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 $G = (V, E)$ است.در این تعریف:

V مجموعه ای شامل راس ها یا گره ها

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

$$E = \{(i, j) | \ i, j \in V\}$$

هر گراف را می توان بایک **ماتریس مجاورت** منحصر به فرد نمایش داد.ماتریس مجاورتA را به این صورت تعریف میکنیم که درایه $A_{ij} = 1$ اگر دو گره iوj به طور مستقیم به هم وصل بودند ودر غیر این صورت $A_{ij} = 0$.

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

زیر گراف Cخود 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 است. انجمن یابی در یک شبکه اجتماعی کاربردهای زیادی دارد برای مثال ازآنجایی 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 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 کرد.

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 حوضه ی شناسایی انجمن ها پذیرفته شده اند.

راداچی[1] برای هر راس از زیرگراف Cمفهوم(درجه راس) را به شکل زیر مطرح کرد.

[1] Radacci

$$k_i = k_i^{in}(C) + k_i^{out}(C)$$

k_i درجه راس iرا نمایش می دهد که مجموع دومقدار است:

-تعداد $k_i^{in}(C) = \sum_{j \in C} A_{ij}$ یال هایی که از راس i به زیرگرافcوصل است.

-یالهایی $k_i^{out}(C) = \sum_{j \notin C} A_{ij}$ از راس i که به زیرگرافcوصل نیستند.(به سایرزیر گراف ها وصل شده اند)

یک انجمن قوی زیرگرافی است که خاصیت زیر را دارد

$$k_i^{in}(C) > \quad , \quad \forall i \in C \qquad (1)$$

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

یک انجمن ضعیف زیرگرافی با خاصیت زیر است:

$$\sum_{i \in C} k_i^{in}(C) > \sum_{i \in C} k_i^{out}(C) \qquad (2)$$

یعنی جمع ارتباط هایی که گره های این گروه با زیر گرافC دارند بزگتر باشد از جمع ارتباط هایی که گره هابا زیر گراف cندارند.

انجمن قوی،انجمن ضعیف هم هست.ولی عکس قضیه همواره برقرار نیست.

در ادامه تعریف دیگری برای انجمن ها توسط رغوان [2]ارایه شده است.

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

$$k_i = \sum_{C \in \Omega} k_i(C)$$

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

که در این تعریف $k_i(C) = \sum_{j \in C} A_{ij}$ است.

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

[2] Raghavan

پس برای انجمن c این تعریف به شکل زیر است:

$$k_i(C) \geq k_i(C'),\ \forall i \in C \quad \forall C' \in \Omega \qquad (3)$$

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 مهترین انجمن ها تقسیم کند.اما مساله مهمی که مطرح می شود داشتن معیار مناسبی برای ارزیبی کیفیت انجمن هاست.

برای این منظور نیومن [3]مفهموم **مادولاریتی** را ارایه دادند.ایده ی مادولاریتی نشات گرفته ازین 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:

$$Q = \frac{1}{2m} \sum_{i,j \in V} (A_{ij} - \frac{k_i k_j}{2m}) \delta(i,j) \qquad (4)$$

m تعداد یالهای گراف

$A_{i,j}$تعداد درایه های ماتریس مجاورت گراف

k_i درجه گره i

$\delta(.)$: تابع دلتای کرونیکر تعمیم یافته ، به این صورت که برای زمانی که i,jیکسان باشند $\delta(i,j) = 1$ودر غیر این صورت صفر است.

$\frac{k_i k_j}{2m}$: این ترم مقدار قابل انتظار ما از درجه ارتباط بین دو گره i,j 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 0می شود.هرچه مقدار Qبزرگتر باشد نشان دهنده اهمیت بیشتر این انجمن در شبکه است.

[3] Newman and Girvan

توجه به این نکته مهم است که مادولاریتی بازاء کل اجتماع ها تعریف می شود و بازاء یک اجتماع خاص نیست. به عبارتی هرچه Qبزرگتر باشد یعنی اجتماع های شناسائی شده بهتر هستند.با این معیار مساله پیدا کردن بهترین انجمن تبدیل به مساله پیدا کردن مقدار بهینه ی Q شد.

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_hard است. نیومن یک الگوریتم حریصا نه 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. [i] کلازت [۴]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. [ii]

تعداد بسیار زیادی از الگوریتم های مبتنی بربهینه سازی مانندCNM ,بهینه سازی افراطی یا extremal ، بهینه سازی جستجوی گروهی والگوریتم ژنتیک مطرح شدند تا اینکه فرتناتو [۵]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. [iii]این قضیه به محدودیت **رزولوشن** معروف شد.

هدف این مقاله تقسیم گراف های بزرگ به انجمن هایی است که بر اساس معیار مادولاریتی کیفیت خوبی داشته و محدودیت رزولوشن هم نداشته باشند.

الگوریتم معرفی شده در این مقاله CLA-netنام دارد که در ان کل شبکه به عنوان اتاماتای سلولی یادگیرنده نامنظم مدل میشود. اتاماتای سلولی یادگیرنده CLAیک مدل ریاضی قوی برای بسیاری از مسایل غیر متمرکز وپدیده های پویا است.ایده اصلی اتاماتای سلولی یادگیرنده استفاده از اتاماتای یادگیری برای تنظیم احتمال انتقال حالت در اتاماتای تصادفی سلولی است.اتاماتای یادگیری سلولی را می توان نوعی از اتاماتای سولی در نظر گرفت که در آآن 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 وبه طور موثر مشکل محدودیت رزولوشن در بهینه سازی ماژولار را حل کند.

الگوریتم روی شبکه های مصنوعی و واقعی اعمال شده است.معیار های ارزیابی نتایج این الگوریتم مادولاریتی واطلاعات مشترک نرمال سازی شده، NMI است. همانطور که بیان شد معیار مادولاریتی Qاهمیت اجتماع را در شبکه بررسی می کند.مقدار بزرگترqنشان می دهد کیفیت انجمن ها بهتر ودورتر از حالت تصادفی مورد انتظار هستند.NMI مخصوص 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 NMIبین [0، 1] است و مقدار بزرگتر نشان می دهد انجمن های به دست آمده مطابقت بیشتری با انجمن های واقعی دارند.

---

[4] Clauset

[5] Fortunato and Barth´elemy

[i] M. E. J. Newman, Fast algorithm for detecting community structure in networks, Phys. Rev. E 69(2004)066133.

[ii] A. Clauset, M. E. J. Newman, C. Moore, Finding community structure in very large networks, Phys. Rev. E 70(2004)066111.

[iii] S. Fortunato, M. Barth´elemy, Resolution limit in community detection, Proc. Natl. Acad. Sci. USA 104(2007)36-41.