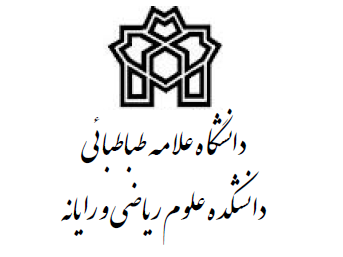
**پایان نامه کارشناسی ارشد علوم کامپیوتر**

**الگوریتم مبتنی بر اتاماتای سلولی یادگیرنده به منظور انجمن یابی در شبکه های پیچیده**

**استاد راهنما**

**دکتر محمدرضا اصغری اسکوئی**

**استاد مشاور**

**دکتر حسین تیموری فعال**

**استاد داور**

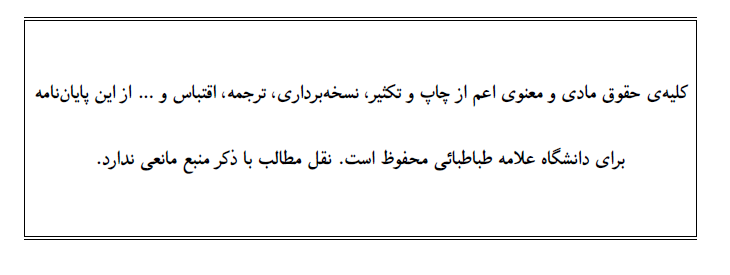
**دکتر محمد بحرانی**

**پژوهش گر**

**سمیه گلمحمدی**

**تابستان1398**

****

****

UniversityPicture

به نام خدا

**منشور اخلاق پژوهش**

با یاری از خداوند سبحان و اعتقاد به این که عالم محضر خداوند است و همواره ناظر به اعمال انسان و به منظور پاس داشت مقام بلند دانش و پژوهش و نظر به اهمیت جایگاه دانشگاه در اعتلای فرهنگ و تمدن بشری ما دانشجویان دانشکده های دانشگاه علامه طباطبائی متعهد می گردیم اصول زیر را در انجام فعالیت های پژوهشی مد نظر قرار داده و از آن تخطی نکنیم:

1. اصل حقیقت جوئی: تلاش در راستای پی جویی حقیقت و وفاداری به آن و دوری از هرگونه پنهان سازی حقیقت،
2. اصل رعایت حقوق: التزام به رعایت کامل حقوق پژوهشگران و پژوهیدگان (انسان، حیوان و نبات) و سایر صاحبان حق،
3. اصل مالکیت مادی و معنوی: تعهد به رعایت کامل حقوق مادی و معنوی دانشگاه و کلیه همکاران پژوهش،
4. اصل منافع ملی: تعهد به رعایت مصالح ملی و در نظر داشتن پیشبرد و توسعه کشور در کلیه مراحل پژوهش،
5. اصل رعایت انصاف و امانت: تعهد به اجتناب از هرگونه جانب داری غیر علمی و حفاظت از اموال، تجهیزات و منابع در اختیار،
6. اصل رازداری: تعهد به صیانت از اسرار و اطلاعات محرمانه افراد، سازمان هاو کشور و کلیه افراد و نهادهای مرتبط با تحقیق،
7. اصل احترام: تعهد به رعایت حریم ها و حرمت ها در انجام تحقیقات و رعایت جانب نقد و خودداری از هرگونه حرمت شکنی،
8. اصل ترویج: تعهد به رواج دانش و اشاعه نتایج تحقیقات و انتقال آن به همکاران علمی و دانشجویان به غیر از مواردی که منع قانونی دارد،
9. اصل برائت: التزام به برائت جوئی از هرگونه رفتار غیر حرفه ای و اعلام موضع نسبت به کسانی که حوزه علم و پژوهش را به شائبه های غیر علمی می آلایند.

نام و نام خانوادگی :**سمیه گلمحمدی**

تاریخ و امضاء:

UniversityPicture

**تعهدنامه ی اصالت پایان نامه / رساله**

اینجانب **سمیه گلمحمدی** دانش آموخته مقطع تحصیلی کارشناسی ارشد در رشته **علوم کامپیوتر** که در تاریخ ............................... از پایان نامه / رساله خود تحت عنوان **الگوریتم مبتنی بر اتاماتای سلولی یادگیرنده یه منظور انجمن یابی در شبکه های پیچیده** با کسب نمره / درجه ................... دفاع نموده ام، متعهد می شوم:

1. این پایان نامه / رساله حاصل تحقیق و پژوهش انجام شده توسط اینجانب بوده و درمواردی که از دستاوردهای علمی و پژوهشی دیگران (اعم از مقاله، کتاب، پایان نامه و غیره) استفاده نموده ام، مطابق ضوابط ورویه موجود، نام منبع مورد استفاده و سایر مشخصات آن را در فهرست مربوط ذکر و درج کرده ام.
2. این پایان نامه / رساله قبلا برای دریافت هیچ مدرک تحصیلی (هم سطح، پایین تر یا بالاتر) در سایر دانشگاه ها و موسسات آموزش عالی ارائه نشده است.
3. چنانچه بعد از فراغت از تحصیل، قصد استفاده از هرگونه بهره برداری اعم از چاپ کتاب، ثبت اختراع و ازین دست موارد از این پایان نامه / رساله را داشته باشم، از حوزه معاونت پژوهشی دانشگاه علامه طباطبائی مجوزهای مربوطه را اخذ نمایم.
4. چنانچه در هر مقطع زمانی خلاف موارد فوق ثابت شود، عواقب ناشی از آن را می پذیرم و دانشگاهی مجاز است با اینجانب مطابق ضوابط و مقررات رفتار نموده و درصورت ابطال مدرک تحصیلی ام هیچ گونه ادعائی نخواهم داشت.

نام و نام خانوادگی :**سمیه گلمحمدی**

تاریخ و امضاء:

**UniversityPicture**

**صورتجلسه دفاع**

با تأييدات خداوند متعال جلسه دفاع از پايان‏نامه کارشناسي‏ارشد دانشجو **سمیه گلمحمدی** رشته **علوم کامپیوتر** با شماره دانشجویی9613139111تحتعنوان **الگوریتم مبتنی بر اتاماتای سلولی یادگیرنده یه منظور انجمن یابی در شبکه های پیچیده** با حضور استاد راهنما، استاد مشاور و استاد داور در دانشگاه علامه طباطبائی در تاريخ **.......................** تشکيل گرديد. در اين جلسه، پايان‏نامه / رساله با موفقيت مورد دفاع قرار گرفت.

نامبرده نمره / درجه -------- (بدون ارائه مقاله) دريافت نمود.

نمره مقاله ----------- + نمره پايان‎نامه ------------ = نمره نهايي

استاد راهنما

نام و نام خانوادگی **دکتر محمدرضا اصغری اسکوئی** تاریخ و امضا

استاد مشاور

نام و نام خانوادگی **دکتر حسین تیموری فعال** تاریخ و امضا

استاد داور

نام و نام خانوادگی **دکتر محمد بحرانی** تاریخ و امضا

نماینده تحصیلات تکمیلی

نام و نام خانوادگی تاریخ و امضا

**فهرست مطالب**

**فهرست جدول ها**

**فهرست شکل ها**

1. **کلیات**
2. **تعاریف و مفاهیم پایه ای**
   1. **شبکه های اجتماعی . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .**

**2.1.1 انواع شبکه های اجتماعی. . . . . . . . . . . . . . .. . . . . . . . . . . . . . . . . . . . .**

**2.1.2 ویژگی های شبکه های اجتماعی. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .**

**2.1.3 نمایش شبکه های اجتماعی. . . . . . . . . . . . . . . . . . .. . . . . . . . . . . . . . . . .**

* 1. **گراف. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .**
     1. **انواع متخلف گراف. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . .**
     2. **زیرگراف. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .**
  2. **انجمن. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .**
  3. **روش های انجمن یابی. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .**
     1. **الگوریتم های تقسیم. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .**
     2. **الگوریتم های براساس بهینه سازی ماژولاریتی. . . . . . . . . . . . . . . . . . . . . .**
     3. **الگوریتم های براساس شعاع طیفی مقادیر ویژه. . . . . . . . . . . . . . . . . . . . . .**
     4. **الگوریتم های پویا. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .**
     5. **الگوریتم های بر اساس استنباط آماری. .. . . . . . . . . . . . . . . . . . . . . . . . .**
     6. **سایر روش ها. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .**
  4. **مروری بر روش های انجمن یابی بر اساس معیار ماژولاریتی. . . . . . . . . . . . . . . . . . . . . .**

1. **الگوریتم انجمن یابی با استفاده از اتاماتای سلولی یادگیرنده**
2. **پیاده سازی الگوریتم انجمن یابی با اتاماتای سلولی یادگیرنده و ارزیابی روش**
3. **جمع بندی و نتیجه گیری**

**مرجع ها**

**واژه نامه فارسی به انگلیسی**

**فصل دوم**

**تعاریف و مفاهیم پایه**

برای ارایه یک الگوریتم جدید انجمن یابی در یک شبکه لازم است ابتدا با مفاهیم این مبحث اشنا شد. به همین منظور در این فصل ابتدا ارتباط بین شبکه و گراف و ماتریس مجاورت بررسی می‌شود. سپس مفاهیم زیر گراف، انجمن و معیارهای اندازه گیری کیفیت انجمن ها یعنی ماژولاریتی[[1]](#footnote-1) و اطلاعات مشترک نرمالسازی شده [[2]](#footnote-2)معرفی شده اند. سرانجام روش های مختلفی که برای حل مساله انجمن یابی مطرح شده اند معرفی و رویکرد های هرکدام مورد بررسی قرار میگیرد.

**1.2 شبکه های اجتماعی**

ساختارهایی که شامل مجموعه ای از افراد و بر اساس رابطه اجتماعی تشکیل شود ، یک شبکه اجتماعی است.هر شبکه اجتماعی شامل دو جز می باشد.موجودیت ها یا همان افراد و رابطه میان آنها. در علوم اجتماعی به مجموعه موجودیت ها بازیگران و به مجموعه دوم ، رابطه گفته می شود. (Wasserman, 1994) با تعمیم تعریف شبکه اجتماعی به انواع شبکه ها می توان این گونه برداشت کرد که شبکه ها مجموعه ای از نه تنها افراد، بلکه اشیا یا هرچیزی که بتوانند باهم رابطه داشته باشند می باشد و رابطه های آنهاست. شبکه ها انواع مختلفی دارند ولی ساختار یکسان تمام انواع شبکه باعث ایجاد مفهمومی به نام توپولوژی شبکه شده است.که این مفهموم چگونگی چیدمان موجودیت ها و همچنین اتصالات بین هرکدام از آنها را نشان می دهد. مفهموم و اصطلاح شبکه های اجتماعی ، توسط جان بارنز[[3]](#footnote-3) در سال 1954 مطرح شد و بعد از آن بود که مطالعه علمی شبکه ها آغاز شد . (MCCRACKEN, 1968)او تحقیقی در مورد گروه‌های اجتماعی در بخشی از نروژ انجام می‌داد و اصطلاح شبکه اجتماعی را در آن تحقیق برای توصیف رابطه بین انسانها و تحلیل مکانیزم‌های ارتباطی و خصوصاً [تصمیم گیری](https://motamem.org/?p=5526) آنها به کار برد. در قرن نوزدهم بود که شبکه های اجتماعی مورد توجه قرار گرفتند. بعد از دهه چهل با تعریف گراف اجتماعی (Moreno, 1953) پژوهش در این زمینه طرفداران بیشتری پیدا کرد. در سال 1994 واسرمن[[4]](#footnote-4) کتاب تحلیل شبکه های اجتماعی (Wasserman, 1994) خود را چاپ کرد و علوم بررسی شبکه های اجتماعی وارد عصر جدیدی شد. چرا که بعد از آن بررسی و تحلیل شبکه های اجتماعی در علوم اجتماعی و ریاضی بصورت جدی مورد توجه قرار گرفت.منظور از تجزیه و تحلیل شبکه های اجتماعی ، روابطی است که به کمک علم گراف می توانند مدل شوند.از آنجایی که روابط بین گره ها می توانند روابط مختلفی باشند پس ساختار های مبتنی بر گرافی که برای یک شبکه یکتا می توان یافت بسیار متنوع و پیچیده است. (Zhang, 2007)

**1.1.2 انواع شبکه های اجتماعی**

شبکه های اجتماعی مختلفی وجود دارند .تعدادی از معروف ترین و پرکاربرد ترین شبکه های اجتماعی در زیر معرفی شده اند:

**شبکه دوستی**: در این نوع شبکه مجموعه ای از انسان ها و دوستی بین آنها وجود دارد.هر گره نشان دهنده ی یک فرد و وجود یال بین آن دو فرد نشان دهنده ی دوستی و ارتباط انهاست. این شبکه معمولا بدون وزن است ولی می توانیم از گراف وزن دار هم برای نشان دادن میزان دوستی هر دو نفر استفاده کنیم.

**شبکه تلفنی**: یکی از مثال های دیگر برای شبکه اجتماعی ،شبکه تلفی است به این صورت که هر شماره تلفن را به صورت یک یال در نظر گرفت و هر تماسی که در بازه مشخص بین دو شماره برقرار شده است ، با یک یال نمایش داده شود.

**شبکه ایمیلی**: با در نظر گرفتن هر آدرس ایمیل به عنوان یک گره، و نشان دادن ارسال پیام بین دو ادرس با یال، می توان شبکه ایمیلی را تشکیل داد.این نوع شبکه می تواند جهت دار یا غیر جهت دار باشد.

شبکه همکاری: چنان چه گره ها را افرادی در نظر بگیریم که با یکدیگر همکاری دارند.وتعداد کارهای مشترک بین آنها را با یال نشان دهیم ، یک شبکه همکاری داریم.برای مثال اگر هردو نفری که باهم مقاله ای منتشر کرده اند را با یال به هم وصل کنیم و وزن یال تعداد مقالات را نشان دهد در این صورت این عدد بیانگر میزان همکاری هر دو نویسنده است.

**2.1.2 ویژگی های شبکه های اجتماعی**

بعد از بررسی ویژگی های توپولوژیکی شبکه ، نتایج زیر بدست آمده است:

* تعداد یالها در شبکه های اجتماعی تنک[[5]](#footnote-5) وشامل اجزای بزرگی است که به هم وصل شده اند.
* انجمن ها در شبکه معمولا بزرگ و تعداد انجمن های کوچک بسیار کم است.
* انجمن های حاظر در یک شبکه ممکن است ویژگی متفاوتی نسبت به دیگر انجمن های شبکه داشته باشند.
* شبکه های اجتماعی بیازتابی از روابط در دنیای واقعی هستند.
* توزیع درجه در شبکه های اجتماعی اصلا همگن نیست.
* شبکه های اجتماعی خاصیت آزادی از مقیاس[[6]](#footnote-6) دارند. به این صورت که تعداد اندکی از افراد می توانند ارتباط زیادی با بقیه گره های شبکه داشته باشند. (Takes, 2011.)

**3.1.2 نمایش شبکه های اجتماعی**

دو ابزار اصلی نمایش شبکه های اجتماعی گراف و ماتریس هستند.در ادامه به بررسی این دو ابزار می پردازیم:

**گراف**: در نمایش گرافی ، افراد به عنوان یک گره نمایش و ارتباط بین آنها، با یک یال بین آنها نمایش داده می‌شود. انواع مختلف گراف این امکان را برای ما فراهم می کند که متناسب با نوع شبکه از گرافها استفاده کنیم.گراف جهت دار برای شبکه هایی مناسب است که یک طرفه هستند.گراف وزن دار برای زمانی مناسب است که مقدار و شدت یا استحکام ارتباط می تواند متفاوت باشد .گراف ساده ،برای شبکه های دوطرفه و بدون وزن قابل استفاده است.این نوع گراف بهترین و مناسب ترین ساختار جهت توصیف شبکه اجتماعی است به همین دلیل الگوریتم انجمن یابی که پیشتر در مورد ان توضیح داده می شود برروی این نوع نمایش تمرکز دارد.

برای هر دو گره ،ارتباط دو گره یعنی یال را با نماد نمایش می دهیم. در گراف غیر جهت دار، و برابر است اما در گراف های جهتدار این مقدار می تواند متفاوت باشد. تعریف دقیقتر در ادامه به تفصیل بیان می شود و اینجا صرفا منظور تفاوت بین گراف جهت دار و غیر جهتدار است. در گراف ساده هر سال باخودش هم در اتباط نیست واین یعنی مقدار است. گراف ساده خاصیت چندگانگی هم ندارد به این صورت که دو گره یا باهم در ارتباط هستند یا نیستند.پس یالهای متعددی از یک گره به گره دیگری وجود ندارد.

مثال: در شبکه اجتماعی دوستی که پیشتر مطرح شد،ارتباط بین افراد رابطه دوستی بود. شکل زیر یک مثال ساده از نمایش یک شبکه اجتماعی شامل7 نفر و ارتباط دوستی میان آنهاست.

اطلاعات زیادی توسط نمایش گرافی قابل مشاهده است.

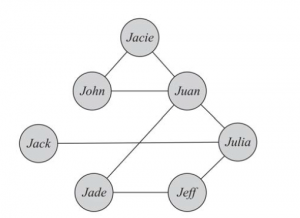
* هر فرد با چه کسانی دوست است؟
* دوستان مشترک دو نفر چه کسانی هستند؟
* ...

Figure 1 نمایش شبکه اجتماعی با کمک ساختار گراف

**ماتریس مجاورت:** نمایش رایج دیگر برای نشان دادن ساختار توپولوژیکی شبکه ها،ماتریس مجاورت آن ها است. شبکه های اجتماعی را می توان به صورت آرایش مربعی از درایه هانشان داد. در این حالت در ماتریس مورد نظر ما هر سطر متعلق به یک راس و هر ستون متعلق به یک راس می باشد. ماتریس مجاورت یک شبکه ساده یعنی غیر جهتدار ، یک ماتریس متقارن با عناصر نامنفی است. اگر تعداد راس یا گره های شبکه برابر با N باشد در اینصورت ارتباط هر زوج گره به صورت عدد نامنفی است که مقدار صفر برای این عدد به معنی عدم وجود یال، مقدار یک به معنی وجود یال و مقدار های بزرگتر از یک نشان دهنده وزن گراف هستند. برای گراف های ساده ، وزن تمامی یال های موجود برابر یک است.

**2.2 مبانی نظری گراف**

**2.2.1 گراف و انواع گراف**

همانطور که اشاره شد گراف یک ابزار مهم و مناسب برای مدل کردن شبکه های اجتماعی است. گراف را با دوتایی  *در نظر می گیریم به این صورت که نشان دهنده مجموعه گره ها یا راس ها و نشان دهنده مجموعه یالهاست. تعداد گره های گراف را با و تعداد یال‏ها را یا مشخص میکنیم.یک گراف می تواند جهت دار،غیر جهت دار ،وزن دار و یا بدون وزن باشد.*

**گراف جهت دار:** در یک گراف جهتدار ، یال به این معنی است که مسیری از راس

به راس وجود دارد و ممکن است عکس آن همواره برقرار نباشد. در گراف غیر جهتدار اگر موجود باشد می توان نتیجه گرفت که هم وجود دارد. یعنی به عبارت دیگر

**2.2.2زیرگراف**: زیرگراف Cخود یک گراف است که مجموعه راس‌های آن زیر مجموعه ای از مجموعه راس‌های گراف بزرگ‌تری باشد، واضح است که در این‌صورت یال‌های زیرگراف هم زیرمجموعه‌ای از یال‌های گراف بزرگ‌تر خواهد بود. تبدیل گراف به زیر گراف‌های معنا دار در حقیقت تلاش برای شناخت بهتر شبکه است و اطلاعات بسیاری در مورد ماهیت آن در اختیار ما قرار می دهد.

**3.2** ا**نجمن**[[7]](#footnote-7)

انجمن یکی از ویژگی‌های مهم شبکه‌های بزرگ انجمن‌ها است. انجمن‌یابی در یک شبکه کاربردهای زیادی دارد برای مثال از آن‏جایی که افراد حاضر در انجمن‌هاي تشکیل شده در یک شبکه اجتماعی به احتمال زیاد علایق مشترکی دارند، می‌توان با یافتن علایق آن‌ها از این اطلاعات در مسایل مربوط به تبلیغات و بازاریابی استفاده کرد .مثال دیگری از این کاربرد مربوط به انتشار اخبار است، اعضای یک انجمن باهم در ارتباط هستند لذا برای انتشار خبر یا تبلیغات می‌توان آن را برای اعضایی ارسال کرد که در یک انجمن نباشند بدین ترتیب هرکدام خبر را در انجمن خود انتشار داده و بجای ارسال آن به تمام اعضا ، به تعداد انجمن‌های موجود در شبکه خبر را ارسال و در هزینه‌های مربوطه صرفه‌جویی می شود. یک انجمن می‌تواند به‌طور‌کلی به‌عنوان مجموعه‌ای از راس‌ها که چگالی بالایی در ارتباط با زیر‌گراف خود (ارتباط داخلی) و ارتباط کمی با سایر زیرگراف‌ها دارند، توصیف شود. این تعریف از جهت‌هایی مبهم است، هنگام برخورد با مساله شناسایی انجمن‌ها ما باید تعریف دقیق و روشن‌تری از مفهوم انجمن داشته باشیم .لذا اینجا چند تعریف دقیق‌تر از آن را ارایه می‌دهیم که در حوزه‌ی شناسایی انجمن‌ها پذیرفته شده اند.

همانطور که پیش‏تر مطرح شد، نمایش ریاضی یک گراف به صورت است.در این تعریف گفتیم مجموعه ای شامل راس ها یا گره ها، زیرمجموعه های دوتایی از گره های گراف است که به آنها یال گفته میشود.

راداچی[[8]](#footnote-8) برای تعریف انجمن ، از دومفهموم محلی استفاده کرد به این‏صورت که یکی از آن‏ها مفهوم قوی بودن و دیگری مفهوم ضعیف بودن انجمن را تشریح کرد . (F. Radicchi, 2004) برای هر راس از زیرگراف مفهوم( درجه راس) به شکل زیر است . درجه راس را نمایش می دهد که مجموع دو مقدار زیر است ، تعداد یال‌هایی که از راس به زیرگراف C وصل است و یال‌هایی از راس که به زیرگراف C وصل نیستند. (به سایر زیر گراف‌ها وصل شده اند.)

یک انجمن قوی زیرگرافی است که تمام گره‌های انجمن ارتباط بیشتری با زیرگراف C (نسبت به سایر زیرگراف ها)دارند.

یک انجمن ضعیف زیرگرافی است که جمع ارتباط‌هایی که گره‌های این گروه با زیرگراف C دارند بزرگتر باشد از جمع ارتباط‌ هایی که گره‌ها با زیرگراف C ندارند.

انجمن قوی، انجمن ضعیف هم هست. ولی عکس قضیه همواره برقرار نیست. در ادامه تعریف دیگری برای انجمن‌ها توسط رغوان[[9]](#footnote-9) ارایه شده است. (U. N. Raghavan, (2007)036106.) اگر Ωمجموعه شامل تمام انجمن‌های گراف باشد .پس |Ω|تعداد انجمن‌های گراف ‌را نشان میدهد.کل درجه‌های هرگره نهایتا به |Ω|بخش تقسیم میشود ، یعنی درجه راس برابر است با یال‌هایی که از آن به هر کدام از انجمن‌ها وصل است.

که درجه تعلق گره به انجمن Cبرابر است با یال هایی که از آن گره به انجمن C وصل است.

پس برای انجمن C این تعریف به شکل زیر است:

یعنی تمام گره‌ها ارتباط بیشتری (یا مساوی) با انجمنی دارند ،که متعلق به آن هستند . زمانی‌که فقط دوتا انجمن داریم این تعریف همان تعریف انجمن قوی است که پیشتر ارایه شد، اما زمانی که گراف بیشتر از دو تا انجمن دارد محدودیت رغوان ضعیف تر از انجمن قوی بودن است.

**روش های انجمن‏یابی انواع4.2**

همانطور که بیان شد ،انجمن یابی اطلاعات زیادی برای تحلیل شبکه و دریافت اطلاعات جدید از روی ساختار شبکه ها در اختیار ما قرار می دهد.لذا مقالات زیادی به دنبال یافتن راه حل های جدید برای این کار بوده اند.

طبق یک تقسیم بندی کلی می توان انواع روش های انجمن یابی را به شش دسته تقسیم کرد

1-الگوریتم های تقسیم

2-الگوریتم های برمبنای بهینه سازی معیار ماژولاریتی

3-الگوریتم های بر اساس شعاع طیفی مقادیر ویژه

4-الگوریتم های پویا

5-الگوریتم های بر اساس استنباط های آماری

6-سایر روش ها

تمام انواع الگوریتم های ارائه شده برای انجمن یابی در یکی از این گروه ها قرار می گیرند. در ادامه این بخش به تعریف این گروه ها پرداخته شده و معروفترین مثال های هر بخش توضیح داده شده است.

**1.4.2 الگوریتم های تقسیم**

وجه مشترک تمام این الگوریتم ها در تلاش آنها برای یافتن یالهای بین انجمنی و حذف آنهاست.چرا که اگر تمام یال های بین انجمن ها را پیدا وآنها را حذف کنیم دراین صورت ،انجمن های جدا از هم خواهیم داشت. معروفترین الگوریتمی که در این دسته قرار می گیرد،الگوریتم گیروان نیومن در مقاله های (Newman M. G., (2004).) و (Girvan, (2002).) است.این الگوریتم از مفهمومی با نام مرکزیت ارتباطی استفاده می کندتا یالهای بین انجمنی را یافته وآنها را حذف کند چرا که این یال ها دارای مرکزیت ارتباطی بزرگتری هستند. ابتدا مفهموم مرکزیت ارتباطی تعریف و سپس الگوریتم تشریح می گردد.

مرکزیت ارتباطی:این مفهموم برای یک یال تعریف می شود و عبارت است از تعداد کوتاهترین مسیر های بین هردو راس در گراف شبکه که از این یال عبور میکنند.اگرتعداد کوتاهترین مسیر بین دوراس بیشتراز یکی باشد برای هر مسیر وزنی در نظر گرفته می شو به طوریکه جمع وزن ها برابر یک باشد. این معیار معیاری برای نشان دادن اهمیت یک یال در گراف است به این ترتیب که هر قدر یال به عنوان قسمتی از کوتاه ترین مسیر گراف بیشتر استفاده شود،نقش مهمتری را در گراف بازی میکند گیروان و نیومن نشان دادند هرقدر یک یال ضریب مرکزیت ارتباطی بیشتری داشته باشد به احتمال زیادتری آن یال پلی بین دوتا از انجمن هاست و بنابراین با حذف ان شانس رسیدن به ساختار مطلوب بیشتر می شود.

این الگوریتم از چهار مرحله تشکیل شده است.

1. محاسبه مرکزیت ارتباطی برای تمام یالها
2. یال با بزرگترین مقدار برای ویژگی مرکزیت ارتباطی را حذف می کنیم.
3. ویژگی مرکزیت ارتباطی برای تمام یالها دوباره محاسبه می شود.
4. مرحله دوم و سوم مجددا اعمال می شود.

این الگوریتم در مقاله (Girvan, (2002).) معرفی شد.اما برای محاسبه مرکزیت ارتباط تمام یالها در هر مرحله و لذا برای کل الگوریتم برابر است.دو سال بعد در مقاله ی (Newman M. G., (2004).) که چاپ شد، معیاری جدیدی برای یافتن بهترین ساختار و تقسیم بندی شبکه ارائه کردند.در الگوریتم اول نمودار سلسه مراتبی کامل محاسبه و بابرش از یک سطح انجمن ها مشخص می شد ، اما در الگوریتم جدید تقسیم بندیی که بزرگترین مقدار ماژولاریتی را نتیجه می داد به عنوان تقسیم بندی بهینه شناخته می شد.این الگوریتم جدید با وجود اینکه نسبت به الگوریتم پیشین بهبود های زیادی یافته بود،هنوزهم پیچیدگی زمانی زیادی داشت و نمی توانست انجمن هایی ک همپوشانی دارند را شناسایی کند.

**2.4.2 الگوریتم های برمبنای بهینه سازی معیار ماژولاریتی**

مقدار ماژولاریتی برای بهترین ساختار انجمن هایش را مقدار ماکسیممی فرض می کنیم و هرقدر الگوریتم بتواند مقدار ماژولاریتی که پیدا می کند به مقدار ماکسیمم ماژولاریتی نزدیکتر باشد،تقسیم حاصل بهتر ولذا عملکرد الگوریتم مناسب تر است.این مساله به عنوان ایده اصلی تمام الگوریتم های این دسته مطرح می شود.

خود این الگوریتم ها چهار دسته هستند.

1-تبرید شبیه سازی شده

2-بهینه سازی خارجی

3-بهینه سازی شعاع طیفی

4-دیگر شیوه های بهینه سازی

الگوریتم های متفاوتی که برای بیشینه سازی ماژولاریتی رائه شده اند ،در یکی از این گروه ها جای میگیرند.اما مهمترین آنها الگوریتم حریصانه ای است که نیومن در مقاله ارائه داد (Newman, 2004)

این الگوریتم به این صورت عمل می کند که در ابتدا هر راس را در یک انجمن مجزا قرار می دهد.وکار خود را شروع میکند.در ابتدا هیچ یالی وجود ندارد ،سپس با اضافه کردن یالهای بین راس ها،انجمن ها کم کم به گونه ای در هم ادغام می شوند که در نهایت ماژولاریتی برای کل ساختار بیشینه شود.

گرافی که یالها به آن اضافه می شودتنها حکم نشانگر انجمن ها را دارد.به اینصورت که اگر اضافه کردن یک یال باعث ادغام حداقل دو انجمن نشود در این صورت این یال درون ساختار انجمن است و یال بین انجمنی نیست.

این یالها تغییری در مقدار ماژولاریتی پدید نمی آورند. تعداد ساختار هایی که الگوریتم در طول فرایند پیدا می کند برابر nیعنی تعداد راس های گراف است که هرکدام دارای یک ماژولاریتی خاص هستند.در نهایت بعد از اضافه کردن یالها،ساختاری که بزرگترین ماژولاریتی را به ما بدهد به عنوان خروجی الگوریتم انجمن یابی در نظر گرفته می شود.پیچیدگی زمانی این الگوریتم برای یک گراف تنک یا خلوت برابر است.

**3.4.2 الگوریتم های بر اساس شعاع طیفی مقادیر ویژه**

در این روش هاابتدا ماتریس های شعاع طیفی مقدارهای ویژه مربوط به ماتریس مجاورت گراف را استخراج و سپس از آنها استفاده می کنند. اولین باری که از شعاع طیفی خوشه ها استفاده شد در مقاله ی (Donath, 1973) نوشته دوناث [[10]](#footnote-10)و هوفمن[[11]](#footnote-11) بود.در این مقاله از بردار ویژه ماتریس مجاورت برای تقسیم بندی گراف استفاده شد. در همان سال فیدلر[[12]](#footnote-12) نشان داد (Fiedler, (1973)) که برداد ویژه دومین مقدار ویژه کوچک ماتریس لاپلاسین به احتمال زیاد تقسیم بندیی ارائه می دهد که کمترین برش این تقسیم، اندازه بسیار کوچکتری دارد.

ماتریس لاپلاسین برای گراف ساده Gبا nیال به صورت زیر تعریف می شود. که در آن D ماتریس درجات راس هاست. یکی از روش های تعریف ماتریس لاپلاسین به شکل زیر است:

برای الگوریتم های شعاع طیفی ، ماتریس لاپلاسین ، پرکاربردترین ماتریس است (Fortunato, (2010)) .دونتی [[13]](#footnote-13)و مونتر[[14]](#footnote-14) در مقاله ی (Donetti, 2004) روش خود را بر مبنای بردار ویژه ماتریس لاپلاسین ارائه داده اند. بردار ویژه مولفه ها برای راس هایی که در یک انجمن قرار دارند، مقدار نزدیکی به هم هستند. پس می توان از آن برای کشف انجمن ها استفاده کرد. این روش به این صورت است که اگر از mبردار ویژه استفاده شود،راس ها را در فضای m بعدی قرار داده و راس هایی که در این فضا نزدیک هم قرار دارند را یک انجمن در نظر می گیریم. هرقدر تعداد بردار های ویژه بکار رفته بیشتر باشد،انجمن ها به صورت واضح تری در فضا مشخص می شود.در الگوریتمی که دونتی و مونز پیشنهاد دادند الگوریتم شامل مرحله گروه بندی نقاط و استخراج انجمن هاست. دونتی و مونز از خوشه بندی سلسله مراتبی است به این صورت که تنها انجمن هایی که حداقل یک یاله بین انجمنی در گراف اصلی است با هم ادغام می شوند. از بین تمامی تقسیم های استخراج شده، ساختاری که بزرگترین ماژولاریتی را دارد به عناون بهترین ساختار انتخاب شده و الگوریتم آن را به عنوان خروجی ارائه می دهد.پیچیدگی زمانی این روش است.

**4.4.2 الگوریتم های پویا**

این الگوریتم ها بصورت مستقیم بر روی گراف شبکه کار می کنند وانجمن ها را شناسایی می کنند. یکی از الگوریتم های مهم این دسته ،الگوریتم عابر تصادفی است که توسط ژو[[15]](#footnote-15) ارائه شد. (Zhou, 2003)

عابر تصادفی نخستین بار توسط هوگز[[16]](#footnote-16) در مقاله (Hughes, 1996) مطرح شد. ایده این مفهوم به این شکل است که عابری تصادفی به صورت تصادفی روی گراف حرکت می کند.در هر راس باتوجه به یالهای موجود به صورت تصادفی به هرکدا از راس های مجاور می رود.ایده این الگوریتمی که ژو مطرح کرد به ایت صورت است که عابر تصادفی به علت چگالی زیاد یالها در داخل انجمن ،زمان بیشتری را داخل انجمن خواهد بود.ژو ازین مفهوم برای تعریف مفهومی به نام فاصله بین دو راس استفاده کرد. به این صورت که فاصله بین دوراس و با نماد نشان داده می شود و به معنی تعداد یالهایی است که یک عابر تصادفی برای رسیدن به از راس اغازین باید از آنها عبور کند. راس هایی که فاصله کمی از هم دارند احتمالا به یک انجمن تعلق دارند. پیچیدگی زمانی این الگوریتم است.

**5.4.2 الگوریتم های بر اساس استنباط های آماری**

از استنباط های آماری می توان برای استخراج ویژگی ها با استفاده از یک سری مشاهدات و مقایسه آنها با مدل های فرضیی که داریم ،استفاده کنیم.اگر مجموعه داده ما گراف باشد،مدل مجموعه ای از راس هاست که به وسیله یالهایی به همدیگر متصل هستند ،این مدل با توپولوژی گراف شبکه منطبق است.این دسته از الگوریتم ها از این مدلها استفاده کرده و انجمن های موجود در گراف را شناسایی می کنند.

یکی از روشهای استنباط های آماری، استنباط بیزی است که در مدل سازی گراف های واقعی مانند شبکه های اجتماعی کاربرد فراوانی دارد،یکی از الگوریتم های معروفی که بر مبنای استنباط بیزی به منظور انجمن یابی مطرح شده است الگوریتم هاستینگ[[17]](#footnote-17) می باشد که در مقاله (Hastings, 2006) معرفی شد.

استنباط بیزی از مشاهدات ما برای تخمین و پیشبینی احتمال درستی یک فرضیه استفاده می کند.برای این کار دو مرحله داریم. مرحله اول مشاهدات است ک عبارت است از اطلاعات D که می توان از یک سیستم بدست آورد و مرحله دوم یک مدل آماری با پارامتر . استنباط بیزی را محاسبه می کند که برابر احتمال مشاهده شواهد مورد نظر ما در مدل کنونی با شرط وجود پارامتر است.هدف در این روش مشخص کردن مقداری برای مشاهدات مورد نظر یعنی ی است که مقدار احتمال را بیشینه کند.

**6.4.2 سایر روش ها**

الگوریتم های دیگری هم برای شناسایی انجمن ها ارائه شده اند که همگی در این دسته قرار می گیرند.یکی از الگوریتم های مهم الگورتم رغوان و همکاران (Raghavan, 2007) است که در ادامه توضیحاتی در مورد آن ارائه شده است.

در این الگوریتم ساده و سریع که به نام الگوریتم گسترش برچسب ها شناخته شده است، در ابتدا به هر راس برچسب های منحصر به فردی تعلق می گیرد. در هر مرحله یک بار ارزیابی و برچسب ها با کمک برچسب راس های همسایه مشخص می شود.یعنی برچسب هر راس برچسبی است که تعداد بیشتری از همسایه هایش آن برچسب را دارند. چنانچه تعداد برچسب ها در بین همسایه هایکسان باشد یکی از آنها به صورت تصادفی برای برچسب راس مورد نظر انتخاب می شود.با این روش در طول تکرار الگوریتم برچسب ها گسترش پیدا کرده و طبیعتا بعضی از برچسب ها به طور کلی ناپدید می شوند و برخی دیگر به برچسب های اصلی گراف تبدیل می شوند/این الگوریتم انقدر ادامه می یابد که دیگر برچسب ها تغییری نمی کنند و در نهایت الگوریتم متوقف و ساختار نهایی ساختار انجمن های گراف است.در این روش هر راس همسایه های بیشتری را در انجمنی دارد که به آن قرار دارد.این الگوریتم در هر اجرا روی یک گراف مشخص ممکن است جواب های مختلفی به ما بدهد اما نکته مهم در مورد آن پیچیدگی زمانی آن است که برای هر مرحله فقط است، که تعداد یالهای گراف است.

**5.2 مروری بر روش‏های انجمن‏یابی بر اساس بهینه سازی معیار ماژولاریتی**

نیومن[[18]](#footnote-18)یک الگوریتم حریصانه به نام FNپیشنهاد داد که در ابتدا هر گره را یک گروه در نظر گرفته سپس هردو گره را طوری ادغام می‌کند که ماژولاریتی بیشتری حاصل شود (Newman, 2004). کلازت [[19]](#footnote-19)از یک ساختمان داده‌ی پیچیده استفاده کرد تا پیچیدگی محاسبات ماژولاریتی را کاهش دهد ، با این کار الگوریتم FN برای شبکه های بزرگ هم قابل استفاده شد. (A. Clauset, 2004) کومار[[20]](#footnote-20)از تکنیک بهینه‌سازی جستجو گروهی برای پیدا کردن انجمن‌های بهینه در شبکه‌های دنیای واقعی استفاده کرد (G. K. Kumar). شانگ [[21]](#footnote-21)یک الگوریتم ژنتیک بهبود یافته به نام MIGAمعرفی کرد و رویکرد جدیدی برای ماکزیمم کردن ماژولاریتی ساخت (] R. Shang, 2013). فرتانتو [[22]](#footnote-22) و همکاران نشان دادند که الگوریتم‌های مبتنی بر بهینه‌سازی ماژولار ممکن است انجمن‌های کوچک‌تر از اندازه مشخصی را نتوانند شناسایی کنند. این اندازه به اندازه کل گراف و درجه ارتباط داخلی بین انجمن‌های مختلف بستگی دارد (] A. Arenas, 2008)این قضیه به محدودیت رزولوشن معروف شد. برای غلبه بر این محدودیت معیار‌های زیادی برای ارزیابی انجمن‌ها در مقیاس‌های مختلفی مطرح شد.

آرناس [[23]](#footnote-23)یک پارامتر برای ماژولاریتی معرفی کرد تا سطح رزولوشن را تنظیم کند (] A. Arenas, 2008). لی[[24]](#footnote-24) نوع دیگر از ماژولاریتی به نام چگالی ماژولاریتی را تعریف کرد تا اهمیت انجمن را اندازه بگیرد (Z. Li, 2008).پیزوتی [[25]](#footnote-25)معیار امتیاز جامعه [[26]](#footnote-26)  (CS) را مطرح کرد تا تضمین کند که ارتباطات داخل انجمن باید زیاد و ارتباط با بقیه انجمن‌ها اندک باشد (] C. Pizzuti, 2008).لانسیچینتی [[27]](#footnote-27)پیشنهاد داد از معیار برازندگی انجمن[[28]](#footnote-28) (CF)استفاده کنیم و مقیاس انجمن ها را اندازه بگیریم (A. Lancichinetti)

یک رویکرد دیگر برای حل محدودیت رزولوشن این است که مساله شناسایی انجمن را به صورت یک مساله بهینه‌سازی چند‌‌منظوره تعریف کرد. الگوریتم بهینه‌سازی چند‌منظوره با بهینه‌سازی همزمان چند هدف که از جهات مختلف مساله را ارزیابی میکند ، میتواند نقطه پرتو [[29]](#footnote-29) بهینه را پیدا کند. شی [[30]](#footnote-30) ماژولاریتی را به دو تابع هدف متضاد تقسیم کرد و با استفاده از الگوریتم تکاملی PESA-II آن‌را بهینه کرد. (] C. Shi, 2012)پیزوتی با در نظرگرفتن دو معیار CS ,CFبه عنوان دو هدف و استفاده از الگوریتم ژنتیک NSGA-IIبه بخش‌بندی بهینه دست‌یافت[11]. هدف این تحقیق ارایه یک راه‌حل متفاوت برای پیدا کردن بهینه ی انجمن‌ها و حل مشکل رزولوشن با رویکرد اتاماتای سلولی یادگیرنده است.[ 12]

**فصل چهارم**

**نتایج پیاده سازی الگوریتم وارزیابی آن ها**

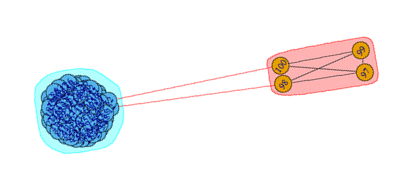
همان طور که قبلا ادعا شد ، استفاده از این رویکرد جدید برای انجمن یابی ، برخی از مشکلات شایع در مساله ی انجمن یابی را به خوبی مدیریت میکند. در این فصل به تفصیل ، این ادعا ها مطرح و نتایج بدست آمده بعد از پیاده سازی الگوریتم برای هر کدام از آنها بررسی و تحلیل می شود. در ادامه به ارزیابی نتایج این الگوریتم روی مجموعه داده های مصنوعی و مجموعه داده های معروف در حوزه انجمن یابی و مقایسه آن با نتایج چند الگوریتم معروف دیگر در این زمینه، پرداخته شده است.

**1.4 نتایج روی شبکه های مصنوعی**

در ابتدا به بیان آزمایشها و بررسی نتایج روی شبکه های مصنوعی می پردازیم.

**1.1.4آزمایش بررسی مساله محدودیت رزولوشن**

یکی از دلایل اصلی ارایه این رویکرد برای انجمن یابی ، برطرف کردن محدودیت مهم رزولوشن یا محدودیت اندازه انجمن های مختلف در شبکه ها بود. به این صورت که زمانی که اندازه یکی از انجمن ها بسیار بزرگ و دیگری بسیار کوچک باشد ، الگوریتم های انجمن یابی نمیتوانند ،انجمن کوچکتر را شناسایی کنند. همانطور که ادعا شد استفاده از اتاماتای سلولی یادگیرنده این محدودیت را برطرف می کند . برای نشان دادن این موضوع، عملکرد الگوریتم روی یک مجموعه داده ساختگی با شرایط زیر نشان داده شده است.

مجموعه داده ی زیر شامل 100 راس می باشد که از آن ها 96 راس در یک انجمن و با احتمال اتصال داخلی یک به همدیگر وصل هستند (گراف کامل 96 راسی) و4 تا راس باقی مانده در یک انجمن دیگر و با احتمال اتصال داخلی یک (گراف کامل4 راسی) و احتمال وجود یال بین این دو انجمن 0.01 می باشد. شکل 4- 1 نمایش گراف مورد نظر می باشد:

شکل 4- 1 : گراف شبیه سازی شده با 100راس که شامل زیرگراف کامل 96 و4 تایی است.

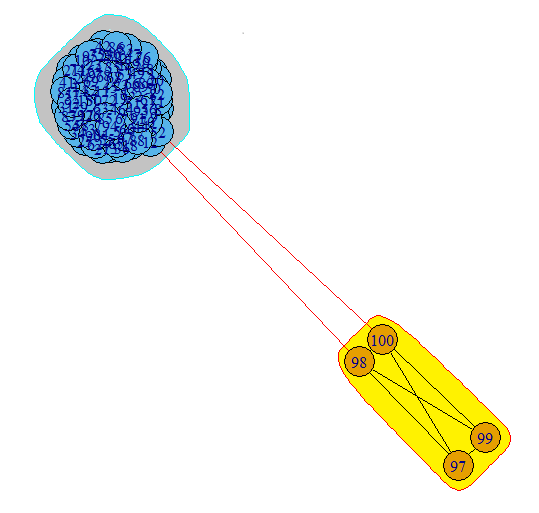
نتایج الگوریتم روی این گراف به صورت زیر است که در ادامه تحلیل آنها بیان شده است.

**Number of nodes: 100**

**Number of edges: 4568.0**

**Modularity is: 0.00262**

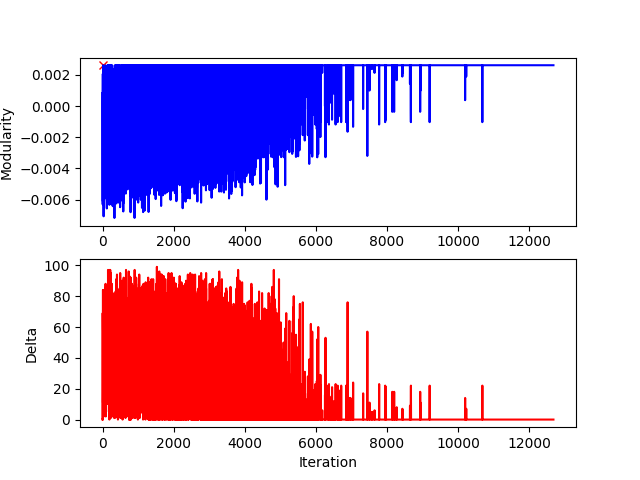
**NMI is: 1.0**



شکل 4-2 انجمن های پیدا شده روی گراف با الگوریتمCLA-net

همان طور که در شکل 4-3 گزارش شده است، الگوریتم به مقدار ماژولاریتی 0.00262همگرا شده و با تکرار انتخاب همسایگی جدید تغیری در نمودار عضویت پدیدار نمیشود لذا الگوریتم متوقف و مقدار نهایی ماژولاریتی را گزارش می دهد.

از آنجایی که گراف ساختگی می باشد،لذا برچسب های اصلی هر گره موجود است.حال برای ارزیابی عملکرد الگوریتم ،برچسب های اصلی با برچسب های پیدا شده توسط الگوریتم و توسط معیار NMI بررسی شده است. مقدار NMI=1 در این مثال نشان دهنده این است که بردار عضویت خروجی الگوریتم و بردار برچسب هایی گراف کاملا بر هم منطبق ولذا الگوریتم بدون مشکل توانسته انجمن های صحیح را تشخیص دهد و مشکل رزولوشن در این الگوریتم مطرح نیست.



شکل4-3 نمودار همگرایی ماژولاریتی و تغییرات دلتا در بررسی مساله محدودیت رزولوشن

**2.1.4آزمایش بررسی تاثیر نویز در عملکرد الگوریتم**

برای بررسی تاثیر نویز در مساله ، الگوریتم با مقدار اولیه alpharate=0.0003 و در حالت های مختلف زیر بررسی شده است.

ماتریس روبه رو مقدار ارتباط بین گراف ساختگی را نشان می دهد.به این صورت که درایه م این ماتریس نشان می دهد احتمال وجود یال بین انجمن و انجمن گروه چه مقداری است.

اندازه ماتریس همیشه مربعی و برابر m\*mمی باشد که mهمان تعداد انجمن های مورد نظر ما برای گراف ساختگی است.

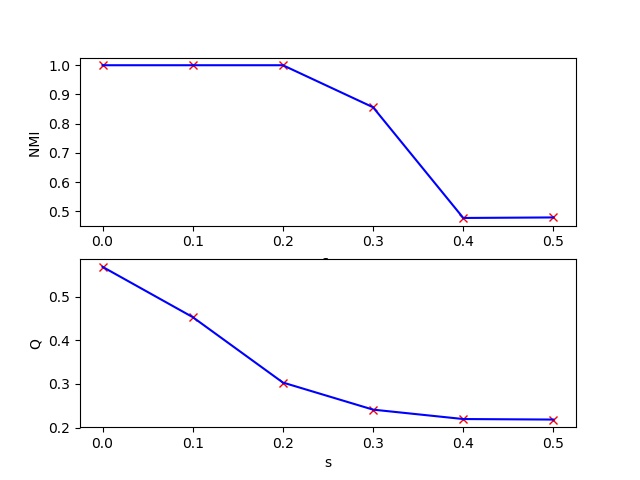
چنانچه مقدار sصفر باشد یعنی نویز وجود ندارد و انجمن ها کاملا مستقل از هم می باشند.

در اینجا برای مقدار های متفاوت sاز 0 تا 0.5 الگوریتم روی گراف ثابت 21 راسی با اندازه انجمن های 8و 4 و 9 تایی ازمایش و نتایج به صورت زیر به دست امده است. برای هر مقدار به دلیل تصادفی بودن ده بار اجزا گرفته شده و میانگین مقدار ها در جدول مورد بررسی قرار گرفته است

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| S | دفعات اجرا | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Avg. |
| 0 | Q | 0.5681 | 0.5681 | 0.5681 | 0.5681 | 0.5681 | 0.5681 | 0.5681 | 0.5681 | 0.5681 | 0.5681 | **0.5681** |
| Iteration | 26910 | 32331 | 30129 | 27589 | 30903 | 26983 | 27812 | 29501 | 26501 | 29741 | **28840** |
| NMI | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | **1** |
| 0.1 | Q | 0.4530 | 0.4530 | 0.4530 | 0.4530 | 0.4530 | 0.4530 | 0.4530 | 0.4530 | 0.4530 | 0.4530 | **0.4530** |
| Iteration | 29622 | 29285 | 28850 | 28823 | 29688 | 30238 | 29255 | 30724 | 29214 | 29525 | **29522** |
| NMI | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | **1** |
| 0.2 | Q | 0.3033 | 0.3033 | 0.3033 | 0.3033 | 0.3033 | 0.3033 | 0.3033 | 0.3033 | 0.3033 | 0.3033 | **0.3033** |
| Iteration | 30535 | 32220 | 33106 | 32059 | 36292 | 31286 | 33296 | 31912 | 32277 | 34371 | **32725** |
| NMI | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | **1** |
| 0.3 | Q | 0.2413 | 0.2413 | 0.2413 | 0.2413 | 0.2413 | 0.2413 | 0.2413 | 0.2413 | 0.2413 | 0.2413 | **0.2413** |
| Iteration | 31895 | 33222 | 35317 | 33452 | 37505 | 34640 | 38316 | 36321 | 31592 | 32030 | **34429** |
| NMI | 0.856 | 0.856 | 0.856 | 0.856 | 0.856 | 0.856 | 0.856 | 0.856 | 0.856 | 0.856 | **0.856** |
| 0.4 | Q | 0.2198 | 0.2198 | 0.2198 | 0.2198 | 0.2198 | 0.2198 | 0.2198 | 0.2198 | 0.2198 | 0.2198 | **0.2198** |
| Iteration | 34240 | 35150 | 36384 | 34501 | 60921 | 33822 | 32655 | 39921 | 31794 | 82928 | **42231** |
| NMI | 0.4779 | 0.4779 | 0.4779 | 0.4779 | 0.4779 | 0.4779 | 0.4779 | 0.4779 | 0.4779 | 0.4779 | **0.4779** |
| 0.5 | Q | 0.2186 | 0.2186 | 0.2186 | 0.2186 | 0.2186 | 0.2077 | 0.2186 | 0.2186 | 0.2186 | 0.2186 | **0.2186** |
| Iteration | 37015 | 39838 | 34747 | 42451 | 36715 | 42329 | 36582 | 45204 | 40705 | 37649 | **39323** |
| NMI | 0.4788 | 0.4788 | 0.4788 | 0.4788 | 0.4788 | 0.4985 | 0.4788 | 0.4788 | 0.4788 | 0.4788 | **0.4788** |

|  |  |  |
| --- | --- | --- |
|  |  | 0 |
|  |  | 0.1 |
|  |  | 0.2 |
|  |  | 0.3 |
|  |  | 0.4 |
|  |  | 0.5 |

همانطورکه در شکل مشخص است با افزایش نویز احتمال خطا افزایش و لذا هردو معیار کاهش می یابند.اما حتی در بیشترین مقدار نویز ممکن هم معیار NMIاز 0.5 کمتر نشده واین پایداری الگوریتم را نسبت به نویز نشان می دهد.



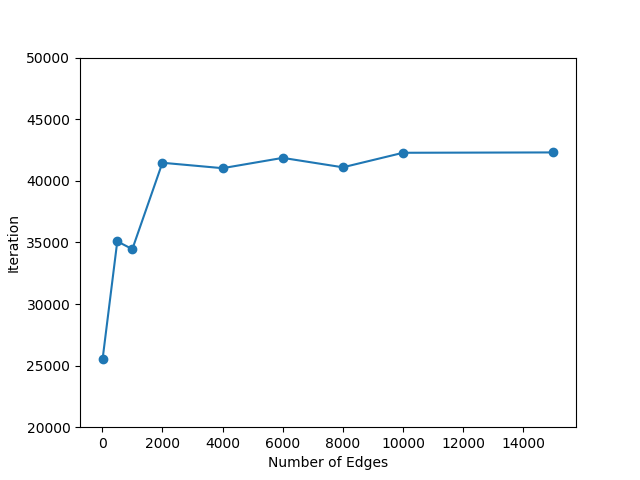
شکل 4-4 .نمودار تاثیرافزایش نویز S ، بر عملکرد الگوریتم با معیار هایNMI و ماژولاریتی

**3.1.4 آزمایش بررسی تاثیر اندازه شبکه در سرعت همگرایی الگوریتم**

آزمایش سوم طراحی شده تا سرعت همگرایی الگوریتم را بررسی کند. برای این منظور از مجموعه داده های ساختگی استفاده شد. از آنجایی که الگوریتم بر مبنای تعداد یالها و نه راس ها کار میکند،گراف های ساختگی باتعداد یال متفاوت اجرا و میانگین تعداد تکرارهای لازم برای همگرایی آن بررسی شد.

در این بررسی شبکه ها کاملا تصادفی و با میزان نویز های متفاوت انتخاب شده اند به همین منظور بهترین جواب لزوما1 NMI= نمی باشد. الگوریتم روی 9 گراف ساختگی تصادفی ،5 بار اجراشده است و نتایج آن در قالب جدول زیر بیان شده که ستون آخر بیانگر میانگین خروجی های الگوریتم است و نمودار تعداد تکرارها بر حسب یال ها در ادامه رسم شده است.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| تعداد  یالها |  | 1 | 2 | 3 | 4 | 5 | **Avg**. |
| 10  10 | Q | 0.4444 | 0.4444 | 0.4444 | 0.4444 | 0.4444 | **0.4444** |
| Iteration | 24963 | 29677 | 27643 | 22249 | 22952 | **25497** |
| NMI | 0.7489 | 0.7489 | 0.7489 | 0.7489 | 0.7489 | **0.7489** |
| 500  50 | Q | 0.3720 | 0.3720 | 0.3720 | 0.3720 | 0.3720 | **0.3720** |
| Iteration | 37609 | 35480 | 37676 | 36619 | 35080 | **35085** |
| NMI | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | **1.0** |
| 1000  75 | Q | 0.6259 | 0.6259 | 0.6259 | 0.6259 | 0.6259 | **0.6259** |
| Iteration | 32513 | 31334 | 39044 | 36764 | 32605 | **34452** |
| NMI | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | **1.0** |
| 2000  100 | Q | 0.3563 | 0.3563 | 0.3563 | 0.3563 | 0.3563 | **0.3563** |
| Iteration | 42868 | 40334 | 36052 | 42622 | 45445 | **41464** |
| NMI | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | **1.0** |
| 4000  150 | Q | 0.4298 | 0.4298 | 0.4298 | 0.4298 | 0.4298 | **0.4298** |
| Iteration | 41928 | 35080 | 39845 | 45215 | 43056 | **41024** |
| NMI | 0.7809 | 0.7809 | 0.7809 | 0.7809 | 0.7809 | **0.7809** |
| 6000  200 | Q | 0.2921 | 0.2921 | 0.2921 | 0.2921 | 0.2921 | **0.2921** |
| Iteration | 42473 | 41256 | 43356 | 39654 | 42569 | **41861** |
| NMI | .72590 | 7259.0 | 7259.0 | 7259.0 | 7259.0 | **7259.0** |
| 8000  225 | Q | 0.3603 | 0.3603 | 0.3603 | 0.3603 | 0.3603 | **0.3603** |
| Iteration | 41250 | 42045 | 45182 | 40952 | 36048 | **41095** |
| NMI | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | **1.0** |
| 10000  250 | Q | 0.4168 | 0.4168 | 0.4168 | 0.4168 | 0.4168 | **0.4168** |
| Iteration | 39938 | 42058 | 41256 | 43465 | 44659 | **42275** |
| NMI | .76600 | .76600 | .76600 | .76600 | .76600 | **.76600** |
| 15000  300 | Q | 0.3816 | 0.3816 | 0.3816 | 0.3816 | 0.3816 | **0.3816** |
| Iteration | 45647 | 42985 | 41595 | 39085 | 42202 | **42303** |
| NMI | 0.7604 | 0.7604 | 0.7604 | 0.7604 | 0.7604 | **0.7604** |

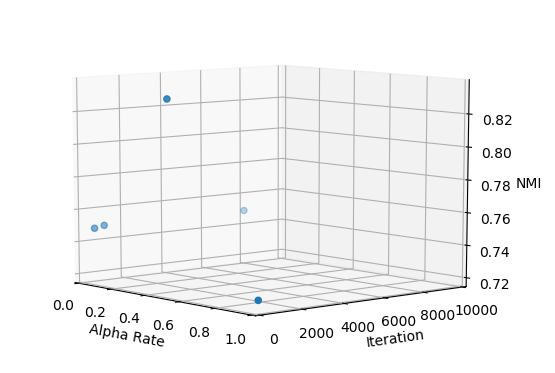


همانطور که در نمودار مشاهده میشود تعداد تکرارهای الگوریتم برای همگرا شدن بر حسب تعداد یال، نسبتا ثابت و در کل الگوریتم نسبت به افزایش یال ها ثبات خوبی نشان داده است.

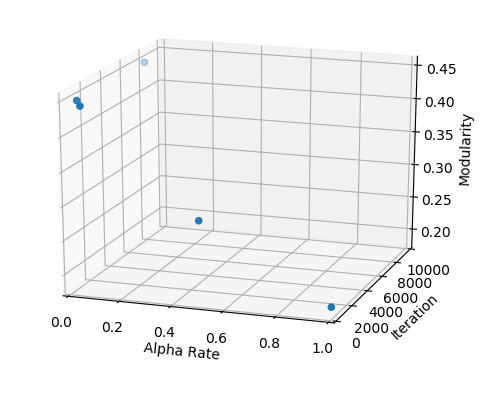
**4.1.4آزمایش بررسی تاثیر نرخ یادگیری در سرعت همگرایی**

نرخ یادگیری برای اتاماتاهای یادگیری در الگوریتم مقداری اختیاری و قابل تنظیم است. اما از انجایی که نرخ های بسیار پایین مانند alpharate=0.0003 باعث تعداد تکرار های بسیار بالای الگوریتم جهت همگرایی می شوند،لذا در این ازمایش روی یک گراف ثابت تصادفی الگوریتم با نرخ های متفاوت اجرا و نتایج مقایسه می شوند.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| نرخ یادگیری |  | 1 | 2 | 3 | **Avg**. |
| 1 | Q | 0.1234 | 0.4444 | 1.85 e-17 | **0.1892** |
| Iteration | 103 | 103 | 103 | **103** |
| NMI | 0.419 | 0.7489 | 1.0 | **0.7226** |
| 0.5 | Q | 0.4444 | 0.4444 | 1.85 e-17 | **0.2962** |
| Iteration | 108 | 108 | 103 | **106** |
| NMI | 0.7489 | 0.7489 | 1.0 | **0.8326** |
| 0.05 | Q | 0.4444 | 0.4444 | 0.4444 | **0.4444** |
| Iteration | 222 | 194 | 203 | **206** |
| NMI | 0.7489 | 0.7489 | 0.7489 | **0.7489** |
| 0.005 | Q | 0.4444 | 0.4444 | 0.4444 | **0.4444** |
| Iteration | 1170 | 937 | 918 | **1008** |
| NMI | 0.7489 | 0.7489 | 0.7489 | **0.7489** |
| 0.0005 | Q | 0.4444 | 0.4444 | 0.4444 | **0.4444** |
| Iteration | 7698 | 8423 | 7904 | **8008** |
| NMI | 0.7489 | 0.7489 | 0.7489 | **0.7489** |
| 0.00005 | Q | 0.4444 | 0.4444 | 0.4444 | **0.4444** |
| Iteration | 64706 | 67140 | 70483 | **67443** |
| NMI | 0.7489 | 0.7489 | 0.7489 | **0.7489** |



این نمودار نشان می دهد که در مقدار 0.5 برای نرخ یادگیری در این مثال خاص، بالاترین مقدار NMI با تقریبا کمترین تعداد تکرار الگوریتم به دست آمده است. از طرفی دیگر با رسم نمودار ماژولاریتی بر حسب نرخ یادگیری و تعداد تکرار مشاهده می شود که باز هم با افزایش مقدار نرخ یادگیری ماژولاریتی دیگر افزایشی نداشته است. لذا افزایش مقدار نرخ یادگیری به جز افزایش تعداد تکرار تاثیر دیگری ندارید.



**5.1.4آزمایش مقایسه نتایج الگوریتم با سایر الگوریتم های انجمن یابی**

**2.4 بررسی نتایج الگوریتم روی مجموعه داده های واقعی**

# منابع

-MCCRACKEN, J. (1968). *POLITICS IN A CHANGING SOCIETY-POLITICAL HISTORY OF FORT JAMESON NGONI-BARNES, JA.*

Moreno, J. L. (1953). *Who shall survive? foundations of sociometry, group psychotherapy and socio-drama.*

Takes, F. W. ( 2011.). Identifying prominent actors in online social networks using biased random walks. *In Proceedings of the 23rd benelux conference on artificial intelligence (BNAIC)*, (pp. 215–222).

1-Wasserman, S. a. (1994). *Social network analysis: Methods and applications, volume 8.* Cambridge university press.

Zhang, J. A. (2007). Expertise networks in online communities: structure and algorithms. *In Proceedings of the 16th international conference on World Wide Web* (pp. 221–230). ACM.

Zhao, Yuxin, Wen Jiang, Shenghong Li, Yinghua Ma, Guiyang Su, and Xiang Lin. "A cellular learning automata based algorithm for detecting community structure in complex networks." *Neurocomputing* 151 (2015): 1216-1226.(مقاله بیس)

1. Modularity [↑](#footnote-ref-1)
2. Normalize mutual information (NMI) [↑](#footnote-ref-2)
3. John Barnes [↑](#footnote-ref-3)
4. Wasserman [↑](#footnote-ref-4)
5. Sparse [↑](#footnote-ref-5)
6. Scale-Free [↑](#footnote-ref-6)
7. community [↑](#footnote-ref-7)
8. Radacci [↑](#footnote-ref-8)
9. Raghavan [↑](#footnote-ref-9)
10. Donath [↑](#footnote-ref-10)
11. Hoffman [↑](#footnote-ref-11)
12. Fiedler [↑](#footnote-ref-12)
13. Donetti [↑](#footnote-ref-13)
14. Munoz [↑](#footnote-ref-14)
15. Zhou [↑](#footnote-ref-15)
16. Hughes [↑](#footnote-ref-16)
17. Hastings [↑](#footnote-ref-17)
18. Newman [↑](#footnote-ref-18)
19. Clauset [↑](#footnote-ref-19)
20. Kumar [↑](#footnote-ref-20)
21. Shang [↑](#footnote-ref-21)
22. Fortunato and Barth´elemy [↑](#footnote-ref-22)
23. Arenas [↑](#footnote-ref-23)
24. Li [↑](#footnote-ref-24)
25. Pizzuti [↑](#footnote-ref-25)
26. Community Score [↑](#footnote-ref-26)
27. Lancichinetti [↑](#footnote-ref-27)
28. Community Fitness [↑](#footnote-ref-28)
29. Pareto [↑](#footnote-ref-29)
30. Shi [↑](#footnote-ref-30)