به نام خداوند جان و خرد

دانشگاه تهران

دانشکدگان فنی

دانشکده مهندسی برق و کامپیوتر

شبکه‌های اجتماعی

تمرین شماره ۳

نام و نام خانوادگی: علی خرم فر

شماره دانشجویی: ۸۱۰۱۰۲۱۲۹

آذرماه ۱۴۰۲

فهرست مطالب

فهرست اشکال

فهرست روابط

۱_ مقدمه‌ای بر فعالیت‌های Girvan-Newman

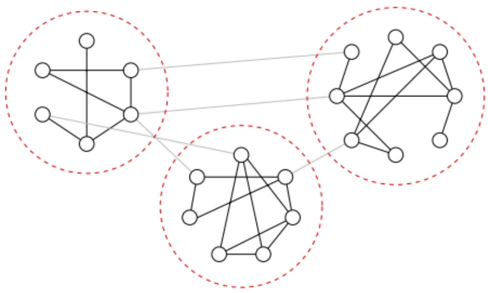
در درس شبکه‌های اجتماعی به بررسی Community یا انجمن‌هایی پرداختیم که چگالی ارتباط داخلی آن‌ها بیشتر از ارتباط خارجی آن‌هاست. در ادامه به بررسی یکی از روش‌های ساخت گراف با ساختار انجمن و همچنین بررسی‌های انجام شده برای تشخیص انجمن توسط این دو نویسنده می‌پردازیم.

۱-۱_ ساختار انجمن در شبکه‌های اجتماعی

در مقاله‌ی (Girvan & Newman, 2002)، به بررسی ساختار انجمن‌ها در شبکه‌های اجتماعی و زیست‌شناسی پرداخته شده‌است و در این پژوهش روشی جدید برای تشخیص ساختار انجمن‌ها معرفی کرده‌اند.نتایج آزمایشات انجام شده توسط نویسندگان این مقاله نشان می‌دهد که این روش با موفقیت بالا، قابلیت شناسایی ساختار انجمن را دارد. آن‌ها از این روش بر روی گراف‌های کامپیوتری تولید شده و همچنین بر روی شبکه‌های واقعی تست کرده‌اند و نتایج خود را با عالی نشان داده‌اند. این مقاله همچنین دو نمونه کاربرد الگوریتم را بر روی شبکه‌هایی با ساختار ناشناخته ارائه داده و نشان داده است که الگوریتم توانایی استخراج انجمن‌های واضحی را دارد.

۲-۱_ یافتن ساختار انجمن در شبکه‌ها

در شکل زیر یک نمونه شبکه کوچک با ساختار جامعه از نوع در نظر گرفته شده در این مقاله را مشاهده می‌کنیم. در این شکل سه انجمن وجود دارد که با دایره‌های خطچین نشان داده می شوند که دارای لینک‌های داخلی با چگالی بالا هستند اما بین آنها چگالی کمتری از لینک‌های خارجی وجود دارد.



شکل ۱ یک نمونه شبکه کوچک با ساختار کامیونیتی

در مقاله (Newman & Girvan, 2004) یک روش جدید برای تجزیه و تحلیل ساختار انجمن در شبکه‌ها مورد بحث قرار گرفته است. این مقاله به بررسی روش‌های مختلف برای یافتن انجمن‌ها در شبکه‌ها می‌پردازد و عملکرد این روش‌ها را ارزیابی می‌کند. تمرکز اصلی بر روی یک روش مبتنی بر ارتباط بین یال‌ها

برای شناسایی حاشیه‌ها و جداسازی انجمن‌ها است. نتایج این روش نشان می‌دهد که این روش با موفقیت ساختار انجمن را در شبکه‌ها شناسایی می‌کند.

۳-۱_ تولید گراف با روش Girvan-Newman

در مقالاتی که خلاصه آن‌ها را مرور کردیم برای سنجش میزان ارزیابی الگوریتم‌های ارائه‌شده با هدف شناسایی انجمن‌ها، از یک شبیه‌سازی برای ساخت گراف با کنترل بر روی انجمن آن استفاده شده‌است. نویسندگان مقاله برای شبیه‌سازی و سنجش الگوریتم خود، یک الگوریتم برای ساخت گراف ارائه دادند که خلاصه آن را در درس شبکه‌های اجتماعی مطالعه کردیم. یکی از مدل‌های گراف رندم، مدل Erdős–Rényi بود که برای تشکیل یال بین دو گره از احتمالات بهره می‌گرفت. برای مدل از نوع G(N,p) در گراف رندم یک عدد p در نظر گرفته می‌شود و برای N گره مراحل زیر انجام می‌شود.

۱. یک جفت گره را انتخاب کن. یک عدد تصادفی بین ۰ و ۱ انتخاب کن. اگر عدد از p کمتر شد بین آن‌ها لینک برقرار کن.
۲. مرحله ۱ را برای تمام $\frac{N(N-1)}{2}$ جفت گره تکرار کن.

از ویژگی‌های مثبت مدل Erdős–Rényi بهره‌مندی از خاصیت Small World بود که به دلیل وجود فاصله متوسط فاصله از مرتبه لگاریتمی در این مدل بود. یکی از نکات منفی که در این مدل وجود دارد سختی تشخیص انجمن‌ها در شبکه است. به این منظور برای نویسندگان مقالات مذکور برای اینکه یک مدل نزدیک به Erdős–Rényi داشته و از طرفی بر انجمن‌های تولیدشده نیز کنترل داشته باشند از روش زیر برای تولید گراف استفاده کردند. روش Girvan-Newman مانند ترکیب چند مدل گراف رندم است.

همانند مدل Erdős–Rényi ابتدا N گره در نظر گرفته و علاوه بر آن به تعداد C انجمن تعیین می‌شود. و سپس مراحل برای تمامی گره‌ها تکرار می‌شود:

۱. یک جفت گره انتخاب کن. اگر که این دو جفت در یک انجمن بودند با احتمال P_{in} بین آن‌ها یال لینک برقرار کن. در غیر این صورت به مرحله ۲ برو.
۲. این دو جفت در یک انجمن نیستند. بین آن‌ها با احتمال P_{out} لینک برقرار کن.

در این روش N نود و C کامیونتی داریم. پس هر انجمن $\frac{N}{C}$ نود دارد. همچنین به معرفی پارامتر P_{in} نمایانگر احتمال ارتباط داخل انجمن و P_{out} احتمال ارتباط یک نود با یک انجمن خارجی است.

۲_تحلیل پارامترهای شبکه با ساختار انجمن

در ادامه به بررسی تحلیلی مقادیر مختلف درخواست‌شده در سوال، در این گراف می‌پردازیم. در این سوال فرض شده که مقدار P_{in} بسیار بزرگتر از مقدار P_{out} باشد.

۱-۲_ درجه متوسط نودهای گراف

درجه متوسط نودهای داخلی هر انجمن

فرض کنیم که گراف یک گراف رندم باشد که احتمال وجود هر لینک بین هر یک جفت از نودها برابر با p است. بنابراین، متوسط درجه هر نود در این گراف برابر با (N-1)p خواهد بود. برای هر نود در داخل یک انجمن، تعداد لینک‌های داخلی آن با سایر نودهای داخل انجمن برابر با (N/C-1) خواهد بود.

اگر احتمال وجود هر یال درون انجمن P_{in} باشد، متوسط درجه داخلی یک نود در انجمن برابر با (N/C-1)p خواهد بود.

$$< K_{in} > = \left(\frac{N}{C} - 1\right) P_{in}$$

درجه متوسط ارتباطات خارجی

برای هر نود در خارج از یک انجمن، تعداد ارتباطات خارجی آن با نودهای انجمن‌های دیگر برابر با (N - N/C) خواهد بود. اگر احتمال وجود هر یال بین یک نود و یک انجمن خارجی P_{out} باشد، میانگین درجه خارجی یک نود از انجمن برابر با P_{out} (N - N/C) خواهد بود.

$$< K_{out} > = \left(N - \frac{N}{C}\right) P_{out}$$

درجه متوسط نودهای کل گراف

$$< K > =< K_{in} > + < K_{out} > = \left(\frac{N}{C} - 1\right) P_{in} + \left(N - \frac{N}{C}\right) P_{out} =$$

$$\left(\frac{N}{C}\right) P_{in} - 1P_{in} + (N)P_{out} - \left(\frac{N}{C}\right) P_{out} = \left(\frac{N}{C}\right) (P_{in} - P_{out}) - P_{in} + (N)P_{out} =$$

$$N \left(\frac{P_{in} - P_{out}}{C} + P_{out}\right) - P_{in}$$

برای تحلیل مقدار متوسط درجه از آن حد گرفته و برای N به سمت بی‌نهایت تحلیل می‌کنیم توجه شود که $P_{in} \gg P_{out}$:

$$\lim_{N\to\infty}(<\boldsymbol{K}>) = \lim_{N\to\infty}\left(N\left(\frac{P_{in}-P_{out}}{C}+P_{out}\right)-P_{in}\right) = \infty$$

حال وقتی N به سمت بی‌نهایت برود درحالی که P_{in} بسیار بزرگتر از P_{out} است، مقدار متوسط درجه نیز افزایش خواهدیافت و به سمت بی‌نهایت میل می‌کند که این مورد در این گراف مشابه با مدل Erdős–Rényi خواهدبود. یعنی عملا N با P_{in} رابطه مستقیم دارد. باتوجه به رابطه بالا هرچه که اختلاف احتمال تشکیل نود داخلی از خارجی بیشتر باشد، متوسط درجه بر اساس تعداد کلاسترها افزایش می‌یابد. در رابطه با پارامتر P_{out} این رابطه هم اثر مستقیم با افزایش N دارد. دلیل این مورد این است که وقتی تعداد نودها زیادشود، تمام نودهای دیگر احتمال برقراری لینک با آن را دارند.

ممکن است تصور کنیم که با افزایش N تاثیر مقدار P_{in} بر افزایش متوسط درجه بسیار بیشتر از P_{out} است. اما باید در نظر داشت که در فرمول بالا، یک فاکتور تاثیرگذار C نیز وجود دارد. در صورتی که انجمن‌هایی که در این گراف تشکیل شده‌اند تعدادشان نسبت به N زیاد باشد، در شبکه حاصل تعداد زیادی کلاستر خواهیم داشت که تعداد گره‌های هرکدام از آن‌ها بسیار کم خواهدبود. پس در این صورت هرکدام از آن‌ها تعداد کمتری لینک نسبت به لینک‌های خارجی خواهند داشت و تاثیر P_{in} بر متوسط درجه ناچیز خواهدبود. پس اینکه فرض کنیم مقدار P_{out} نسبت به P_{in} ناچیز است، پس تاثیر آن نیز کم خواهدبود، باور غلطی است.

۲-۲_ توزیع درجه نودهای گراف

در این قسمت به دنبال توزیع درجه نودهای گراف هستیم که به معنی این است مقدار PMF توزیع را پیدا کنیم. سپس احتمال اینکه یک نود انتخابی درجه برابر با K داشته باشد، با کمک آن قابل محاسبه است. در مدل Erdős–Rényi توزیع زیر را برای درجه نود های گراف داریم که همان توزیع پواسون و برای راحتی بیشتر از محاسبات به عنوان جایگزین توزیع دوجمله‌ای استفاده می‌شود و در درس اثبات شد که در N های بزرگ این توزیع رفتاری مشابه با توزیع دوجمله‌ای دارد.

رابطه ۱ توزیع درجه در مدل Erdős–Rényi

$$P(K) = \frac{e^{-\lambda}\lambda^K}{K!} = \frac{e^{-<K>}<K>^K}{K!}$$

نکته‌ی مهمی که درباره مدل جدیدی که در قصد تحلیل آن را داریم وجود دارد این است که از جمع دو توزیع پواسون مستقل از هم تشکیل شده‌است. و از آنجا که جمع دو توزیع پواسون نیز خواهدبود، سعی داریم که این مورد را اثبات کنیم.

در مدل Erdős–Rényi توزیع درجه تمامی نودها مشابه یکدیگر است و در این مدل برای تولید اعداد تصادفی با هدف برقراری لینک بر حسب p از توزیع یکنواخت استفاده می‌شود. پس تمامی نودها در گراف از توزیع درجه یکسانی پیروی می‌کنند. (Kaiser, 2011)

حال قصد داریم این موضوع را در مباحث آماری استفاده کنیم و تابع توزیع برای مدل Girvan-Newman به عنوان یک تابع توزیع توام در نظربگیریم. احتمال درجه برابر با K در هر یک از نودهای گراف به دو عامل بستگی دارد. یکی P_{in} و دیگری P_{out} . حال قصد داریم احتمال در هر گره را درنظر گرفته و این احتمال را برای تمامی گره‌ها حساب کنیم تا تابع توزیع آن بدست آید. همانطور که گفته شد پیشامد آنکه یک گره درجه برابر K داشته باشد، یک کسر از این K مربوط به لینک‌های درون انجمن و متمم آن برابر با لینک‌های بیرون از انجمن خواهدبود. هرچند که این موضوع احتمالاتی است و ممکن است تمامی لینک‌های یک گره مربوط به داخل انجمن و یا برعکس باشد. پس اگر تعداد تمامی لینک‌های یک گره برابر K بوده و تعداد لینک‌های درون یک انجمن برای یک آن گره برابر j باشد، به تعداد K-j لینک خارج از انجمن نیز برای آن گره برقرار خواهدبود. پس پیشامد $P(K_{in}=j, K_{out}=K-j)$ را برای تمامی Kهای ممکن یک گره جمع میزنیم تا تابع توزیع بدست آید.

$$P(K)=\sum_{J=0}^{K} P(K_{in}=j, K_{out}=K-j)$$

در حالت استقلال، توزیع توام متغیرهای تصادفی برابر با ضرب توزیع های شرطی آنها است. به عبارت دیگر، برای هر مجموعه‌ای از مقادیر ممکن برای متغیرهای تصادفی، احتمال اینکه آن مقادیر رخ دهند برابر است با حاصل ضرب احتمال اینکه هر متغیر تصادفی به صورت مستقل با آن مقدار رخ دهد.

$$P(X=x, Y=y)=P(X=x)\ \times\ P(Y=y)$$

در نتیجه، باتوجه به اینکه درجه ورودی و درجه خروجی بر حسب P_{in} و P_{out} تعیین می‌شوند و هردوی این‌ها مستقل از هم هستند، توزیع توام آنها به صورت زیر تبدیل می شود:

$$P(K_{in}=j, K_{out}=K-j)=P(K_{in}=j)\ \times\ P(K_{out}=K-j)$$

پس رابطه اولیه به صورت زیر بازنویسی می‌شود:

رابطه ۲توزیع درجه به صورت ضرب دو توزیع

$$P(K) = \sum_{J=0}^{K} P(K_{in} = j, K_{out} = K - j) = \sum_{J=0}^{K} P(K_{in} = j) \times P(K_{out} = K - j)$$

باتوجه به رابطه 1 باتوجه به اینکه هرکدام از توزیع‌های بالا از مدل Erdős–Rényi پیروی کرده و در این سوال هدف ما بررسی برای N های بزرگ است، مقدار هرکدام از احتمالات را برابر با توزیع آن قرار می‌دهیم. پس داریم:

$$P(K_{in}) = \frac{e^{-<K_{in}>} < K_{in} >^{K_{in}}}{K_{in}!} \rightarrow P(K_{in} = j) = \frac{e^{-<K_{in}>} < K_{in} >^{j}}{j!}$$

و برای احتمال لینک‌های خارج از انجمن داریم:

$$P(K_{out}) = \frac{e^{-<K_{out}>} < K_{out} >^{K_{out}}}{K_{out}!} \rightarrow P(K_{out} = K - j) = \frac{e^{-<K_{out}>} < K_{out} >^{K-j}}{(K - j)!}$$

مقادیر فوق را در رابطه 2 جایگذاری کرده و رابطه را ساده‌تر می‌کنیم.

$$\sum_{J=0}^{K} P(K_{in} = j) \times P(K_{out} = K - j) = \sum_{J=0}^{K} \frac{e^{-<K_{in}>} < K_{in} >^{j}}{j!} \times \frac{e^{-<K_{out}>} < K_{out} >^{K-j}}{(K - j)!}$$

$$= \sum_{J=0}^{K} \frac{< K_{in} >^{j}}{j!} \times \frac{< K_{out} >^{K-j}}{(K - j)!} \times e^{-<K_{in}>-<K_{out}>}$$

$$= \sum_{J=0}^{K} \frac{< K_{in} >^{j} < K_{out} >^{K-j}}{j!\,(K - j)!} \times e^{-(<K_{in}>+<K_{out}>)}$$

$$= \sum_{J=0}^{K} \frac{1}{j!\,(K - j)!} \times < K_{in} >^{j} < K_{out} >^{K-j} \times e^{-(<K>)}$$

رابطه بالا به قضیه دوجمله‌ای شباهت دارد.

رابطه ۳قضیه دو جمله‌ای

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$$

که در این رابطه برای عدد طبیعی n ، $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ ضریب دوجمله‌ای است.

کافی است که برای تشکیل آن $e^{-(<K>)}$ را از سیگما خارج کنیم($<K>$ وابسته به اندیس سیگا نیست) و سپس یک مقدار K! در صورت و مخرج ضرب کنیم. مقدار K! وابسته به اندیس سیگما نبوده و قسمت مخرج را از سیگما خارج می‌کنیم.

$$\sum_{J=0}^{K} \frac{1}{j!\,(K-j)!} \times < K_{in} >^{j} < K_{out} >^{K-j} \times e^{-(<K>)}$$

$$= \frac{e^{-<K>}}{K!} \sum_{J=0}^{K} \frac{K!}{j!\,(K-j)!} < K_{in} >^{j} < K_{out} >^{K-j} = \frac{e^{-<K>}}{K!} \sum_{J=0}^{K} \binom{K}{j} < K_{in} >^{j} < K_{out} >^{K-j}$$

حال قسمت راست معادله را با رابطه 3 به نحوی که n=K و k=j و همچنین x و y به ترتیب برابر با $<K_{in}>$ و $<K_{out}>$ است جایگذاری می‌کنیم.

$$\frac{e^{-<K>}}{K!} \sum_{J=0}^{K} \binom{K}{j} < K_{in} >^{j} < K_{out} >^{K-j} = \frac{e^{-<K>}}{K!} (< K_{in} > + < K_{out} >)^{K}$$

$$\rightarrow P(K) = \frac{e^{-<K>} < K >^{K}}{K!}$$

که به توزیع پواسون با پارامتر $\lambda = < K >$ رسیدیم. این توزیع همانند توزیع درجه در مدل گراف رندم Erdős–Rényi برای N های بزرگ برقرار است که در این سوال هم فرض بر N به سمت بی‌نهایت لحاظ شده‌است.

۳-۲_ ضریب خوشه‌بندی نودهای گراف

ضریب خوشه‌بندی[1] یک معیار است که میزان اتصال همسایگان یک گره خاص در یک شبکه را اندازه‌گیری می‌کند. برای یک گره به نام i با درجه ki (تعداد یال‌های متصل به گره) و Li به عنوان تعداد لینک‌های واقعی بین همسایگان یک گره، ضریب خوشه‌بندی به صورت زیر تعریف می‌شود:

رابطه ۴ضریب خوشه‌بندی گراف رندم

$$C_i = \frac{L_i}{\frac{K_i(K_i-1)}{2}} = \frac{2L_i}{K_i(K_i-1)}$$

ضریب خوشه‌بندی بین ۰ تا ۱ متغیر است، به این صورت که:

- Ci=0 نشان‌دهنده این است که هیچ یک از همسایگان گره i به یکدیگر متصل نیستند.
- Ci=1 نشان‌دهنده این است که تمام همسایگان گره i به یکدیگر متصل هستند.

در گراف رندم Ci به p بستگی داشته و ارتباطی با اینکه گره‌های همسایه با گره مورد نظر i همسایه باشند یا خیر، ندارد. حال این ویژگی را در گراف‌های تولیدشده به روش Girvan-Newman بررسی می‌کنیم.

ابتدا به محاسبه Li می‌پردازیم که تعداد لینک‌های برقرارشده بین همسایگان یک گره است. در این مدل همسایگان گره i حالت‌های مختلفی بر حسب P_{in} و P_{out} دارند که باید تمامی آن‌ها بررسی و جمع شوند. برای بدست آوردن تعداد لینک‌های برقرارشده، ابتدا احتمال اینکه نوع لینکی که آن همسایه‌ها برقرار می‌کنند را بدست آورده و سپس در میزان احتمال آن حالت (ارتباط داخل انجمن و یا خارج از انجمن) در P_{in} ویا P_{out} ضرب شده و مجموع آن‌ها برابر با Li خواهدبود.

احتمال هم‌انجمن بودن هردو همسایه

اگر دو گره انتخابی در یک انجمن قرار داشته باشند، احتمال این اتفاق برابر است انتخاب یک انجمن از کل انجمن‌ها به تعداد C ، و احتمال اینکه گره اول و گره دوم در یک انجمن باشند که برابر است با:

رابطه ۵ احتمال هم‌انجمن بودن هردو همسایه

$$P_{Same\ Community} = \binom{C}{1} \times \frac{1}{C} \times \frac{1}{C} = \frac{C!}{1(C-1)!} \times \frac{1}{C^2} = C \times \frac{1}{C^2} = \frac{1}{C}$$

---

[1] Clustering coefficient

احتمال غیرهم‌انجمن بودن دو همسایه

بدیهی است که یا دو همسایه انتخابی، در یک انجمن قرار دارند یا ندارند. پس احتمال اینکه همسایه‌ها در انجمن های متفاوتی باشند برابر مکمل این است که در یک انجمن باشند:

رابطه ۶ احتمال غیرهم‌انجمن بودن دو همسایه

$$P_{Different\ Community} = 1 - P_{Same\ Community} = 1 - \frac{1}{C} = \frac{C-1}{C}$$

حال برای بدست آوردن احتمال لینک‌های بین همسایگان، احتمال هم‌انجمن بودن را در $\mathrm{P_{in}}$ و احتمال غیرهم‌انجمن بودن را در $\mathrm{P_{out}}$ ضرب کرده و باهم جمع می‌کنیم.

$$P_{Link\ Between\ Neighbors} = \mathrm{P_{in}}\left(\frac{1}{C}\right) + \mathrm{P_{out}}\left(\frac{C-1}{C}\right)$$

برای بدست آوردن تعداد لینک‌های بین همسایگان، احتمال برقراری لینک بین هردو همسایه انتخابی را در تعداد جفت همسایه‌های ممکن ضرب می‌کنیم. می‌دانیم که درجه هر گره برابر K است و این عدد به معنی این است که هر گره K همسایه دارد. پس انتخاب ۲ از K یعنی انتخاب هر دو همسایه از K که تعداد تمامی همسایگان یک گره است.

$$L_i = \binom{K_i}{2}\left(\mathrm{P_{in}}\left(\frac{1}{C}\right) + \mathrm{P_{out}}\left(\frac{C-1}{C}\right)\right)$$

حال این مورد را در رابطه 4 جایگذاری می‌کنیم.

$$C_i = \frac{L_i}{\frac{K_i(K_i-1)}{2}} = \frac{L_i}{\binom{K_i}{2}} = \frac{\binom{K_i}{2}\left(\mathrm{P_{in}}\left(\frac{1}{C}\right) + \mathrm{P_{out}}\left(\frac{C-1}{C}\right)\right)}{\binom{K_i}{2}} = \mathrm{P_{in}}\left(\frac{1}{C}\right) + \mathrm{P_{out}}\left(\frac{C-1}{C}\right)$$

حال از مقدار ضریب خوشه‌بندی نسبت به N به سمت بی‌نهایت حد می‌گیریم:

$$\lim_{N\to\infty}(C_i) = \lim_{N\to\infty}\left(\mathrm{P_{in}}\left(\frac{1}{C}\right) + \mathrm{P_{out}}\left(\frac{C-1}{C}\right)\right) = \mathrm{P_{in}}\left(\frac{1}{C}\right) + \mathrm{P_{out}}\left(\frac{C-1}{C}\right)$$

هیچ‌کدام از پارامتر ها به N وابسته نیست پس این مقدار در N به سمت بی‌نهایت ثابت می‌ماند.

باتوجه به رابطه بدست‌آمده ضریب خوشه‌بندی، متوجه می‌شویم که این مدل از گراف نیز دارای خاصیت *Clustering Coefficient* بالاست. دلیل این موضوع $\mathrm{P_{in}}$ بالاست و هرچه $\mathrm{P_{in}}$ بالاتر باشد، مقدار C_i هم بالاتر می‌رود.

یک حالتی هم وجود دارد که در صورتی که تعداد C بالا باشد، تاثیر $\mathrm{P_{in}}$ خیلی کم می‌شود. به دلیل اینکه در این حالت کلاسترها بسیار زیاد هستند و تعداد گره‌های در هر کلاستر خیلی کم خواهدبود. که در رابطه هم واضح است که این دو پارامتر با یکدیگر رابطه عکس دارند. پس در این حالت استثنا خاصیت *Clustering Coefficient* کم خواهدبود. هرچند که این شرایط زمانی حاصل می‌شود که مقدار $\mathrm{P_{out}}$ را پایین فرض کرده‌باشیم و فقط بحث اختلاف بین $\mathrm{P_{in}}$ و $\mathrm{P_{out}}$ مطرح نباشد.

۲-۴_ متوسط فاصله نودهای گراف

متوسط فاصله ، با نماد $\langle \mathrm{d} \rangle$، میانگین فاصله بین تمام جفت گره در یک شبکه است. برای یک شبکه با N گره، میانگین مسافت معمولاً به صورت زیر محاسبه می‌شود:

$$< d > = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} d(i,j)}{N(N-1)}$$

فرمول شامل جمع کردن طول‌های کوتاهترین مسافت‌ها بین تمام جفت گره‌هاست و سپس تقسیم بر تعداد کل جفت‌های گره. می‌توان فاصله متوسط را از طریق دیگر نیز حساب کرد. باتوجه به اینکه تمامی گره‌ها در این مدل از توزیع یکسانی استفاده کرده و درجه آن‌ها وابسته به اعداد تصادفی یکنواخت است می‌توانیم فاصله متوسط یک گره را تمامی دیگر گره‌ها محاسبه کرده و آن را به متوسط کل تعمیم دهیم.

رابطه ۷فاصله متوسط بین دو گره

$$< d > = < d_i > \ = \frac{\sum_{j=1}^{N} d(i,j)}{(N-1)}$$

حال باید فاصله بین گره انتخابی i با تمامی گره‌ها محاسبه شود. همانند قسمت قبل برای محاسبه این فاصله‌ها مسئله را به دو زیرمسئله تقسیم می‌کنیم.

اگر گره i و گره j در یک انجمن باشند، تعداد گره‌های موجود در هر کلاستر(تعداد گره‌های گراف رندم تشکیل شده به عنوان انجمن) به جز خود گره را در متوسط فاصله ضرب می‌کنیم. باتوجه به اینکه هر یک از انجمن‌ها یک گراف رندم مدل Erdős–Rényi است و در این گراف متوسط فاصله از مرتبه لگاریتمی بوده $(\frac{\log(N)}{\log(K)})$ و ویژگی جهان کوچک برقرار است، این متوسط فاصله را در احتمال اولیه ضرب می‌کنیم تا متوسط فاصله برای ارتباطات درون انجمنی گراف محاسبه شود:

رابطه ۸متوسط فاصله برای ارتباطات درون انجمنی

$$< d_{intra\ community} > = \left(\frac{N}{C} - 1\right) \times \ln\left(\frac{N}{C}\right)$$

اگر گره i و گره j در یک انجمن نباشند، قضیه متفاوت می‌شود. ما باید فاصله متوسط بین دو گره در در گراف تشکیل شده به روش Girvan-Newman را محاسبه کنیم. این فاصله از ۳ زیرفاصله تشکیل شده‌است. یک مسیر اولیه برای طی مسیر تا لبه‌ی انجمن، یک مسیر میان دو انجمن و در آخر نیز مسیر نهایی از لبه انجمن تا گره هدف قرار دارد. مسیر اولیه و نهایی از مرتبه $\ln\left(\frac{N}{C}\right)$ است که دلایل آن در قسمت قبل ذکر شد. مسیر میانی بستگی به وضعیت گراف دارد. اگر که بین هر دو انجمن یالی وجود داشته باشد، بدیهی است که فاصله برابر ۱ خواهدبود. اگر که یالی نباشد علاوه بر اضافه شدن یک مسیر میانی تا انجمن دیگر، یک مسیر درون انجمن هم از مرتبه لگاریتمی خواهیم داشت که البته این مورد تاثیر زیادی ندارد زیرا که عدد C یک ثابت است و باز هم مسیر از مرتبه لگاریتمی خواهدبود.هرچند که در این سوال فرض شده که N های بزرگ بررسی شوند.پس اگر که مقدار N به سمت بی‌نهایت میل کند، احتمال اینکه دو انجمن به هم متصل نباشند بسیار کم است و البته مقدار C هم تاثیرگذار است. پس به طور کلی فاصله بین انجمنی را ۱ فرض می‌کنیم. هرچند که احتمال لگاریتمی شدن آن نیز وجود دارد ولی این مورد استثنا، تاثیری روی جمع با مرتبه لگاریتمی نداشته و به طور کلی جمع هر ۳ زیرفاصله از مرتبه لگاریتمی خواهدبود.

رابطه ۹متوسط فاصله برای ارتباطات بین انجمنی

$$< d_{inter\ community} > = \left(N - \frac{N}{C}\right) \times \left(\ln\left(\frac{N}{C}\right) + 1 + \ln\left(\frac{N}{C}\right)\right)$$

پس مقدار متوسط فاصله بین یک گره و دیگر گره‌ها و به طور کلی فاصله متوسط برابر با جمع فاصله متوسط درون انجمنی و بین انجمنی تقسیم بر تعداد کل این روابط خواهدبود:

$$< d > = < d_i > = \frac{\sum_{j=1}^{N} d(i,j)}{(N-1)} = \frac{\left(\frac{N}{C} - 1\right) \times \ln\left(\frac{N}{C}\right) + \left(N - \frac{N}{C}\right) \times \left(\ln\left(\frac{N}{C}\right) + 1 + \ln\left(\frac{N}{C}\right)\right)}{(N-1)}$$

$$= \frac{\left(\frac{N}{C} - 1\right) \times \ln\left(\frac{N}{C}\right) + N\left(1 - \frac{1}{C}\right) \times \left(2\ln\left(\frac{N}{C}\right) + 1\right)}{(N-1)}$$

حال برای تحلیل حالتی که N به سمت بی‌نهایت میل می‌کند از آن حد می‌گیریم:

$$\lim_{N\to\infty} (< d >) = \lim_{N\to\infty} \left(\frac{\left(\frac{N}{C} - 1\right) \times \ln\left(\frac{N}{C}\right) + N\left(1 - \frac{1}{C}\right) \times \left(2\ln\left(\frac{N}{C}\right) + 1\right)}{(N-1)} \right) =$$

برای N های بزرگ کسر ۱ عدد تاثیری ندارد پس داریم:

$$\lim_{N\to\infty} \left(\frac{N\left(\frac{1}{C}\right) \times \ln\left(\frac{N}{C}\right) + N\left(1 - \frac{1}{C}\right) \times \left(2\ln\left(\frac{N}{C}\right) + 1\right)}{N} \right) =$$

$$\lim_{N\to\infty} \left(\left(\frac{1}{C}\right) \times \ln\left(\frac{N}{C}\right) + \left(1 - \frac{1}{C}\right) \times \left(2\ln\left(\frac{N}{C}\right) + 1\right) \right) = O\left(\ln\left(\frac{N}{C}\right)\right)$$

پس متوسط فاصله کل در این نوع گراف نیز از مرتبه لگاریتمی و بر حسب N بوده و باتوجه به فاصله متوسط لگاریتمی، این نوع گراف خاصیت Small World نیز دارد.

۳_ نتیجه‌گیری

در این تمرین با روش جدیدی برای تولید گراف با کنترل بر روی کلاسترها آشنا شدیم که از مدل گراف رندم Erdős–Rényi الهام گرفته شده‌است. این گراف با حفظ ویژگی‌های مثبت گراف رندم، ویژگی‌های دیگری نیز ارائه میدهد. با بررسی‌هایی که بر روی ویژگی‌های مختلف این روش برای تولید گراف انجام شد متوجه شدیم که درجه متوسط این گراف نیز مانند گراف رندم، با افزایش N ، افزایش خواهد یافت که موارد موثر بر روی آن از جمله P_{in} ویا P_{out} و حالت خاص آن بررسی شد. توزیع درجه در این مدل گراف نیز مانند گراف Erdős–Rény از نوع پواسون است. همچنین متوجه شدیم این گراف دارای ضریب خوشه‌بندی بالاست و حالت استثنای آن نیز بررسی شد. و در آخر نیز به بررسی فاصله متوسط پرداخته‌شد که با بررسی کوتاه‌ترین مسیرهای ممکن بین دو گره و محاسبه متوسط فاصله متوجه شدیم خاصیت دنیای کوچک نیز در این مدل برقرار است. هرچند که این گراف نیز همانند مدل Erdős–Rény تفاوت‌هایی با گراف‌های دنیای واقعی دارند ولی برای مواردی از جمله الگوریتم‌های تشخیص انجمن کاربرد دارند.

۴_مراجع

Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, *99*(12), 7821-7826 .

Kaiser, M. (2011). A tutorial in connectome analysis: topological and spatial features of brain networks. *Neuroimage*, *57*(3), 892-907 .

Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, *69*(2), 026113 .