**شبکه** مجموعه ای از اجزا و ارتباط های بین آنهاست. در دنیای واقعی با شبکه های متفاوتی روبه رو هستیم ، شبکه های بیولوژیکی ،شبکه ترافیک،شبکه های زیستی وشبکه اطلاعات و... تنها انواع مختلفی از شبکه ها هستند. یکی ازقالب های قابل کاربرد برای مدل کردن یک شبکه ، گراف است.نمایش گرافی شبکه را ملموس تر وانجام اعمال مختلفی را روی آن ممکن می سازد.

نمایش ریاضی یک **گراف** به صورت است.در این تعریف:

مجموعه ای شامل راس ها یا گره ها

زیرمجموعه های دوتایی از گره های گراف است که به آنها یال گفته میشود.

هر گراف را می توان بایک **ماتریس مجاورت** منحصربهفرد نمایش داد.ماتریس مجاورتA را به این صورت تعریف میکنیم که درایه اگر دو گره به طور مستقیم به هم وصل بودند ودر غیر این صورت .

یکی از راه های تجزیه و تحلیل شبکه و درک بهتر ماهیت آن پیدا کردن زیر گراف های گراف شبکه است.

**زیر گراف C**خود یک گراف است که مجموعه راس های آن زیر مجموعه ای از مجموعه راس های گراف بزرگتری باشد، واضح است که دراینصورت یالهای زیرگراف هم زیرمجموعه ای از یالهای گراف بزرگتر خواهد بود.

تبدیل گراف به زیر گراف های معنا دار در حقیقت تلاش برای شناخت بهتر شبکه است واطلاعات بسیاری درمورد ماهیت آن در اختیار ما قرار می دهد

یکی از ویژگی های مهم شبکه ها ی بزرگ **انجمن ها[[1]](#footnote-1)** است. انجمن یابی در یک شبکه کاربردهای زیادی دارد برای مثال ازآنجایی که افراد حاضر در انجمن هاي تشکیل شده دریک شبکه اجتماعی به احتمال زیاد علایق مشترکی دارند، می توان با یافتن علایق آن ها از این اطلاعات در مسایل مربوط به تبلیغات وبازاریابی استفاده کرد .مثال دیگری از این کاربرد مربوط به انتشار اخبار است،از آنجایی که اعضای یک انجمن باهم در ارتباط هستند لذا برای انتشار خبر یا تبلیغات می توان 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 انجمن های موجود در شبکه خبر را ارسال و در هزینه های مربوطه صرفه جویی می شود.

یک انجمن می تواند به طور کلی به عنوان مجموعه ای از راس ها که چگالی بالایی در ارتباط با زیر گراف خود (اتباط داخلی)و وارتباط بسیار کمی با سایر زیرگراف هادارند، توصیف شود.این تعریف از جهت هایی مبهم است.هنگام برخورد با مساله شناسایی انجمن ها ما باید تعریف دقیق و روشن تری از مفهوم اجتماع داشته باشیم .لذا اینجا چند تعریف دقیق تر از آن را ارایه میدهیم که در حوزه ی شناسایی انجمن ها پذیرفته شده اند.

راداچی[[2]](#footnote-2) برای هر راس از زیرگرافمفهوم( درجه راس) را به شکل زیر مطرح کرد.

درجه راس را نمایش می دهد که مجموع دومقدار است:

-تعداد یال هایی که از راس به زیرگرافc وصل است.

- یال هایی از راس که به زیرگراف cوصل نیستند.(به سایر زیر گراف ها وصل شده اند.)

یک انجمن قوی زیرگرافی است که خاصیت زیر را دارد:

یعنی تمام گره های انجمن ارتباط بیشتری با زیرگراف C (نسبت به سایر زیرگراف ها)دارند.

یک انجمن ضعیف زیرگرافی با خاصیت زیر است:

یعنی جمع ارتباط هایی که گره های این گروه با زیر گرافC دارند بزرگتر باشد از جمع ارتباط هایی که گره ها با زیر گرافc ندارند.

انجمن قوی،انجمن ضعیف هم هست.ولی عکس قضیه همواره برقرار نیست.در ادامه تعریف دیگری برای انجمن ها توسط رغوان[[3]](#footnote-3) ارایه شده است.

اگر Ωمجموعه شامل تمام انجمن های گراف باشد.پس |Ω|تعداد انجمن های گراف را نشان میدهد.کل درجه های هرگره نهایتا به |Ω|بخش تقسیم میشود:

یعنی درجه راس برابر است با یال هایی که از آن به هر کدام از انجمن ها وصل است.

که

یعنی درجه تعلق گره به انجمن cبرابر است با یالهایی که از آن گره به انجمن C وصل است.

پس برای انجمن c این تعریف به شکل زیر است:

یعنی تمام گره ها ارتباط بیشتری(یامساوی) با انجمنی دارند،که متعلق به آن هستند.زمانی که فقط دوتا انجمن داریم این تعریف همان تعریف اجتماع قوی است که پیشتر ارایه شد. اما زمانی که گراف بیشتر از دوتا انجمن دارد،محدودیت رغوان ضعیف تر انجمن قوی بودن است.

هدف **الگوریتم های انجمن یابی** آن است که گراف را به مهمترین انجمن ها تقسیم کند. پس مساله مهمی که مطرح می شود داشتن معیار مناسبی برای ارزیابی کیفیت انجمن هاست.

برای این منظور نیومن [[4]](#footnote-4)مفهموم **مادولاریتی[[5]](#footnote-5)** را ارایه داد.ایده ی مادولاریتی نشات گرفته از این است که در یک گراف تصادفی بخاطر توزیع یکسان درجات ، چگالی یال در قسمت خاصی بیشتر نمیشود ولذا انتظار نداریم در این گراف هیچ انجمنی وجود داشته باشد.

فرمول مادولاریتی را به شکل زیر می توان نوشت:

(4)

تعداد یالهای گراف

تعداد درایه های ماتریس مجاورت گراف

درجه گره

: تابع دلتای کرونیکر تعمیم یافته ، به این صورت که برای زمانی که یکسان باشند ودر غیر این صورت صفر است.

: این ترم مقدار قابل انتظار ما از درجه ارتباط بین دو گره در یک شبکه تصادفی با اندازه توزیع درجه یکسان است.

اگر نسبت یالها در انجمن داده شده بزرگتراز مقدار قابل انتظار در شبکه تصادفی بود مقدار Qبزرگتر از 0 می شود. توجه به این نکته مهم است که مادولاریتی بازاء کل انجمن ها تعریف می شود و بازاء یک انجمن خاص نیست. به عبارتی هرچه Q بزرگتر باشد یعنی انجمن های شناسائی شده بهتر هستند.با این معیار مساله پیدا کردن بهترین انجمن تبدیل به مساله پیدا کردن مقدار بهینه ی Qشد.

باتوجه به اینکه تعداد تقسیم بندی های ممکن رشد بیشتری از هر توانی از اندازه گراف دارد. ثابت شده است مساله پیدا کردن مادولاریتی بهینه یک مساله NP\_hardاست. نیومن 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 مادولاریتی بیشتری حاصل شود.[[6]](#endnote-1) کلازت [[7]](#footnote-6)از یک ساختمان داده ی پیچیده استفاده کرد تا پیچیدگی محاسبات مدولاریتی را کاهش دهد وبا این کار الگوریتم FN برای شبکه های بزرگ هم قابل استفاده شد.[[8]](#endnote-2) کومار[[9]](#footnote-7)از تکنیک بهینه سازی جستجو گروهی برای پیدا کردن اجتماع های بهینه در شبکه های دنیای واقعی استفاده کرد[[10]](#endnote-3). شانگ [[11]](#footnote-8)یک الگوریتم ژنتیک بهبود یافته به نام MIGAمعرفی کرد و رویکرد جدیدی برای ماکزیمم کردن مادولاریتی ساخت[[12]](#endnote-4). فرتانتو [[13]](#footnote-9)وهمکارانش نشان دادند که الگوریتم های مبتنی بر بهینه سازی مدولار ممکن است انجمن های کوچکتر از اندازه مشخصی را نتوانند شناسایی کنند. این اندازه به اندازه کل گراف ودرجه ارتباط داخلی بین انجمن های مختلف بستگی دارد[[14]](#endnote-5) .این قضیه به محدودیت **رزولوشن** معروف شد. برای غلبه بر این محدودیت معیار های زیادی برای ارزیابی اجتماع ها در مقیاس های مختلفی مطرح شد.

آرناس [[15]](#footnote-10)یک پارامتر مقیاس برای مادولاریتی معرفی کرد تا سطح رزولوشن را تنظیم کند[[16]](#endnote-6). لی[[17]](#footnote-11) نوع دیگر از مادولاریتی به نام چگالی مادولاریتی را تعریف کرد تا اهمیت اجتماع را اندازه بگیرد[[18]](#endnote-7).

پیزوتی [[19]](#footnote-12)معیار امتیاز جامعه [[20]](#footnote-13)(CS)را مطرح کرد تا تضمین کند که بین اجتماع باید ارتباطات زیاد وارتباط بین اجتماع ها اندک باشد.[[21]](#endnote-8)

لانسیچینتی [[22]](#footnote-14)پیشنهاد داد ازمعیاربرازندگی انجمن[[23]](#footnote-15) (CF)استفاده کنیم ومقیاس اجتماع ها را اندازه بگیریم.[[24]](#endnote-9)

یک رویکرد دیگر برای حل محدودیت رزولوشن این است که مساله شناسایی اجتماع را به صورت یک مساله بهینه سازی چند منظوره تعریف کرد.الگوریتم بهینه سازی چند منظوره با بهینه سازی همزمان چند هدف که از جهات مختلف مساله را ارزیابی میکند ،میتواند نقطه پرتو بهینه را پیدا کند.

شی [[25]](#footnote-16)مدولاریتی را به دوتابع هدف متضاد تقسیم کرد و با استفاده از الگوریتم تکاملی PESA-II آنرا بهینه کرد.[[26]](#endnote-10)پیزوتی با در نظرفتن دو معیار CS ,CFبه عنوان دوهدف واستفاده از الگوریتم ژنتیک NSGA-IIبه بخش بندی بهینه دست یافت.[[27]](#endnote-11)

هدف این مقاله حل مشکل پیدا کردن بهینه ی انجمن ها وحل مشکل رزولوشن با رویکرد اتاماتای یادگیری است.

الگوریتم ارایه شده بر مبنای اتاماتای سلولی یادگیرنده[[28]](#footnote-17) است وCLA-net نام دارد که در آن کل شبکه به عنوان اتاماتای سلولی یادگیرنده نامنظم مدل میشود. اتاماتای سلولی یادگیرنده CLAیک مدل ریاضی قوی برای بسیاری از مسایل غیر متمرکز و پدیده های پویا است. ایده اصلی اتاماتای سلولی یادگیرنده استفاده از اتاماتای یادگیری برای تنظیم احتمال انتقال حالت در اتاماتای تصادفی سلولی است.اتاماتای سلولی یادگیرنده را می توان نوعی از اتاماتای سلولی در نظر گرفت که در آن هر کدام از سلول های اتاماتای سلولی مجهز به یک اتاماتای یادگیری است. اتاماتای یادگیری مقیم در هر سلول ، حالت آن را بر اساس بردار احتمال اقدام تعیین میکند.هر اتاماتای یادگیری تلاش میکند اقدام بهینه را با تعامل با محیط محلی(اتاماتای یادگیری سلول های همسایه اش) یادبگیرد. پروسه انقدر تکرار می شود تا حالت بهینه هرکدام از سلول ها بدست بیاید و به طور موثر مشکل محدودیت رزولوشن در بهینه سازی ماژولار را حل کند.

الگوریتم روی داده های تولید شده ی فرضی و مجموعه داده های واقعی کاراته[[29]](#footnote-18)[[30]](#endnote-12) و دلفین ها [[31]](#footnote-19)[[32]](#endnote-13) و فوتبال[[33]](#footnote-20)وSFI[[34]](#endnote-14) وشبکه علم [[35]](#footnote-21)[[36]](#endnote-15) و داده قدرت گرید [[37]](#footnote-22)[[38]](#endnote-16) اعمال شده است و نتایج آن با الگوریتم های CNMو MIGAو GA-net و MOCD و Meme-net و MOGAnetو D-net/MOEA مقایسه شده است.

معیار های ارزیابی نتایج این الگوریتم ،معیار مادولاریتی و معیار اطلاعات مشترک نرمال سازی شده،NMI است[[39]](#endnote-17). همانطور که بیان شد معیار مادولاریتی Qاهمیت اجتماع را در شبکه بررسی می کند. مقدار بزرگتر Qنشان می دهد کیفیت انجمن ها بهتر و دورتر از حالت تصادفی مورد انتظار هستند.معیار NMIمخصوص شبکه های با انجمن های شناخته شده است. به این صورت مقدار شباهت بین انجمن واقعی و انجمن بدست امده توسط الگوریتم را می سنجد. ارزش NMI بین [0، 1] است و مقدار بزرگتر نشان می دهد انجمن های به دست آمده مطابقت بیشتری با انجمن های واقعی دارند.

1. community [↑](#footnote-ref-1)
2. Radacci [↑](#footnote-ref-2)
3. Raghavan [↑](#footnote-ref-3)
4. Newman [↑](#footnote-ref-4)
5. modularity [↑](#footnote-ref-5)
6. M. E. J. Newman, Fast algorithm for detecting community structure in

   networks, Phys. Rev. E 69(2004)066133. [↑](#endnote-ref-1)
7. Clauset [↑](#footnote-ref-6)
8. A. Clauset, M. E. J. Newman, C. Moore, Finding community structure in

   very large networks, Phys. Rev. E 70(2004)066111. [↑](#endnote-ref-2)
9. Kumar [↑](#footnote-ref-7)
10. چک کن کومار [↑](#endnote-ref-3)
11. Shang [↑](#footnote-ref-8)
12. شانگ [↑](#endnote-ref-4)
13. Fortunato and Barth´elemy [↑](#footnote-ref-9)
14. S. Fortunato, M. Barth´elemy, Resolution limit in community detection,

    Proc. Natl. Acad. Sci. USA 104(2007)36-41. [↑](#endnote-ref-5)
15. Arenas [↑](#footnote-ref-10)
16. ارناس [↑](#endnote-ref-6)
17. Li [↑](#footnote-ref-11)
18. li [↑](#endnote-ref-7)
19. Pizzuti [↑](#footnote-ref-12)
20. Community Score [↑](#footnote-ref-13)
21. پیزوتی [↑](#endnote-ref-8)
22. Lancichinetti [↑](#footnote-ref-14)
23. Community Fitness [↑](#footnote-ref-15)
24. لانسینچیتی [↑](#endnote-ref-9)
25. Shi [↑](#footnote-ref-16)
26. شی [↑](#endnote-ref-10)
27. پیزوتی [↑](#endnote-ref-11)
28. cellular learning automata [↑](#footnote-ref-17)
29. Karate [↑](#footnote-ref-18)
30. W. W. Zachary, An information flow model for conflict and fission in small

    groups, J. Anth. Res. 33(1997)452-473. [↑](#endnote-ref-12)
31. Dolphins [↑](#footnote-ref-19)
32. D. Lusseau, The emergent properties of a dolphin social network, Proc. R.

    Soc. Lond. B 270(2003)S1860-1888. [↑](#endnote-ref-13)
33. Football [↑](#footnote-ref-20)
34. M. Girvan, M. E. J. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. USA 99(2002)7821-7826. [↑](#endnote-ref-14)
35. Netscience [↑](#footnote-ref-21)
36. M. E. J. Newman, Finding community structure in networks using the eigenvectors of matrices, Phys. Rev. E 74(2006)036104. [↑](#endnote-ref-15)
37. Powergrid [↑](#footnote-ref-22)
38. D. J. Watts, S. H. Strogatz, Collective dynamics of ’small-world’ networks,

    Nature 393(1998)440-442. [↑](#endnote-ref-16)
39. A. Lancichinetti, S. Fortunato, J. Kertesz, Detecting the overlapping and hierarchical community structure in complex networks, New J. Phys.

    11(2009)033015. [↑](#endnote-ref-17)