گذار های فاز در SBM ها

کورش صدری
دانشجوی کارشناسی مهندسی برق
دانشگاه صنعتی شریف

۵ مرداد ۱۴۰۰

چکیده

در این گزارش، تلاش میکنیم به مطالعه ی نظری گذار های فاز در مساله ی تشخیص جماعت در گراف های تولید شده با استفاده از مدل بلوک های تصادفی بپردازیم. تمرکز اصلی ما بر روی رژیم مربوط به گراف های با میانگین درجه ی ثابت خواهد بود که در مرجع [۶] بعنوان مساله ی باز مطرح شده است. نشان میدهیم حدس گاوسی معرفی شده، در این رژیم برقرار نخواهد بود. در ادامه و به کمک شبیه سازی های عددی، نشان میدهیم که الگوریتم طیفی برای گراف های با اندازه ی متوسط، عملکرد بهینه دارد.

واژگان کلیدی: گذار فاز، الگوریتم طیفی

فهرست مطالب

۱ مقدمه

گراف های تصادفی در دهه های اخیر، مورد توجه بسیاری از پژوهشگران قرار گرفته اند. همچنین کاربردهای عملی جالبی برای این مدل ها در شاخه های مختلف علوم از حوزه ی یادگیری ماشین به کمک شبکه های عصبی تا نظریه ی میدان های کوانتومی پیدا شده است. ساده ترین مدل برای یک گراف تصادفی، در ابتدای دهه ی ۶۰ میلادی، توسط اردوش و رنیی ارائه شد. [۱] در این مدل، ماتریس مجاورت (برای گراف های بدون جهت و ساده) دارای درایه های مستقل و همتوزیع با پارامتر مشخص p است. راه حل های هوشمندانه ای برای پاسخ به پرسش های اغلب مجانبی [۱] درباره ی گراف های بسیار بزرگ از آنسامبل اردوش–رنیی ارائه شده اند. آنچه که عموماً در تعیین رفتار یک کمیت مجانبی اهمیت پیدا میکند، نرخ رشد پارامتر p با تعداد رئوس گراف N است. برای مثال نشان داده شده است که در رژیم $p \sim \log N$ گراف های اردوش رنیی، با احتمال بالا، همیلتونی هستند. [۲]

آنسامبل اردوش–رنیی را میتوان با افزودن ساختار جماعت [۲] تعمیم داد که در این حالت، مدل را، مدل بلوکی تصادفی یا SBM مینامیم. در این جا، هر کدام از راس ها یک نوع [۳] مشخص دارد و احتمال وجود یک یال بین دو راس، مستقل از دیگر راس ها و به صورت متقارن به نوع راس ها بستگی دارد. در ساده ترین حالت، میتوان گرافی شامل دو جماعت با علامتهای + و - را در نظر گرفت به طوریکه احتمال حضور یک یال بین دو راس همعلامت، با مقدار p و احتمال حضور یک یال بین دو راس مختلف العلامت، با مقدار q برابر باشند. همانطور که در آنسامبل اردوش–رنیی، تعداد رئوس، N ، بعنوان یک فراپارامتر در نظر گرفته شده است ، در مدل های بلوکی تصادفی نیز، تعداد و نوع رئوس فراپارامتر هایی هستند که باید از روی یک احتمال پیشینی تعیین شوند.

علاوه بر پرسش های مجانبی مشابه با آنسامبل اردوش–رنیی، یک مساله که به طور خاص درباره ی SBM ها مطرح میشود، مساله ی تشخیص جماعت است: آیا الگوریتمی وجود دارد که با مشاهده ی یک گراف داده شده، نوع رئوس گراف را بدرستی تخمین بزند؟ در حدود ۲۰ سال گذشته، کاربردهای فراوانی برای این مساله، از جمله در تبلیغات هوشمند، مدلسازی همه گیری ها، فعالیت های ضدتروریستی، تشخیص وبسایت های قانون شکن و ... معرفی شده اند. [۳]

برای ارزیابی عملکرد یک الگوریتم مشخص در حل مساله ی تشخیص جماعت، معیار های

---

[۱] Asymptotic
[۲] Community
[۳] Type

زیادی تعریف شده اند، متداول ترین آنها، معیار توافق [۴] است. [۴]

$$A \equiv \max_{\pi} \frac{1}{N} \sum_{i} \mathbb{I}\big[X_i = \pi(\hat{X}_i)\big] \qquad (۱)$$

که در این تعریف، N تعداد رئوس، X_i نوع راس i ام، و $\hat{X}_i$ تخمین الگوریتم از نوع راس i پس از مشاهده ی گراف هستند. با بهینه سازی روی جایگشت های مختلف از نامگذاری نوع راس ها، π ، در واقع تنها به تفکیک صحیح جماعت های مختلف از یکدیگر توجه میکنیم.

در طول سالها، الگوریتم های متنوعی برای حل این مساله در رژیم های مختلف پیشنهاد شده اند. در بعضی از رژیم ها، عملکرد های بهینه از لحاظ نظری محاسبه شده اند و همچنین الگوریتم های بهینه نیز ارائه شده اند. [۴] در نگاه اول، ممکن است به نظر برسد که مانند یک مساله ی تخمین آماری در آمار پارامتری، استفاده از یک رهیافت مبتنی بر درست نمایی [۵] میتواند عملکرد بهینه ی مجانبی داشته باشد. باید دقت داشت که در حد گرافهای بزرگ و با افزایش تعداد رئوس، همانطور که میزان اطلاعات دریافتی افزایش پیدا میکند، تعداد پارامتر های مورد تخمین نیز افزایش پیدا میکند. بعبارت دقیقتر، مساله ی تشخیص جماعت، یک مساله ی تخمین تک–اندازه گیری [۶] محسوب میشود و بنابراین استفاده از درست نمایی بهینه نخواهد بود. [۵]

مانند بسیاری از مدل های احتمالاتی و در حد N بزرگ، عملکرد الگوریتم های مختلف در رژیم های مختلف، میتوانند دستخوش گذار های فازی، به معنای ترمودینامیکی شوند. [۵] این بدان معناست که مقدار مجانبی کمیت های ماکروسکوپی مانند توافق، با تغییر پارامتر های مساله، به صورت غیر تحلیلی تغییر میکنند. از آنجاییکه کمیت توافق، علاوه بر پارامتر های مساله، به الگوریتم مورد استفاده برای تشخیص جماعت نیز بستگی دارد، درباره ی جهان شمول بودن یک گذار مشاهده شده، ابهام بوجود میآید. آیا شکست ناکهانی یک الگوریتم در یک بازه ی خاص پارامتری، ناشی از ضعف همین الگوریتم خاص است یا میان تمام الگوریتم های قابل استفاده عمومیت دارد؟

یک راه برای رهایی از این ابهام و پیدا کردن گذار های جهان شمول، تمرکز بر روی کمیت های ماکروسکوپیکی است که مستقل از الگوریتم هستند و در عین حال میتوانند تضمین کننده ی موفقیت یا شکست در حل مساله ی تشخیص جماعت باشند. با در نظر گرفتن فرایند نوفه ای [۷] تولید یک گراف از روی داده های مربوط به نوع رئوس، بعنوان یک کانال مخابراتی، میتوان نتیجه گرفت که کمیت اطلاعات متقابل بر راس میتواند چنین نقشی را به خوبی ایفا کند.تمرکز بر روی محاسبه ی تحلیلی این کمیت در پی یک مبنای نظری مستحکم برای توجیه گذار های

---

[۴] Agreement
[۵] Likelihood
[۶] Single-shot
[۷] Noisy

مشاهده شده، روند متاخری در تحقیقات مربوطه است. [۴] به طور خاص و در سال ۲۰۱۵ میلادی، این کمیت برای مدل های دارای دو جماعت هم اندازه (متقارن)، و در رژیم هایی که در آن ها، میانگین درجه ی هر راس بدون کران رشد میکند، محاسبه شده و گذار های مشاهده شده ی متناظر، بازشناسایی و بازیابی شده اند. [۶، ۴]

تمرکز ما در ادامه ی این گزارش، بر مدل های دو جماعتی متقارن با میانگین درجه ی ثابت خواهد بود. ابتدا عبارتی (غیر تحلیلی) برای اطلاعات متقابل پیدا خواهیم کرد، سپس به پیامدهای آن میپردازیم و در ادامه گذار های الگوریتم طیفی در این رژیم را بررسی میکنیم.

۲ اطلاعات متقابل بر راس

۱.۲ طرح مساله

در این پروژه، یک گراف تصادفی N راسی را در نظر میگیریم که از یک آنسامبل بلوکی تصادفی با دو جماعت متقارن تولید شده است. اندازه ی احتمال پیشینی برای تولید نوع راس ها، بصورت متقارن و مستقل در نظر گرفته میشود. بعبارت دیگر، N متغیر تصادفی دوحالته X_i به صورت مستقل وهمتوزیع از توزیع یکنواخت روی مجموعه ی $\{\pm 1\}$ تولید شده اند. سپس ماتریس مجاورت (G) مربوط به گراف ساده ی بدون جهت، با درایه ها (یال ها) ی مستقل از هم و با توجه به نوع رئوس تولید میشود. به طوریکه احتمال حضور یک یال بین دو راس متفاوت i و j با رابطه ی $(\alpha + \beta X_i X_j)$ داده میشود. [۸] در نهایت، یک الگوریتم با مشاهده ی ماتریس مجاورت G اقدام به بازسازی و تخمین علامت های X_i میکند.

$$\mathbf{X} \longrightarrow G \longrightarrow \hat{\mathbf{X}} \tag{۲}$$

مطابق مرجع [۴] میگوییم یک الگوریتم، بازیابی جزیی [۹] انجام میدهد اگر داشته باشیم

$$\mathbb{P}\big[A \geq \frac{1+t}{2}\big] = 1 - o(1) \tag{۳}$$

بعبارت دیگر، در حد گراف های بزرگ، الگوریتم تضمین بازیابی حداقل $\frac{1+t}{2}$ از انواع رئوس را ارائه میکند. واضح است که در این تعریف داریم: $0 < t < 1$.

با توجه به این تعریف و با استفاده از نامساوی پردازش اطلاعات [۱۰] نتیجه میگیریم که برای بازیابی جزیی، لازم است که کمیت اطلاعات متقابل بر راس، از مقدار معینی بیشتر باشد

$$i(\mathbf{X}, G) \equiv \frac{I(\mathbf{X}, G)}{N} \geq \frac{I(\mathbf{X}, \hat{\mathbf{X}})}{N} \geq 1 - h\Big(\frac{1+t}{2}\Big) + o(1) \tag{۴}$$

با توجه به این رابطه، علاقمند به یک تخمین از اطلاعات متقابل هستیم به طوریکه خطای آن $o(N)$ باشد.

برای محاسبه ی اطلاعات متقابل، از تعریف زیر استفاده میکنیم.

$$i(\mathbf{X}, G) \equiv \frac{H(G) - H(G|\mathbf{X})}{N} \tag{۵}$$

[۸] البته ضرایب α و β ثابت نیستند و بسته به تعداد رئوس گراف، تغییر میکنند.

[۹] Partial recovery

[۱۰] Data processing inequality

محاسبه ی جمله ی دوم، سر راست است

$$H(G|\mathbf{X}) = \sum_{\mathbf{x}} 2^{-N} \sum_{i<j} \Big[h(\alpha+\beta)\frac{1+x_i x_j}{2} + h(\alpha-\beta)\frac{1-x_i x_j}{2} \Big] \\ = \frac{N(N+1)}{4} \big[h(\alpha+\beta) + h(\alpha-\beta) \big] \tag{۶}$$

با این حال، محاسبه ی جمله ی اول کار ساده ای نیست. داریم

$$H(G) = -\mathbb{E}\big[\log P(G) \big] \tag{۷}$$

$$P(G) = (\alpha+\beta)^{|G|}(1-\alpha-\beta)^{|\bar{G}|}\mathbb{E}_{\mathbf{X}}\Big[\big(\frac{\alpha-\beta}{\alpha+\beta}\big)^{|G^-|} \big(\frac{1-\alpha+\beta}{1-\alpha-\beta}\big)^{|\bar{G}^-|} \Big] \tag{۸}$$

که در آن

$$|G| \equiv \sum_{i<j} G_{ij}; \quad |\bar{G}| \equiv \sum_{i<j} \bar{G}_{ij} \tag{۹}$$

$$|G^-| \equiv \sum_{i<j} G_{ij}\frac{1-X_i X_j}{2}; \quad |\bar{G}^-| \equiv \sum_{i<j} \bar{G}_{ij}\frac{1-X_i X_j}{2} \tag{۱۰}$$

همچنین ماتریس $\bar{G}$ مربوط به گراف مکمل است.

$$\bar{G}_{ij} \equiv (1-\delta_{ij})(1-G_{ij}) \tag{۱۱}$$

تذکر: دقت داریم که همه ی کمیت های آنتروپیک، تحت تبدیل مکملیت متقارن هستند.

$$\alpha \to 1-\alpha; \quad \beta \to -\beta \tag{۱۲}$$

با عنایت به این تقارن مکملیت، ضریب λ را بعنوان معیاری از تمایز بین جماعت ها، چنین تعریف میکنیم

$$\lambda \equiv \log\Big(\frac{\alpha+\beta}{\alpha-\beta}\Big) \tag{۱۳}$$

به طریق مشابه و برای سازگاری نمادگذاری تعریف میکنیم

$$\bar{\lambda} \equiv \log\Big(\frac{1-\alpha-\beta}{1-\alpha+\beta}\Big) \tag{۱۴}$$

با استفاده از این تعاریف میتوانیم بنویسیم

$$P(G) = (\alpha^2-\beta^2)^{|G|/2}\big((1-\alpha)^2-\beta^2\big)^{|\bar{G}|/2}\mathbb{E}_{\mathbf{X}'}\Big\{\exp\Big[\frac{1}{2}\mathbf{X}'^T(\lambda G+\bar{\lambda}\bar{G})\mathbf{X}'\Big]\Big\} \quad (۱۵)$$

که در آن $\mathbf{X}'$ یک بردار از متغیر های مستقل و همتوزیع با توزیع متقارن روی $\{\pm 1\}$ است. در نهایت و با استفاده از (۵) ، (۶) ، (۷) و (۱۵) داریم

$$i(\mathbf{X}, G) = \frac{N+1}{4}\beta(\lambda-\bar{\lambda}) - \frac{1}{N}\mathbb{E}_G \log \mathbb{E}_{\mathbf{X}'}\Big\{\exp\Big[\frac{1}{2}\mathbf{X}'^T(\lambda G+\bar{\lambda}\bar{G})\mathbf{X}'\Big]\Big\} \quad (۱۶)$$

تا اینجا، هنوز رژیم مشخصی از مساله را انتخاب نکرده ایم؛ ضرایب α و β میتوانند هر رفتاری را بر حسب N داشته باشند. همانطور که قبلاً هم اشاره کردیم، بنا داریم تمرکز خود را بر رژیمی قرار دهیم که در آن

$$\alpha = \frac{a}{N}, \quad \beta = \frac{b}{N} \quad (۱۷)$$

که در آن، a و b ثابت هستند. در این رژیم و با صرف نظر کردن از جملات از مرتبه ی $o(1)$ در نهایت به رابطه ی زیر برای اطلاعات متقابل بر راس میرسیم.

$$\boxed{i(\mathbf{X}, G) \overset{o(1)}{=} \frac{\lambda b}{4} - \frac{1}{N}\mathbb{E}_G \log \mathbb{E}_{\mathbf{X}'}\Big\{\exp\Big[\frac{\lambda}{2}\mathbf{X}'^T G\mathbf{X}'\Big]\Big\}} \quad (۱۸)$$

فرع: با توجه به اینکه اطلاعات متقابل بر راس در نامساوی $0 \leq i(\mathbf{X}, G) \leq 1$ صدق میکند، داریم

$$\frac{\lambda b}{4} - 1 \leq \lim_{N\to\infty}\frac{1}{N}\mathbb{E}_G \log \mathbb{E}_{\mathbf{X}'}\Big\{\exp\Big[\frac{\lambda}{2}\mathbf{X}'^T G\mathbf{X}'\Big]\Big\} \leq \frac{\lambda b}{4} \quad (۱۹)$$

۲.۲ حدس گاوسی

برای ارزیابی حاصل معادله ی (۱۸) ابتدا ماتریس مجاورت G را به صورت $G = QDQ^T$ قطری میکنیم. در این صورت، میتوانیم فرضیه ی زیر را فرمولبندی کنیم:

فرضیه ی گاوسی: در حد گراف های بزرگ، تقریباً همه ی ویژه بردار های ماتریس مجاورت با ویژه مقدار غیر صفر، نرم صفر بسیار بزرگتر از یک دارند. بعبارت دیگر، ویژه بردار های ماتریس مجاورت، تُنُک نیستند!

در صورتی که این فرضیه درست باشد، با استفاده از قضیه ی حد مرکزی، میتوانیم توزیع بردار $\mathbf{X}'$ در (۱۸) را با یک توزیع گاوسی استاندارد عوض کنیم. در این صورت به رابطه ی ساده شده ی زیر برای اطلاعات متقابل بر راس میرسیم

$$i(\mathbf{X}, G) \overset{G\ Conj.}{=\!=} \frac{\lambda b}{4} + \frac{1}{2N}\mathbb{E}_G \log\left|\mathbf{I} - \lambda G\right| \tag{۲۰}$$

در نهایت و برای محاسبه ی عمل متوسط گیری، لازم است که به توزیع عملی طیف G دسترسی داشته باشیم.

۳.۲ طیف G

درباره ی طیف ماتریس های تصادفی صحبت های بسیاری شده است. برای مثال، طیف ماتریس های منتظم [۱۱] در یک مقاله ی کلاسیک در سال ۱۹۸۱ میلادی محاسبه شده است. [۷] همچنین در مرجع [۸] به یک الگوریتم شناخته شده برای مساله ی تشخیص جماعت تحت عنوان الگوریتم طیفی اشاره شده است. درباره ی این الگوریتم در ادامه بیشتر صحبت خواهیم کرد. بهرحال، یک نکته که بعنوان یک تعمیم از نتیجه ی موجود در [۸] منتج میشود، این است که طیف ماتریس مجاورت G شامل دو ویژه مقدار تعینی و $N-2$ ویژه مقدار تصادفی است که توزیع توده ای [۱۲] ویژه مقدار ها را تشکیل میدهند. با مقایسه با [۷] نتیجه میگیریم که در حد $a \gg 1$ توزیع توده ای طیف به توزیع نیمدایره ی ویگنر با شعاع $2\sqrt{a}$ نزدیک میشود. بنابراین، تمرکز خود را به توزیع توده ای طیف ماتریس $\frac{G}{\sqrt{a}}$ معطوف میکنیم. گشتاور های این توزیع از رابطه ی زیر بدست میآیند

$$\begin{aligned}\langle x^n \rangle &= \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\Big[a^{-n/2}\operatorname{Tr} G^n\Big] \\ &= \lim_{N\to\infty} \frac{1}{Na^{n/2}} \sum_{i_1\cdots i_n} \mathbb{E}\Big[G_{i_1 i_2} G_{i_2 i_3}\cdots G_{i_n i_1}\Big]\end{aligned} \tag{۲۱}$$

۴.۲ شمارش درخت ها و گشتاور های توزیع طیفی

یک محاسبه ی دقیق نشان میدهد که $\mathbb{E}\Big[G_{i_1 i_2} G_{i_2 i_3}\cdots G_{i_n i_1}\Big]$ متناظر با مجموع روی تمام مسیرهای بسته (و احیاناً خود متقاطع) به طول n است. تعداد زیرگراف های شامل هر مسیر با V

[۱۱] Regular
[۱۲] Bulk distribution

راس و E یال، شامل یک ضریب $\binom{N}{V}$ و همچنین E ضریب از مرتبه ی $1/N$ است. بنابراین، تنها جملاتی که مربوط به یک دور بسته روی یک درخت میشوند، ($V = E + 1$) در حد $N \to \infty$ سهم میدهند. یک نتیجه ی فوری این است که:

گزاره: توزیع توده ای طیف ماتریس مجاورت، مستقل از پارامتر b (یا λ) است و تنها به a بستگی دارد.

برای اثبات کافیست توجه کنیم که در یک درخت، عدم وجود حلقه ها، منجر به استقلال یال ها از یک دیگر میشود. هر یال، با احتمال میانگین a/N وجود دارد و پارامتر b اثری در نتیجه ی نهایی ندارد. همچنین با توجه به آنکه هیچ مسیر بسته با طول فرد روی یک درخت وجود ندارد، نتیجه میگیریم که

گزاره: توزیع توده ای طیف ماتریس مجاورت، زوج است. بعبارت دیگر

$$\langle x^n \rangle_a = 0; \quad \forall \ odd \ n \tag{۲۲}$$

در نهایت میتوان نشان داد که

$$\langle x^{2n} \rangle_a \sum_{p=0}^{n-1} T_{np} a^{-p} \tag{۲۳}$$

$$T_{np} \equiv \sum_{|\Gamma|=n-p} \frac{\mathtt{CL}(\Gamma, 2n)}{\mathtt{S}(\Gamma)} \tag{۲۴}$$

که در آن، $\mathtt{SL}(\Gamma, 2n)$ تعداد مسیرهای به طول $2n$ است که از یک راس دلخواه روی درخت Γ شروع میشوند و و در نهایت به نقطه ی آغاز برمیگردند. (قیدی روی خود متقاطع بودن یا تعداد دفعات استفاده از هر یال وجود ندارد) همچنین ضریب تقارن [۱۳] برای یک گراف را تعداد نامگذاری های متفاوت برای رئوس گراف تعریف میکنیم به طوری که ماتریس مجاورت دست نخورده باقی بماند؛ این همچنین برابر با تعداد ماتریس های جایگشت Π است به طوریکه

$$\Pi \Gamma \Pi^T = \Gamma \tag{۲۵}$$

در نهایت، منظور از $|\Gamma|$ مانند گذشته، تعداد یال های درخت Γ است.

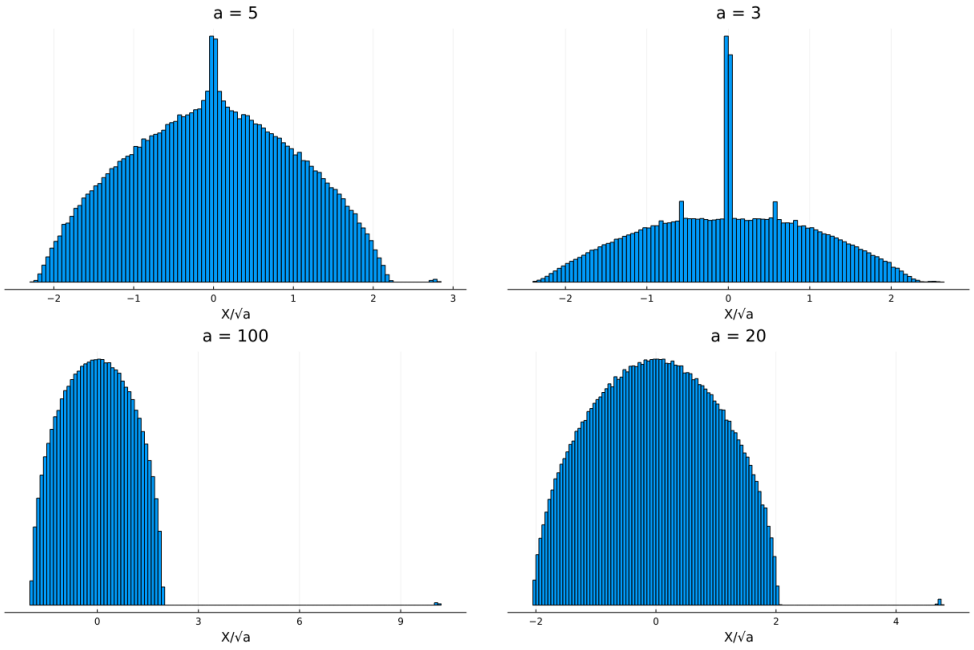
با کمک کامپیوتر و شمارش درخت های مختلف و محاسبه ی مجموع بالا، ضرایب T_{np} مطابق جدول ۱ محاسبه میشوند.

[۱۳] Symmetry factor

	$p=0$	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$	$p=7$	$p=8$
$n=1$	۱								
$n=2$	۲	۱							
$n=3$	۵	۶	۱						
$n=4$	۱۴	۲۸	۱۴	۱					
$n=5$	۴۲	۱۲۰	۱۱۰	۳۰	۱				
$n=6$	۱۳۲	۴۹۵	۶۸۲	۳۷۵	۶۲	۱			
$n=7$	۴۲۹	۲۰۰۲	۳۷۳۱	۳۲۴۸	۱۱۹۰	۱۲۶	۱		
$n=8$	۱۴۳۰	۸۰۰۸	۱۸۹۲۸	۲۳۰۲۰	۱۴۰۶۲	۳۶۲۸	۲۵۴	۱	
$n=9$	۴۸۶۲	۳۱۸۲۴	۹۱۳۹۲	۱۴۴۰۲۴	۱۲۷۰۲۹	۵۷۵۱۶	۱۰۸۰۵	۵۱۰	۱

جدول ۱: چند مقدار اولیه از T_{np}؛ مقادیر مربوط به $p=0$ با قانون نیمدایره ی ویگنر سازگار هستند. همچنین برای $p \geq n-2$ اثبات یک رابطه ی صریح، آسان است.

در ادامه چند هیستوگرام که به کمک شبیه سازی طیف ماتریس $G/\sqrt{a}$ بدست آمده اند را رسم میکنیم. این هیستوگرام ها از روی هم قرار دادن ویژه مقدار های ۵۰ ماتریس مجاورت گراف هایی با $N=2000$ که با پارامتر های $a=3,5,20,100$ و $\lambda=0$ تولید شده اند. برای مقادیر بزرگ a یکی از ویژه مقادیر ماتریس مجاورت از توزیع توده ای خارج میشوند.



شکل ۱. هیستوگرامهای مربوط به توزیع توده ی طیفی ماتریس $G/\sqrt{a}$ به ازای چند مقدار مختلف پارامتر a

۵.۲ شکست حدس گاوسی

با استفاده از مقادیر محاسبه شده برای ضرایب T_{np} میتوان بسادگی نتیجه گرفت که فرضیه ی گاوسی که در قسمت های قبلی به آن اشاره کردیم، نادرست است. کافیست برای مقادیر کوچک λ مقدار اطلاعات متقابل را محاسبه کنیم. معادله ی (۲۰) میدهد

$$\frac{a}{4}\lambda \tanh(\lambda/2) + \frac{1}{2}\Big\langle \log\left(1-\lambda\sqrt{a}x\right)\Big\rangle = -a\frac{\lambda^2}{8} + \mathcal{O}(\lambda^4) \qquad (۲۶)$$

که به وضوح با مثبت بودن کمیت اطلاعات متقابل در تناقض است.

۳ روش طیفی

همانطور که گفتیم، علاوه بر ویژه مقادیر توده ای ماتریس مجاورت، دو ویژه مقدار تعینی هم وجود دارند. در رژیمی که میانگین درجه ی رئوس به صورت خطی با اندازه ی گراف رشد میکنند، نشان داده شده است که ویژه مقدار مربوط به دومین ویژه بردار، همبستگی بالایی با نوع رئوس پیدا میکند. [۸] واضح است که برای تشخیص ویژه بردار مربوط، لازم است که دومین ویژه مقدار بزرگ ماتریس مجاورت، خارج از توده ی طیفی قرار بگیرد. در رژیم مورد نظر ما به نظر میرسد که توزیع توده ی طیفی، به یک بازه ی خاص محدود نمیشود. بنابراین و در حد N های بسیار بزرگ، استفاده از الگوریتم طیفی هرگز موفقیت آمیز نخواهد بود. با این حال و برای N های متوسط، چنان که خواهیم دید، این الگوریتم عملکرد قابل قبولی ارائه میدهد. همچنین، باید توجه داشت که آنچه در ادامه بعنوان گذار فاز معرفی میشود، ازآنجایی که فقط برای N های محدود قابل مشاهده است، به معنای واقعی کلمه نمیتواند گذار فاز نام بگیرد. احتمالاً شایسته است که این پدیده را یک شبه گذار بنامیم.

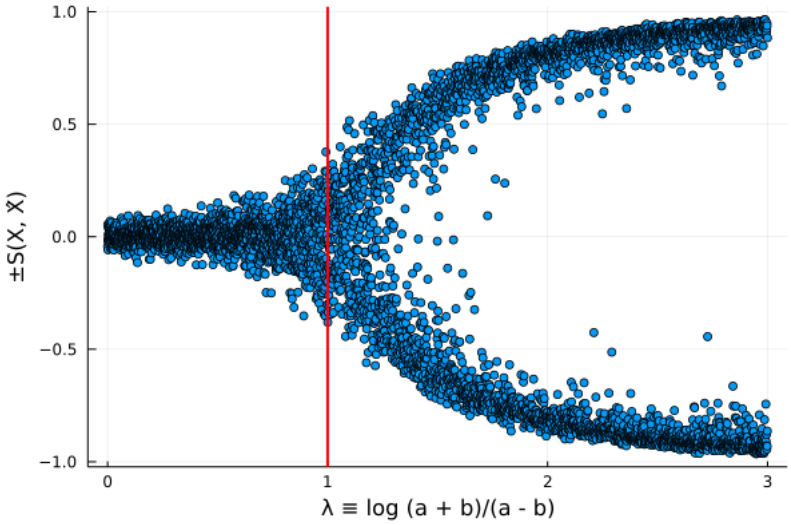
تعریف زیر از سیگنال به نویز متداول است [۴]

$$\sigma \equiv \frac{(a-b)^2}{2(a+b)} = a\frac{e^{2\lambda}}{1+e^{\lambda}} \tag{۲۷}$$

ثابت شده است که برای $\sigma < 1$ بازیابی جزیی امکان پذیر نخواهد بود. با استفاده از این رابطه، مقدار بحرانی λ برای مشاهده ی یک گذار احتمالی در عملکرد الگوریتم طیفی چنین خواهد بود

$$\lambda_c = \log\left(a\frac{1+\sqrt{1+4/a}}{2}\right) \tag{۲۸}$$

شکل زیر، عملکرد الگوریتم طیفی در بازیابی یک مدل بلوکی تصادفی با دو جماعت متقارن را نشان میدهد. مقدار a چنان انتخاب شده است که گذار در $\lambda_c = 1$ رخ دهد. همچنین محور عمودی، نمایانگر مثبت یا منفی کمیت $S \equiv 2A - 1$ است.



شکل۲. شبه گذار در عملکرد الگوریتم طیفی با عبور پارامتر λ از مقدار بحرانی.

۴ خلاصه نتایج

در این گزارش در ابتدا با تمرکز بر روی کمیت اطلاعات متقابل بر راس بعنوان یک کمیت ماکروسکوپی مستقل از الگوریتم، سعی کردیم به توجیه گذارهای فاز شناخته شده در مدل های بلوکی تصادفی با دو جماعت متقارن بپردازیم. نشان دادیم که حدس گاوسی معرفی شده در [۶] نمیتواند در این رژیم درست باشد. همچنین درباره ی توزیع توده ای طیف ماتریس مجاورت در این آنسامبل بحث کردیم و راه حلی برای محاسبه ی گشتاور های این توزیع عملی ارائه دادیم. در پایان با استفاده از شبیه سازی کامپیوتری عملکرد الگوریتم طیفی در مساله ی تشخیص جماعت مورد بررسی قرار گرفت. نشان دادیم که استفاده از این الگوریتم، اگرچه نمیتواند برای گراف های بسیار بزرگ موفقیت آمیز باشد، اما در گراف های با اندازه ی متوسط، عملکرد بهینه دارد و تا قبل از گذار معرفی شده در [۴] میتواند در بازیابی جزیی جماعت ها موفق عمل کند.

با پایان فرصت انجام پروژه ی کارشناسی، همچنان مسائل باز جالبی در همین مسیر وجود دارند که میتوان در ادامه به آن ها پرداخت، برای مثال کمی کردن حیطه ی عملکرد مناسب الگوریتم طیفی (N های متوسط) یا اندازه گیری نماهای بحرانی در گذار های مورد بحث میتوانند بعنوان مسائل جدید پیگیری شوند.

۵ سپاسنامه

در اینجا لازم است از استاد راهنمای خود، دکتر محمدحسین یاسایی که در انجام این پروژه، راهنمایی های گرانبهایی را در اختیارم قرار دادند تشکر کنم. همچنین، از اساتید محترم درس های پروژه ی کارشناسی، آقایان دکتر جاهد و آشتیانی که درس های ارزنده ای درباره ی روش تحقیق، گزارش نویسی و ارائه ی علمی به من آموختند سپاسگزارم. همچنین و در پایان مایلم از دکتر امین زاده بابت آنچه از ایشان آموختم قدردانی کنم.

مراجع

[1] P. Erdos, A Renyi, *On the evolution of random graphs* (1960), THE MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES, 3047

[2] B. Bollobas, *Random Graphs* (2nd edition), Cambridge University Press, 2001

[3] M. S. Ashuja, *et al*, Practical Applications of Community Detection, *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 6, Issue 4.

[4] E. Abbe, *Community detection and stochastic block models: recent developments*, arXiv:1703.10146

[5] A Decelle, *et al.*, *Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications*, arXiv: 1109.3041

[6] Y. Deshpande, *et al*, *Asymptotic Mutual Information for the Balanced Binary Stochastic Block Model*, arXiv: 1507.08685

[7] B. D. McKay, The Expected Eigenvalue Distribution of a Large Regular Graph, *LINEAR ALGEBRA AND ITS APPLICATIONS* 40:203-216 (1981)

[8] R. Vershynin, *High-Dimensional Probability: An Introduction with Application sin Data Science*, math.uci.edu/rvershynin/