهند معروف ترین و محبوب ترین الگوریتم که در دسته الگوریتم های تقسیم قرار می گیرد الگوریتم، گیروان نیومن [۳[]۱۰]است. این الگوریتم برای یافتن یال های بین انجمنی از مفهوم مرکزیت ارتباطی استفاده می کند. ایدهکلی این الگوریتم این است یال هایی که بین انجمن ها قرار می گیرند، دارای مقدار مرکزیت ارتباطی بزرگ تریهستند. این الگوریتم از چهار مرحله تشکیل می شود:

1محاسبه مرکزیت ارتباطی هر یال،

2.یال با بزرگترین مقدار مرکزیت ارتباطی را حذف می کنیم،

3.مرکزیت ارتباطی یال ها را دوباره محاسبه می کنیم

4.به مرحله دوم باز می گردیم

نخست این الگوریتم در مقاله [۱۰]معرفی شد. با توجه به اینکه زمان محاسبه مرکزیت ارتباطی برای هر یال درهر مرحله برابر($O(n^2)$برای ماتریس های تنک) پیچیدگی زمانی این الگوریتم برابر($O(n^3)$بود. در مقاله ای کهنیومن[۱] و گیروان[۲] در سال 2004 [8]دو سال پس از مقاله اول منتشر کردند؛ معیاری برای یافتن بهترین تقسیمارائه دادند. در نسخه اولیه تمام نمودار سلسله مراتبی الگوریتم بدست می آمد و با برش از یک سطح انجمن هامشخص می شد اما در الگوریتم اصلاح شده، تقسیمی که بزرگترین مقدار ماجولاریتی را ارائه می داد، به عنوانتقسیم مورد نظر مشخص می کردند.

این الگوریتم با وجود بهبودهای حاصل شده پیچیدگی زمانی زیادی داشت و مهمتر از همه نمی توانستانجمن های همپوشان را کشف کند.

[۱] Newman
[۲] Girvan

۲-۴ الگوریتم های بر اساس ماجولاریتی

هرچقدر مقدار ماجولاریتی یک تقسیم به مقدار ماجولاریتی ماکزیمم نزدیک تر شود، تقسیم حاصل تقسیمخوب تری خواهد بود. همین مساله ایده اصلی این دسته از الگوریتم ها هستند. الگوریتم های مربوط به این دستهبه چهار نوع زیر تقسیم می شوند

- الگوریتم تبرید شبیه سازی شده
- بهینه سازی خارجی
- بهینه سازی شعاع طیفی
- دیگر شیوه های بهینه سازی

الگوریتم های متفاوتی برای بیشینه سازی ماجولاریتی ارائه شده که به یکی از چهار دسته بالا تعلق دارند. اما از مهمترین الگوریتم های معرفی شده الگوریتم حریصانه ای است که توسط نیومن [۱۱]ارائه شد. الگوریتم با قرار دادن هر راس در یک انجمن مجزا کار خود را شروع می کند. در شروع کار هیچ یالی وجود ندارد، با افزودن یکبه یک یال ها انجمن هایی که در دو سر این یال قرار دارند در صورت افزایش ماجولاریتی تقسیم در هم ادغاممی شوند. ماجولاریتی تقسیم از روی گراف کامل محاسبه می شود یعنی گرافی که یال ها به آن اضافه می شود،تنها حکم نشانگر انجمن ها را دارد. اگر افزودن یک یال ادغامی در انجمن ها به وجود نیاورد، آن یال یک یالدرون انجمنی است. لذا مقدار ماجولاریتی را تغییر نخواهد داد. تعداد تقسیم های یافت شده در طول فرایند برابر nیعنی تعداد راس هاست. هرکدام از این تقسیم ها دارای یک مقدار ماجولاریتی هستند در نهایت بعد از افزودن یال ها تقسیمی که بزرگترین ماجولاریتی را دارد به عنوان خروجی مشخص می شودپیچیدگی زمانی این الگوریتم برای یک گراف تنک یعنی گراف هایی که تعداد یال های آن از تعداد کل یال های ممکن که برای nراس برابر $n(n-1)/2$ است بسیار کمتر(از مرتبه خطی) می باشد، مساوی $O(n^2)$است.

۳-۴ الگوریتم های براساس شعاع طیفی مقادیر ویژه

این الگوریتم ها از شعاع طیفی مقادیر ویژه ماتریس مربوط به ماتریس های مجاورت)یعنی ماتریس هایی که بهطریقی از ماتریس مجاورت استخراج می شوند، استفاده می کنند. نخستین تحقیقات بر روی شعاع طیفی خوشه هاتوسط دوناث [۱] و هوفمن[۲][۱۲]انجام شد. در این مقاله از بردار ویژه ماتریس مجاورت برای تقسیم بندیگراف استفاده شد. در همان سال فیدلر[۳] [۱۳]نشان داد که بردار ویژه دومین مقدار ویژه کوچک ماتریسلاپلاسین به احتمال زیاد تقسیمی ارائه می دهد که مینیمم برش این تقسیم اندازه بسیار کوچک تری دارد.گراف ساده Gرا با nیال داریم، ماتریس لاپلاسین آن $L = (l_{i;j})_{n\times n}$[۵]به صورت زیر تعریف می شود:

$$L = D - A$$

که در آن Dماتریس درجات راس هاست. برحسب کاربرد می توان ماتریس لاپلاسین را به صورت زیر نیز تعریفکرد

[۱] Donath
[۲] Hoffman
[۳] Feidler
[۴] Donetti
[۵] Munoz

$$l_{i,j} = \begin{cases} deg(v_i) & \text{اگر } i = j \\ -1 & \text{اگر } i \neq j \text{ and } v_i \text{ adjacent to } v_j \\ 0 & otherwise \end{cases}$$

تاکنون لاپلاسین پرکاربردترین ماتریس برای الگوریتم های شعاع طیفی بوده است[۴] . دونتی[۴] و مونز[۵][۱۴] روشی، رابرپایه بردار ویژه ماتریس لاپلاسین ارائه دادند. چون مقدار بردار ویژه مولفه ها برای راس هایی که دریک انجمن قرار دارند، مقدار نزدیکی به هم دارند، می توان با استفاده از بردارهای ویژه انجمن ها را کشف کرد. بدین معنی که اگر از mبردار ویژه استفاده کنیم، می توانیم راس ها را در یک فضای mبعدی قرار دهیم که انجمن ها به صورت راس هایی که در گروه های نزدیک به هم در این فضا قرار دارند مشخص می شوند. هرچقدرتعداد بردارهای ویژه به کار رفته بیشتر باشد، انجمن ها به صورت واضح تری در فضا مشخص می شود. الگوریتمارائه شده توسط دونتی و مونز شامل گروه بندی نقاط و استخراج تقسیم می باشد. دونتی و مونز از خوشه بندیسلسله مراتبی، با این محدودیت که تنها خوشه هایی که حداقل یک یال بین خوشه ای در گراف اصلی دارندبا هم ادغام می شوند، استفاده می کند. از بین همه تقسیم های استخراج شده، تقسیمی که بزرگ ترین مقدارماجولاریتی را دارد به عنوان خروجی مشخص می شود. پیچیدگی زمانی این روش برابر $O(n^3)$ می باشد.

۴-۵ الگوریتم های پویا

این دسته از الگوریتم ها بصورت مستقیم بر روی گراف کار می کنند و به این طریق انجمن ها را استخراج می کنند.در اینجا به تشریح الگوریتم عابر تصادفی که توسط ژو[۱][۱۵]ارائه شده است و در این دسته از الگوریتم هاطبقه بندی می شود، می پردازیم. عابر تصادفی که نخستین بار توسط هوگز[۲][۱۶]معرفی شد، عابری تصادفیاست که به صورت تصادفی روی گراف حرکت می کند و در هر راس با توجه به یال های موجود به هر کداماز راس های مجاور بصورت تصادفی می رود. ایده الگوریتم ژو این است که عابر تصادفی به علت چگالی زیادیال ها در داخل انجمن زمان بیشتری را داخل انجمن مصرف خواهد کرد. ژو از عابر تصادفی برای تعریف فاصلهبین دو راس استفاده کرد: فاصله dijبین دو راس iو jمیانگین تعداد یال هایی است که یک عابر تصادفیبرای رسیدن از jبه iباید از آن ها عبور کند. راس ها نزدیک به هم احتمالا به یک انجمن تعلق دارند و جذبکننده سراسری راس iرا به عنوان نزدیک ترین همسایه این راس تعریف می کند (راسی که کوچکترین مقدارdijرا داشته باشد). او همچنین جذب کننده محلی راس iرا به عنوان راسی تعریف می کند که iنزدیک ترینهمسایه آن باشد. دو نوع انجمن براساس جذب کننده سراسری و جذب کننده محلی وجود دارد: راس iبا تمامراس هایی که جذب کننده سراسری (محلی) هستند در یک انجمن قرار می گیرد و همه راس هایی که iجذبکننده سراسری(محلی) آن هستند نیز در انجمنی قرار می گیرند که iدر آن قرار دارد. انجمن زیرگراف کمینهاست، یعنی هیچ زیرگراف کوچک تری که شرایط مورد نظر در آن صدق کند وجود ندارد. پیچیدگی زمانی اینالگوریتم برابر $O(n^3)$است.

۴-۶ الگوریتم های براساس استنباط های آماری

استنباط های آماری به منظور استخراج ویژگی های مجموعه ای از داده ها با استفاده از یک دسته مشاهدات ومقایسه آن ها با مدل های فرضی

تولید شده، مورد استفاده قرار می گیرنداگر مجموعه داده یک گراف باشد،مدل مجموعه ای از راس هاست ء که به وسیله یال هایی به هم متصل می شوند، این مدل تولید شده با توپولوژیگراف اصلی منطبق است. این دسته از الگوریتم ها از این مدل ها استفاده کرده و انجمن های موجوددر گراف راپیش بینی می کند.استنباط بیزی یکی از روش های استنباط های آماری است که در مدل سازی گراف های واقعی مانند شبکه هایاجتماعی به کار می رود، در این بخش به معرفی الگوریتم هاستینگ[۱] [۱۷]می پردازیم که از روش استنباطبیزی استفاده می کند.استنباط بیزی از مشاهدات به منظور تخمین احتمال درستی یک فرض استفاده می کند. استنباط بیزی شاملدو جزء است: شواهد، که عبارت است از اطلاعات Dکه می توان از سیستم بدست آورد؛ و یک مدل آماری باپارامتر $\{\theta\}$ استنباط بیری از محاسبه احتمال $P(D|\{\theta\}$که برابر احتمال مشاهده شواهد در مدل مورد نظربا پارامتر $\{\theta\}$است، شروع می شود. هدف مشخص کردن مقدار $\{\theta\}$ای است که مقدار($P(D|\{\theta\}$ را حداکثرکند.در ارتباط با گراف، شواهد توسط ساختار گراف (ماتریس مجاورت یا ماتریس وزن) بدست می آید. در این موردیک جزء دیگر علاوه بر اجزاء ذکر شده وجود دارد و آن گروه بندی گراف ها به وسیله قرار دادن راس ها داخلگرو ه هاست. این گروه بندی اطلاعات پنهانی است که انتظار داریم از مدل انتخابی به وسیله پارامتر بدست آوریم .در تمام روش هایی که از استنباط بیزوی استفاده می کنند، هدف بیشینه کردن $P(D|\{\theta\}$می باشد که در آنمدل شامل ساختار گراف مشاهده شده، با مقداری محدویت اعمال شده می باشد. در ارتباط با گراف ها، پارامتر $\{\theta\}$ توسط سه گانه$(\{q\},\{\square\},\{k\})$ مشخص می شود که در آن $\{q\}$ انجمن هایی که توسط انتساب راس ها بهیال ها مشخص شده را نشان می دهد $\{\theta\}$ پارامتر مدل و kتعداد خوشه هاست.الگوریتم هاستینگ از مدلی با نام مدل طراحی شده استفاده می کند که در این مدل، nراس به qگروه منتسبمی شوند: راس هایی که در یک گروه قرار دارند با احتمال pinبا هم در ارتباط هستند (یالی بینشان قرار دارد)،در حالی که راس هایی متعلق به گروه های مجزا با احتمال $pout$با هم ارتباط دارند. اگر $pin > pout$گرافدارای انجمن می باشد. کلاس بندی گراف با مجموعه برچسب های $\{q_i\}$مشخص می شود. احتمال اینکه با توجهبه گراف داده شده کلاس بندی $\{q_i\}$ یک کلاس بندی مناسب، متناسب با مدل داده شده باشد برابر است با

$$p(\{q_i\}) \propto \{\exp[-\sum_{\langle ij\rangle} J\delta_{q_iq_j} - \sum_{i\neq j} J'\delta_{q_iq_j}/۲]\}^{-۱}$$

که در آن $J = \log\{[p_{in}(۱-p_{out})]/[p_{out}(۱-p_{in})]\}$ و $J' = \log\{[(۱-p_{in})]/[(۱-p_{out})]\}$

واولین مجموع روی نزدیک ترین همسایه ها اجرا می شود.

۷-۴ سایر روش ها

بقیه الگوریتم های ارائه شده برای یافتن انجمن ها در این دسته قرار می گیرند. از بین این الگوریتم ها، الگوریتمراقوان[۲] و همکاران[۱۸]را معرفی خواهیم کرد.راقوان و همکاران الگوریتم ساده و سریعی طراحی کردند که با نام الگوریتم گسترش برچسب ها شناخته می شود.راس ها در ابتدای کار با برچسب های منحصر به فردی از هم مجزا می شوند. در هر مرحله یک بار مورد بررسی قرارمی گیرند و برچسب شان با کمک برچسب راس های اطراف مشخص می شود؛ بدین ترتیب که هر راس برچسبخود را به برچسبی که بیشترین تعداد تکرار را در بین راس های

۱ Zhou-2Hughes
۱ Hastings
۲ Raghavan

همسایه دارد، تغییر می دهد. اگر چندینبرچسب با تعداد تکرار مساوی در همسایگی راسی قرار داشته باشد، یکی به تصادف انتخاب می شود. بدینوسیله برچسب ها در گراف گسترش می یابند: اغلب برچسب ها ناپدید می شوند و برخی دیگر به برچسب غالبگراف تبدیل می شوند. در مرحله ای که دیگر تغییری رخ نمی دهد، الگوریتم پایان می یابد. به صورت ساختاریهر راس همسایه های بیشتری را در بین راس های انجمنی که در آن قرار دارد، نسبت به راس های متعلق بهدیگر انجمن ها داراست. این الگوریتم احتمالا جواب های مختلفی را برای یک گراف در تکرار (الگوریتم) بدستمی دهد. الگوریتم برای هر مرحله زمان $O(m)$را مصرف می کند .

۵–نتیجه گیری

در این پژوهش تلاش کردیم الگوریتمهایی را که در بحث کاوش شبکه های اجتماعی به منظور جستجوی وجود انجمن ها بطور معمول وگسترده بکار میروند،معرفی کنیم. . پس ازمعرفی ماجولاریتی یک دسته از الگوریتم ها معرفی شدند که به کمک آن ها می توان انجمن های یک گراف رابدست آورد. بهینه سازی مقدار ماجولاریتی یکی از پایه های اصلی این الگوریتم هاست. کاربرد دوم مقایسه مقدارماجولاریتی بیشینه با مقدار ماجولاریتی انجمن حاصل از یک الگوریتم به منظور بررسی کارایی الگوریتم وکیفیت انجمن بدست آمده می باشد. ماجولاریتی مشکلاتی را به همراه دارد. یکی از مشکلات مربوط به بیشینه مقدار می باشد و اینکه این بیشینه مقدار تا چه اندازه می تواند مورد اعتماد باشد. اما مشکل بنیادی تر به این برمی گردد که ماجولاریتی تا چه اندازه قابلیت کشف و شناسایی انجمن های خوب رادارد این مشکل ماجولاریتی به خاطر تعریف مدل محض می باشد. نقطه ضعف مدل محض در این است کهراس ها با همه راس های دیگر قادر به تعامل هستند، بدین معنا که هر قسمت از گراف همه اطلاعات مربوط بهدیگر قسمت های گراف را در گستره دیدش دارد با وجود همه این موارد که به عنوان نواقص تابع کیفیت ماجولاریتیبرشمردیم تا کنون ماجولاریتی پرکاربردترینو بهترین تابع کیفیت ارائه شده می باشد.[4]

مراجع:

[1] Wasserman, S. (1994). Social network analysis: Methods and applications (Vol. 8). Cambridge university press.
[2] Moreno, J. L. (1953). Who shall survive? Foundations of sociometry, group psychotherapy and socio-drama.
[3] Newman, M. E., Girvan, M. (2004). Finding and evaluating community structure in networks. Physical review E, 69(2), 026113.

[4] Fortunato, S. (2010). Community detection in graphs. Physics Reports, 486(3), 75-174..
[5] Dorogovtsev, S. N., Mendes, J. F. F., Samukhin, A. N. (2000). Structure of growing networks with preferential linking.Physical Review letters, 85(21), 4633.
[6] Krapivsky, P. L., Redner, S., Leyvraz, F. (2000). Connectivity of growing random networks.Physical review letters, 85(21), 4629.
[7] Toivonen, R., Kovanen, L., Kivelä, M., Onnela, J. P., Saramäki, J., Kaski, K. (2009). A comparative study of social network models: Network evolution models and nodal attribute models. Social Networks, 31(4), 240-254.
[8] Coleman, J. S. (1964). Introduction to mathematical sociology.London Free Press Glencoe.
[9] Weiss, R. S., Jacobson, E. (1955). A method for the analysis of the structure of

complex organizations. American Sociological Review, 661-668.
[10] Girvan, M., Newman, M. E. (2002). Community structure in social and biological networks. Proceedings of the National Academy of Sciences, 99(12), 7821-7826.
[11] Newman, M. E. (2004). Fast algorithm for detecting community structure in networks.Physical review E, 69(6), 066133.
[12] Donath, W. E., Hoffman, A. J. (1973). Lower bounds for the partitioning of graphs. IBM Journal of Research and Development, 17(5), 420-425.
[13] Fiedler, M. (1973). Algebraic connectivity of graphs. Czechoslovak Mathematical Journal, 23(2), 298-305.
[14] Donetti, L., Munoz, M. A. (2004). Detecting network communities: a new systematic and efficient algorithm. Journal of Statistical Mechanics: Theory and Experiment, (10), P10012.
[15] Zhou, H. (2003). Distance, dissimilarity index, and network community structure. Physical review e, 67(6), 061901.
[16] Hughes, B. D. (1996). Random walks and random environments. Oxford: Clarendon Press.
[17] Hastings, M. B. (2006). Community detection as an inference problem.Physical Review E, 74(3), 035102.
[18] Raghavan, U. N., Albert, R., Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks.Physical Review E, 76(3), 036106.