دانشکده علوم ریاضی

گروه ریاضی کاربردی

عنوان

جستجوی گراف مخفی

پژوهشگر

الهه شریفی

استاد راهنما

دکتر میثم علیشاهی

استاد مشاور

سید رضا موسوی

پایان‌نامه جهت اخذ درجه کارشناسی ارشد
ریاضی کاربردی، گرایش گراف و ترکیبیات

تیر ۱۳۹۲

نام خانوادگی دانشجو: شریفی نام: الهه

عنوان: جستجوی گراف مخفی

استاد راهنما: دکتر میثم علیشاهی

استاد مشاور: سید رضا موسوی

مقطع تحصیلی: کارشناسی ارشد رشته: ریاضی کاربردی گرایش: گراف و ترکیبیات

دانشگاه: دانشگاه صنعتی شاهرود دانشکده علوم ریاضی

تاریخ فارغ‌التحصیلی: تیر ۱۳۹۲ تعداد صفحات: ۸۳

واژگان کلیدی: آزمون گروهی، الگوریتم غیرانطباقی، توالی‌یابی ژنوم

چکیده

در این پایان نامه، مسأله‌ی جستجوی گرافی مخفی از یک خانواده از گراف‌ها روی n رأس، بررسی می‌شود. در مدلی که از آن استفاده می‌کنیم، تنها عمل مجاز عبارت است از این پرسش که: "آیا یک زیرمجموعه از رئوس، دارای یالی از این گراف مخفی است؟". هدفی که در این راستا مطرح می‌شود، تخمین کم‌ترین تعداد پرسشی است که برای شناسایی گراف مخفی مورد نیاز است. برای رسیدن به این هدف عموماً از الگوریتم‌های غیرانطباقی استفاده کرده و مسأله را برای خانواده‌ی تطابق‌ها، خانواده‌ی ستاره‌ها و خانواده‌ی خوشه‌ها بررسی می‌کنیم. همچنین، کران‌هایی برای اندازه‌ی خانواده‌ای که مسأله را برای گراف‌های کلی‌تر حل می‌کند، ارائه می‌دهیم.

تقدیم به مهربان فرشتگانی که

از نگاهشان صلابت

از رفتارشان محبت

و از صبرشان ایستادگی را آموختم

تقدیم به پدر و مادر عزیزم

سپاس گزاری...

سپاس بی‌کران پروردگار یکتا را که هستی‌مان بخشید و بــه طریـق علـم و دانش رهنمونمان شد و به همنشینی رهروان علم و دانش مفتخرمان نمــود و خوشه چینی از علم و معرفت را روزیمان ساخت.

خدا را شاکرم که به من این توان را داد که این مرحله از زندگی را با سلامتی و سر بلندی پشت سر بگذارم و سر بر آستان پربرکت او می‌سایم که هرچه عزت و سرافرازی هست، از اوست. پس از خدای بلند مرتبه سر تعظیم در برابر پدر و مادرم فرود می‌آورم، بر پایشان بوسه می‌زنم و ایشان را می‌ستایم که تجلی مهر و لطف خداوندی برای من می‌باشند.

لازم می‌دانم از استاد با کمالات و شایسته، جناب آقای دکتر میثم علیشاهی که در کمال سعه‌ی صدر، با حسن خلق و فروتنی، از هیچ کمکی در این عرصه بر من دریغ ننمودند و با راهنمایی‌های ارزنده و گهربارشان نقش مهمی در به ثمر رسیدن این پایان نامه داشتند، صمیمانه تقدیر و تشکر نمایم. همچنین از استاد صبور و با تقوا، جناب آقای سید رضا موسوی که زحمت مشاوره‌ی این پایان نامه را متقبل شدند و از اساتید فرزانه و دلسوز، آقایان دکتر نادر جعفری راد و دکتر صادق رحیمی شعرباف که زحمت داوری این پایان نامه را متقبل شدند، کمال تشکر و قدردانی را نمایم. باشد که این خردترین، بخشی از زحمات آنان را سپاس گوید.

الهه شریفی

تیر ۱۳۹۲

پیشگفتار

فرض کنید n ماده‌ی شیمیایی در اختیار داریم که برخی از این مواد با یکدیگر واکنش می‌دهند. لازم است با انجام آزمایش تعیین شود که کدام مواد با یکدیگر واکنش می‌دهند. از طرفی، بنا به دلایلی از جمله تعداد زیاد مواد، هزینه‌ی بالای بعضی از آزمایش‌ها و فسادپذیری برخی مواد، نمی‌توان این مواد را دوبه‌دو آزمایش کرد. بنابراین، به دنبال روشی هستیم که تعداد آزمایش‌ها را کاهش دهد. آزمون گروهی، یکی از این روش‌هاست. در این راستا، مسأله را به مسأله‌ی آزمون گراف یعنی، جستجوی یک گراف مخفی روی n رأس، تعمیم می‌دهیم. در این تعمیم، مواد شیمیایی را متناظر با رئوس، واکنش میان مواد را متناظر با یال‌ها و یک آزمایش را متناظر با این پرسش در نظر می‌گیریم که: "آیا یک زیرمجموعه از رئوس شامل یال است یا خیر؟". اگر F زیرمجموعه‌ای از رئوس باشد، آنگاه می‌توان F را مورد پرسش قرار داد. اگر پاسخ "منفی" بود یعنی، در F هیچ یالی واقع نمی‌شود و اگر پاسخ "مثبت" بود یعنی، حداقل دو رأس در F مجاورند.

آنچه مطلوب مسأله است، تخمین کم‌ترین تعداد پرسشی است که مسأله را حل می‌کند. برای این منظور و با توجه به ارجحیت الگوریتم‌های غیرانطباقی در کاربرد، غالباً از این الگوریتم‌ها بهره جسته و کران‌هایی را برای کم‌ترین تعداد پرسش‌های مورد نیاز ارائه می‌دهیم.

در فصل اول این پایان نامه به معرفی آزمون گروهی پرداخته و کاربردهای این آزمون را در علوم مختلف زیستی بیان می‌کنیم. سپس مسأله را در غالب آزمون گراف، مدل‌سازی می‌کنیم.

در فصل دوم، با در نظر گرفتن این فرض که گراف مخفی یک تطابق روی n رأس است، جستجو را در خانواده‌ی تطابق‌ها انجام می‌دهیم و کوچک‌ترین اندازه‌ی خانواده‌ای که مسأله‌ی تطابق روی n رأس را حل می‌کند، تخمین می‌زنیم. برای این منظور، مسأله‌ی جستجوی تطابق روی n رأس را به پیدا کردن تطابقی روی گراف H تعمیم می‌دهیم. سپس یک گراف H ارائه می‌دهیم به‌طوری‌که می‌توان مسأله‌ی تطابق را با خانواده‌ای از اندازه‌ی تقریباً نصف تعداد یال‌های H حل کرد. با استفاده از نتایج به‌دست آمده و قضیه‌ی افراز ویلسون، مسأله‌ی تطابق روی n رأس را با خانواده‌ای از اندازه‌ی $(\frac{1}{2} + o(1))\binom{n}{2}$ حل می‌کنیم. در ادامه الگوریتم‌های احتمالاتی و الگوریتم‌های چندمرحله‌ای را برای حل این مسأله ارائه می‌دهیم. مقاله‌ی [۲] مرجع اصلی این فصل است.

در فصل سوم، روش‌های قبل را تعمیم داده و مسأله را برای خانواده‌ای از گراف‌ها با ویژگی معلوم در

نظر می‌گیریم. از جمله خانواده‌هایی که در این فصل مطالعه می‌شوند، خانواده‌ای شامل همه‌ی گراف‌های یکریخت با گراف G است و کران پایینی را که به عدد استقلال G وابسته است، برای اندازه‌ی خانواده‌ای که این مسأله را حل می‌کند، اثبات می‌کنیم. در پایان این فصل، خانواده‌ی گراف تصادفی $G(n, \frac{1}{2})$ را در نظر می‌گیریم. نتایج این فصل براساس مراجع [۱] و [۲] است.

در فصل چهارم، براساس مرجع [۱]، مسأله‌ی جستجوی گراف مخفی را، زمانی که گراف مخفی یک ستاره است، مطالعه و مسأله را برای هر دو حالتی که اندازه‌ی ستاره معلوم یا نامعلوم است حل می‌کنیم.

آنچه که در فصل پنجم مورد بررسی قرار می‌گیرد، خانواده‌ی خوشه‌ها و همچنین خانواده‌ی خوشه‌ها با اندازه‌ی مشخص است و کران‌هایی برای کوچک‌ترین اندازه‌ی خانواده‌ای که این مسأله را حل می‌کند اثبات می‌کنیم. مرجع این فصل نیز [۱] است.

فهرست مطالب

لیست جداول

لیست تصاویر

فصل ۱

مقدمه و تاریخچه

۱.۱ تاریخچه‌ی آزمون گروهی

برخلاف بسیاری از مسائل ریاضی که به قرن‌ها پیش و منابع مختلف برمی‌گردند منشأ آزمون گروهی [۱] تا حد زیادی به حادثه‌ی نسبتاً اخیر جنگ جهانی دوم (۱۹۴۲-۴۳) پیوند خورده است و به روبرت دورفمَن [۲] نسبت داده می‌شود [۶]. میلیون‌ها سرباز در جنگ حضور داشتند که در بین آن‌ها فقط تعداد اندکی، ۱۰۰۰ مورد، مبتلا به ویروس سیفیلیس شده بودند. دورفمن و دِیوید روزن‌بلات [۳] در یکی از شعبه‌های تحقیقاتی اداره‌ی آمار مالی واشنگتن کار می‌کردند، در جلسه‌ای که در این اداره با حضور تعدادی اقتصاددان برگزار شده بود تمام حضار به این نتیجه رسیدند که اگر برای پیدا کردن تعداد اندکی بیمار، نمونه خون میلیون‌ها سرباز را به‌صورت انفرادی مورد تجزیه و تحلیل آزمایشگاهی قرار دهیم عملی بی‌فایده است. آزمون گروهی توسط یکی از افراد در جلسه مطرح شد ولی از طرف دیگر فردی اظهار داشت که ممکن است گروه‌بندی کردن نمونه‌های خون مشکلات اقتصادی داشته باشد و این ایده رد شد. اما دورفمن و همکارش روزن‌بلات (دورفمن در مقالاتش به همکاری روزن‌بلات اشاره می‌کند) این ایده را به‌صورت جدی پیگیری کردند. بعد از مدت کوتاهی دورفمن یک یادداشت چهار صفحه‌ای را در جلسه‌ی انجمن آمار واشنگتن ارائه نمود که در مجله‌ی سالیانه‌ی آمار ریاضی آمریکا منتشر شد.

با توجه به گزارش اصلی [۵] از دورفمن:

[۱] group testing
[۲] Robert Dorfman
[۳] Devid Rosenblatt

آزمایش خون میکروب سیفیلیس برای هر سربازی که مورد آزمایش قرار می‌گیرد، به دو مرحله تقسیم می‌شود:

۱. نمونه خون افراد دریافت می‌شود.

۲. نمونه خون در معرض تجزیه و تحلیل آزمایشگاهی قرار می‌گیرد که وجود یا عدم وجود آنتی‌ژن سیفیلیس را نشان می‌دهد.

وقتی این فرآیند استفاده شود n تجزیه و تحلیل آزمایشگاهی به‌منظور شناسایی افراد مبتلا به ویروس، از میان جمعیتی به اندازه‌ی n نیاز است.

با توجه به روش پیشنهاد شده، میکروب به‌صورت زیر شناسایی می‌شود.

مقداری از نمونه خون دریافت شده از افراد را به گروه‌های پنج تایی تقسیم می‌کنیم. نمونه خون‌های هر گروه را با هم ترکیب کرده و در یک مخزن می‌ریزیم. حال نمونه خون داخل مخزن را در معرض تجزیه و تحلیل آزمایشگاهی قرار می‌دهیم. اگر هیچ یک از پنج نمونه خون داخل مخزن حاوی میکروب سیفیلیس نباشد جواب آزمایش "منفی" است، ولی اگر لااقل یکی از نمونه خون‌های داخل مخزن حاوی این میکروب باشد جواب آزمایش "مثبت" است. آزمایش‌های منفی کنار گذاشته می‌شوند ولی آزمایش‌های مثبت باید به‌صورت انفرادی مجدداً بررسی شوند.

در آن روز این ایده‌ی امیدبخش (گروه‌بندی کردن نمونه‌های خون) به‌صورت کاربردی مطرح نشده بود ولی امروزه ایده‌ی دورفمن تأثیرات جدید و مهمی بر روی جامعه‌ی پزشکی و صنعت گذاشته است.

مسأله‌ی آزمایش خون دورفمن در سال ۱۹۵۰ باعث تدوین کتاب معروف فلر [۴] در زمینه‌ی احتمالات، تحت عنوان "مقدمه‌ای بر نظریه‌ی احتمال و کاربرد آن"، شد [۸].

سوبل [۵] و گرول [۶] [۱۴]، دو دانشمند از آزمایشگاه بل، در مقاله‌های خود به مفهوم و کاربرد جدیدی از آزمون گروهی اشاره کردند، برای مثال می‌توان به کاربرد آزمون گروهی در صنعت اشاره کرد. دورفمن، سوبل و گرول آزمون گروهی را تحت مدل احتمالی بررسی کردند، برای مثال یک توزیع احتمال بر روی تمام

[۴] Feller
[۵] Sobel
[۶] Groll

نواقص تعریف کردند و به حداقل رساندن تعداد آزمایش‌های لازم برای شناسایی نواقص را به‌عنوان هدف قرار دادند.

آزمون گروهی ترکیباتی ابتدا توسط لی [۷] مورد مطالعه قرار گرفت [۱۳].

۲.۱ آزمون گروهی ترکیباتی

جامعه‌ای از افراد بیمار و سالم را درنظر می‌گیریم به‌طوری‌که تعداد افراد بیمار نسبت به تعداد کل افراد جامعه عدد بسیار کوچکی است. دقت شود که با توجه به داده‌های تجربی در یک جامعه، می‌توان نسبت افراد بیمار به کل جامعه را عددی ثابت فرض کرد.

هدف، شناسایی افراد بیمار است و برای این کار آزمایش خون از تمامی افراد جامعه گرفته می‌شود. حال اگر قرار باشد نمونه خون‌های گرفته شده از افراد گروه به‌صورت انفرادی مورد آزمایش قرار بگیرند ممکن است به دلایل ذیل مشکلاتی را به دنبال داشته باشد:

الف. برخی از آزمایش‌ها وجود دارند که انجام آن‌ها هزینه‌ی بالایی را دربردارد.

ب. تجزیه و تحلیل بعضی از آزمایش‌ها به زمان زیادی نیاز دارد و همچنین کمبود دستگاه‌های آزمایشگاهی، آزمون انفرادی در زمان کوتاه را غیرممکن می‌سازد.

بنابراین، یک ابزار مهم برای شناسایی بیماران آزمایش گروهی است.

فرض کنید یک جامعه‌ی n-نفره داریم و می‌خواهیم افراد بیمار را در بین آن‌ها شناسایی کنیم. کل افراد این جامعه را با N نمایش می‌دهیم. هر زیرمجموعه‌ی ناتهی از N مانند F را یک "**گروه**" می‌نامیم. نتیجه‌ی آزمایش روی یک گروه "منفی" است اگر همه‌ی اعضای گروه سالم باشند و "مثبت" است اگر لااقل یک فرد بیمار در F قرار داشته باشد.

یک **آزمون گروهی** شامل t گروه $F_1, F_2, \ldots, F_t$ است به‌طوری‌که نتیجه‌ی آزمایش این گروه‌ها همه‌ی افراد بیمار را مشخص می‌کند.

[۷] Li

در آزمون گروهی هدف این است که با فرض وجود حداکثر d بیمار در میان جمعیتی از اندازه‌ی مشخص n، با حداقل آزمایش ممکن بتوانیم افراد بیمار را شناسایی کنیم. در نظر گرفتن دو نکته‌ی زیر در کاهش تعداد آزمایش‌ها اهمیت دارد:

الف. از آنجایی که تعداد بیماران به‌طور نسبی خیلی کوچک‌تر از تعداد کل افراد است، اگر اندازه‌ی گروه‌ها بزرگ باشد آنگاه برای به‌دست آوردن نتیجه‌ی مثبت نیاز به یک آزمون خیلی حساس داریم، بنابراین، اندازه‌ی گروه‌ها در عمل دارای محدودیت است.

ب. با توجه به نوع آزمایش برای هر شخص، تعداد گروه‌های شامل این شخص دارای محدودیت است. برای مثال مقدار نمونه خون دریافتی از هر شخص محدود است و لذا تعداد گروه‌هایی که می‌توانند از این نمونه خون استفاده کنند نمی‌تواند زیاد باشد.

به‌طور کلی الگوریتم‌های آزمون گروهی به دو دسته تقسیم می‌شوند:

۱. الگوریتم‌های ترتیبی. [۸]

۲. الگوریتم‌های غیرانطباقی. [۹]

یک الگوریتم آزمون گروهی را **الگوریتم ترتیبی** می‌نامیم هرگاه آزمایش روی گروه‌ها به‌ترتیب (یکی یکی) انجام شود و در تعیین یک گروه برای آزمایش، می‌توان از نتایج آزمایش‌های قبلی استفاده کرد. یک الگوریتم آزمون گروهی را **الگوریتم غیرانطباقی** می‌نامیم اگر تمام گروه‌ها به‌طور هم‌زمان مورد آزمایش قرار بگیرند. بنابراین، از اطلاعات آزمایش‌های قبلی در تعیین گروه‌های بعدی استفاده نمی‌شود.

از آنجا که داشتن اطلاعات اضافی کمک می‌کند تا گروه‌بندی بهتری داشته باشیم، یک الگوریتم ترتیبی به تعداد گروه‌های کم‌تری برای آزمایش نیاز دارد و چون الگوریتم‌های غیرانطباقی به‌طور هم‌زمان انجام می‌شوند بنابراین، در زمان صرفه‌جویی می‌شود. همان‌طور که گفته شد هدف اصلی آزمون گروهی کم کردن تعداد آزمایش‌های مورد نیاز برای شناسایی همه‌ی افراد بیمار است، بنابراین الگوریتم‌های ترتیبی از نظر

[۸] Sequential algorithms
[۹] Nonadaptive algorithms

تئوری غالب هستند. اما در عمل، به‌خصوص در مولکول شناسی، آزمایش متناظر با یک گروه ممکن است چندین ساعت یا حتی چندین روز به طول انجامد لذا انجام آزمایش‌ها به روش ترتیبی غیرعملی است.

بین الگوریتم‌های ترتیبی و غیرانطباقی، **الگوریتم چندمرحله‌ای** [۱۰] وجود دارد که در آن آزمایش‌ها به چندین مرحله تقسیم می‌شوند به‌طوری‌که تمامی گروه‌های موجود در هر مرحله به‌طور هم‌زمان آزمایش می‌شوند ولی مراحل الگوریتم به ترتیب انجام می‌شوند. یعنی، در تعیین یک گروه می‌توان از نتایج آزمایش‌هایی که در مراحل قبلی انجام شده است استفاده کرد اما نه از آزمایش‌هایی که در همان مرحله موجودند. اگر تعداد مراحل الگوریتم عدد ثابت s باشد، می‌توان آن را یک الگوریتم s-مرحله‌ای [۱۱] نامید.

۳.۱ آزمون گروهی و کاربرد آن در زیست‌شناسی مولکولی

آزمون گروهی ابزاری است مهم که برای مسائل مختلفی مانند آزمایش خون، کانال‌های ارتباطی، تئوری کدگذاری و سایر مسائل به‌کار می‌رود.

اخیراً آزمون گروهی در نظریه‌ی زیست‌شناسی مولکولی [۱۲] مفید واقع شده است تا جایی که امروزه از آزمایش‌های غربال‌گری [۱۳] در زیست‌شناسی مولکولی به عنوان مهم‌ترین کاربرد آزمون گروهی یاد می‌شود. زیست‌شناسی مولکولی، مطالعه‌ی زیست‌شناسی در سطح مولکولی است. این حوزه دارای وجوه مشترکی با زیست‌شناسی، شیمی و به‌طور خاص، با علم ژنتیک و بیوشیمی است. بحث عمده در زیست‌شناسی مولکولی استنباط برهم‌کنش بین سیستم‌های درون سلولی، از جمله، برهم‌کنش‌های RNA، DNA و پروتئین‌سازی است. همچنین، چگونگی تنظیم این برهم‌کنش‌ها مورد بررسی قرار می‌گیرد.

در این پایان نامه، به بررسی یک مسأله‌ی مهم در زیست‌شناسی محاسباتی [۱۴] می‌پردازیم که از توالی‌یابی ژنوم [۱۵]، به روش تعیین توالی شکاری [۱۶]، ناشی می‌شود. تعیین توالی شکاری تکنیکی با عملکرد بالا است که در تعیین توالی تعداد زیادی از ژنوم‌های میکروبی (حدود ۷۰ نوع) و همچنین دروسفیلا (مگس

[۱۰] multi stage algorithm
[۱۱] s-stage algorithm
[۱۲] molecular biology
[۱۳] screening experiments
[۱۴] computational biology
[۱۵] genome
[۱۶] shotgun sequencing

سرکه)، موش و حتی ژنوم انسان استفاده شده است. در همه‌ی این پروژه‌ها، مجموعه‌ای از کانتیگ‌ها [۱۷] (رشته‌های بلند DNA) در اختیار داریم که به دلایل مختلف محاسباتی یا بیولوژیکی نمی‌توانند حتی با بهترین الگوریتم‌های تعیین توالی، توالی‌یابی شوند. کانتیگ‌ها باید به‌صورت منظم درآمده و فاصله‌های [۱۸] بین آن‌ها با به کار بردن روش‌های آهسته و ملالت‌آور توالی‌یابی شود. وقتی تعداد فاصله‌ها کم است (یعنی کم‌تر از ده مورد) زیست‌شناسان اغلب از روش ترکیبی PCR استفاده می‌کنند. این روش مجموعه‌ای از گشت‌های مولکولی دو جهتی [۱۹] را در طول فاصله‌ها در توالی‌یابی ایجاد می‌کند. این گشت‌ها توسط دستگاه PCR آسان می‌شوند.

به منظور شروع گشت‌های مولکولی، زیست‌شناسان از پرایمر [۲۰] استفاده می‌کنند. پرایمرها رشته‌های کوتاه و پیش‌ساخته‌ای از نوکلئوتید هستند و طوری طراحی می‌شوند که آن‌ها به قالب‌های منحصر به فردی که در انتهای هر کانتیگ قرار دارد می‌چسبند. در این روش ابتدا نمونه‌ای از DNA دو رشته‌ای را که می‌خواهند آن را تکثیر کنند حرارت می‌دهند. حرارت باعث جدا شدن دو رشته‌ی DNA از هم می‌شود. سپس این مخلوط را سرد کرده و پرایمر را به آن اضافه می‌کنند. پرایمر (در دما و غلظت مناسب) به نقاطی از DNA که همانندسازی از آنجا آغاز می‌شود متصل می‌شود و یک گشت یک جهتی در طول فاصله‌ی ایجاد شده در رشته‌ی DNA آغاز می‌شود. نتیجه‌ی این فرآیند تولید دو مولکول DNA است که با یکدیگر و با مولکول اولیه مشابه هستند. فرآیند گرمادهی و همانندسازی بارها و بارها تکرار می‌شود و هر چند دقیقه یک‌بار نمونه‌ی DNA دو برابر می‌شود و در زمان کوتاهی میلیون‌ها نسخه از نمونه‌ی اولیه به‌دست می‌آید. یک واکنش PCR [۲۱] اتفاق می‌افتد و یک بسط DNA [۲۲] را می‌توان دید هرگاه دو پرایمری که در دو طرف یک فاصله‌ی یکسان متصل شده‌اند در یک لوله‌ی آزمایش یکسان قرار گیرند [۷، ۲].

اگر N کانتیگ در اختیار داشته باشیم، روش ترکیبی PCR تمامی جفت‌های ممکن از $2N$ پرایمر را با قرار دادن هر دو پرایمر و رشته‌ی DNA اصلی جدا نشده در یک لوله‌ی آزمایش، آزمایش می‌کند.

[۱۷] contig
[۱۸] gaps
[۱۹] bi-directional molecular walks
[۲۰] primer
[۲۱] PCR reaction
[۲۲] DNA ladder

می‌توان فرآورده‌های PCR را با استفاده از ژل‌ها شناسایی و با استفاده از تکنولوژی توالی‌یابی یا طیف‌سنج جرمی DNA، ترجمه کرد. وقتی تعداد فاصله‌ها زیاد است، تعداد آزمایش‌های PCR نیز زیاد است، بنابراین، پرایمرها به‌صورت $k > ۲$ پرایمر در هر لوله مخلوط می‌شوند. این روش PCR چندگانه [۲۳] نامیده می‌شود. در ادامه چند استراتژی مطلوب برای آمیختن پرایمرها، به‌منظور به حداقل رساندن تعداد آزمایش‌های بیولوژیکی مورد نیاز در توالی‌یابی فاصله‌ها، بررسی می‌شود.

۴.۱ مدل‌سازی مسأله

مسأله‌ی توالی‌یابی فاصله می‌تواند به‌طور کلی به‌صورت زیر بیان شود. یک مجموعه از مواد شیمیایی در اختیار داریم، می‌دانیم که هر ماده‌ی شیمیایی با حداکثر یکی از مواد دیگر واکنش می‌دهد (چون تنها پرایمرهایی که در دو طرف یک فاصله‌ی یکسان قرار دارند با هم واکنش می‌دهند)، و یک مکانیزم آزمایشگاهی که تعیین می‌کند وقتی چندین ماده در یک لوله‌ی آزمایش قرار داده می‌شوند آیا واکنشی اتفاق می‌افتد یا خیر. قصد داریم با کم‌ترین تعداد آزمایش، تعیین کنیم کدام جفت از مواد شیمیایی با یکدیگر واکنش می‌دهند.

این مسأله می‌تواند به‌صورت مسأله‌ی شناسایی یا جستجوی یک تطابق مخفی روی مجموعه‌ای از رئوس، با عملی به‌صورت یک پرسش مجاز مدل‌سازی شود. در این مدل یک رأس، یک ماده‌ی شیمیایی و یک یال تطابق، یک واکنش را نمایش می‌دهد و یک پرسش متناظر است با بررسی یک واکنش وقتی مجموعه‌ای از مواد شیمیایی در یک لوله‌ی آزمایش قرار داده می‌شوند. فرض کنید $V = \{۱, ۲, \ldots, n\}$. می‌خواهیم تطابق مخفی (نه لزوماً کامل) M را روی V توسط تعداد کمی پرسش به‌شکل زیر شناسایی کنیم.

$$Q_F: \quad \text{آیا } F \text{ شامل حداقل یک یال از } M \text{ است؟} \qquad (۱.۱)$$

که F زیرمجموعه‌ای از V است و آن را ”گروه“ می‌نامیم. این مسأله حتی در موارد قطعی و غیرانطباقی مورد توجه است. می‌گوییم خانواده‌ی $\mathcal{F}$ از زیرمجموعه‌های V، مسأله‌ی تطابق روی V را حل می‌کند اگر برای هر دو تطابق متمایز $M_۱$ و $M_۲$ روی V حداقل یک $F \in \mathcal{F}$ موجود باشد که شامل یالی از یکی از تطابق‌ها باشد و هیچ یالی از دیگری نداشته باشد. به‌وضوح، هر چنین خانواده‌ای ما را قادر می‌سازد تا به‌طور قطعی و

[۲۳] multiplex PCR

غیرانطباقی، تطابق مخفی را با پرسش‌های Q_F تعریف شده در (۱.۱) برای هر $F \in \mathcal{F}$ بیابیم. هدفی که در این راستا مطرح می‌شود عبارت است از تخمین کوچک‌ترین اندازه‌ی ممکن خانواده‌ای که مسأله‌ی تطابق را روی یک مجموعه‌ی n رأسی حل می‌کند [۲].

۱.۴.۱ تعمیم مسأله

با توجه به کاربرد این مدل در شاخه‌های دیگری مانند علوم زیستی و شیمی، می‌توان آن را تعمیم داد. مسأله‌ی بیولوژیکی به این شرح است که مجموعه‌ای از مولکول‌ها داده شده است که برخی از جفت مولکول‌ها با یکدیگر واکنش می‌دهند. مجدداً رأس‌ها را متناظر با مولکول‌ها و یال‌ها را متناظر با واکنش‌ها و پرسش‌ها را متناظر با آزمایش‌هایی در نظر می‌گیریم که مجموعه‌ای از مولکول‌ها را با هم در یک لوله‌ی آزمایش قرار داده و تعیین می‌کنند که آیا واکنشی اتفاق می‌افتد یا خیر. مانند قبل مسأله را به‌صورت زیر مدل‌سازی می‌کنیم. فرض کنید $\mathcal{H}$ خانواده‌ای از گراف‌های برچسب‌دار روی مجموعه‌ی $V = \{۱, ۲, \dots, n\}$ باشد و فرض کنید $\mathcal{H}$ نسبت به یکریختی بسته است. یک کپی مخفی از $H \in \mathcal{H}$ داده شده است و باید با پرسش‌هایی به‌صورت زیر این کپی را شناسایی کنیم. برای گروه $F \subseteq V$،

$$Q_F: \quad \text{آیا } F \text{ شامل حداقل یک یال از } H \text{ است؟} \qquad (۲.۱)$$

هدف، شناسایی H با کم‌ترین تعداد پرسش ممکن است. می‌گوییم خانواده‌ی $\mathcal{F}$، $\mathcal{H}$-مسأله را حل می‌کند اگر برای هر دو عضو متمایز $H_۱$ و $H_۲$ از $\mathcal{H}$، حداقل یک $F \in \mathcal{F}$ موجود باشد که شامل یالی از یکی از گراف‌های H_i باشد و هیچ یالی از دیگری نداشته باشد. به‌وضوح، هر چنین خانواده‌ای ما را قادر می‌سازد تا به‌طور قطعی و غیرانطباقی، با پرسش‌های Q_F برای هر $F \in \mathcal{F}$، عضو مخفی $\mathcal{H}$ را شناسایی کنیم. دقت شود که برای هر خانواده‌ی $\mathcal{H}$، مجموعه‌ی همه‌ی دوتایی‌ها از رئوس، $\mathcal{H}$-مسأله را حل می‌کند. همچنین، کران پایین اطلاعات دلالت می‌کند بر این که حداقل $\log|\mathcal{H}|$ پرسش نیاز داریم، که در اینجا و در سراسر پایان نامه، تمامی لگاریتم‌ها در مبنای ۲ هستند مگر اینکه غیر این ذکر شود. همچنین، از همه‌ی علامت‌های سقف و کف، در جایی که این علامت‌ها ضروری نیستند، صرف نظر می‌کنیم [۱].

مسأله‌ی بالا برای برخی از خانواده‌ها مطالعه شده است، مانند خانواده‌ی تطابق‌ها [۲]، دورهای همیلتونی

[۹، ۱۰]، خانواده‌ی همه‌ی ستاره‌ها با اندازه‌ی مشخص و خانواده‌ی همه‌ی خوشه‌ها با اندازه‌ی مشخص [۱] و همچنین خانواده‌هایی شامل گراف‌های کلی‌تر. در همه‌ی این مقالات، کم‌ترین تعداد پرسش‌هایی که برای شناسایی گرافی مخفی از خانواده‌ها نیاز است، بررسی شده است اما در بعضی از آن‌ها، پرسش‌های متفاوتی با آنچه مطرح شد در نظر گرفته‌اند، به‌عنوان مثال، گربینسکی [۲۴] و کاچرو [۲۵] در [۹] و [۱۰]، خانواده‌ی دورهای همیلتونی را در نظر گرفتند و چند مدل پرسش در این مقاله‌ها مطرح کردند. علاوه بر مدل (۲.۱)، آن‌ها مدلی را بررسی کردند که در آن پاسخ هر پرسش تنها "مثبت" یا "منفی" نیست بلکه تعداد یال‌های موجود در گروه را نیز مشخص می‌کند. هر دو مدل حتی برای حالتی که اندازه‌ی گروه‌ها محدود است، در نظر گرفته شد. آن‌ها یک الگوریتم انطباقی [۲۶] با $O(n\log n)$ پرسش برای دورهای همیلتونی به‌دست آوردند. همچنین، جواب کاملاً غیرانطباقی $O(n)$ را با استفاده از پرسش‌های قوی‌تر (یعنی، پرسش‌هایی که تعداد یال‌های تولید شده توسط یک مجموعه از رئوس را گزارش می‌کنند) ارائه دادند. سپس کران‌های بالا و پایینی برای مسأله‌ی تطابق، تحت هر یک از این مدل‌ها ارائه دادند که بعضی از کران‌های بالا توسط یک الگوریتم ۲-مرحله‌ای به‌دست می‌آید و دیگر الگوریتم‌ها کاملاً انطباقی هستند. در [۱۱]، گربینسکی و کاچرو مسأله را برای گراف‌هایی با درجه‌ی پایین مطالعه کردند و کران‌های بالا و پایین مسأله‌ی تطابق را تحت مدلی غیرانطباقی اثبات کردند. فرتنو [۲۷] و اپیدین [۲۸] در [۴]، الگوریتمی انطباقی و تصادفی ارائه دادند که مسأله‌ی تطابق را در ۸ مرحله، با به‌طور میانگین $۰/۷۲n\log n$ پرسش، حل می‌کند. در [۲]، آلن و همکارانش [۲۹] با الگوریتمی تصادفی، تعداد مراحل را به ۱ مرحله کاهش دادند (به بهای دو برابر شدن تعداد پرسش‌ها، اما با ۲ مرحله می‌توان تعداد کلی پرسش‌ها را به‌طور تقریبی $۰/۷۲n\log n$ حفظ کرد). همچنین کران بالای $(\frac{۱}{۲}+o(۱))\binom{n}{۲}$ و کران پایین $۰/۳۲\binom{n}{۲}$ را به‌طور قطعی و غیرانطباقی، برای مسأله‌ی تطابق ارائه دادند. آلن و ازُدی [۳۰] در [۱]، این مسأله را برای دو خانواده‌ی ستاره‌ها یا خوشه‌ها و خانواده‌ی ستاره‌ها

[۲۴] Grebinski
[۲۵] Kucherov
[۲۶] adaptive algorithm
[۲۷] Fortnow
[۲۸] Apaydin
[۲۹] Alon et. al.
[۳۰] Asodi

یا خوشه‌ها از اندازه‌ی مشخص، مطرح کردند. آن‌ها در این مقاله تنها از الگوریتم‌های قطعی و غیرانطباقی استفاده کردند.

فصل ۲
جستجوی تطابق

۱.۲ مقدمه

در این فصل خانواده‌ی تطابق‌ها روی n رأس را در نظر گرفته و کوچک‌ترین اندازه‌ی ممکن خانواده‌ای که مسأله‌ی تطابق را حل می‌کند، تخمین می‌زنیم. در راستای این هدف، آلن و همکارانش در [۲]، مسأله‌ی جستجوی تطابق روی n رأس را به پیدا کردن یک تطابق (نه لزوماً کامل) روی گراف H تعمیم داده و یک گراف H ارائه دادند به‌طوری‌که می‌توان مسأله‌ی تطابق را با خانواده‌ای از خوشه‌ها، از اندازه‌ی تقریباً نصف تعداد یال‌های H حل کرد. با توجه به این تعمیم و با استفاده از قضیه‌ی افراز ویلسون [۱]، نشان دادند که می‌توان مسأله‌ی تطابق روی n رأس را با خانواده‌ای از اندازه‌ی $(\frac{1}{2}+o(1))\binom{n}{2}$ حل کرد.

آن‌ها در ادامه، الگوریتم‌های غیرانطباقی و تصادفی را در نظر گرفتند و در مقایسه با الگوریتم قطعی ۱-مرحله‌ای، جواب جالب $O(n\log n)$ را برای این مسأله ارائه کردند. در محاسبات این الگوریتم‌ها از تخمین استاندارد دوجمله‌ای استفاده شده است.

گزاره ۱.۱.۲. [۳]**(تخمین استاندارد دو جمله‌ای)** *فرض کنید X یک متغیر تصادفی دارای توزیع $B(n,p)-np$ است که $B(n,p)$ توزیع دوجمله‌ای است و $a>0$، در این صورت*

$$Pr(X<-a)<e^{-\frac{a^2}{2pn}}.$$

سرانجام الگوریتم‌های قطعی k-مرحله‌ای را با $O(kn^{1+\frac{1}{2(k-1)}}\operatorname{poly}\log n)$ پرسش ارائه دادند، که

[۱] Wilson

به‌طور کلی، منظور از $\mathrm{poly} f(n)$ چند جمله‌ای برحسب $f(n)$ است.

الگوریتم قطعی ۲-مرحله‌ای، $\frac{5}{4}n^{\frac{3}{2}}(1+o(1))$ گروه می‌سازد که هر گروه از اندازه‌ی حداکثر $n^{\frac{1}{4}}$ است. این تعداد در میان همه‌ی الگوریتم‌ها، با گروه‌هایی از اندازه‌ی حداکثر $n^{\frac{1}{4}}$، با عامل $\frac{5}{4}$ بهینه می‌شود. بنابراین، می‌تواند به‌دلیل محدودیت‌های کاربردی در گروه‌بندی، مفید واقع شود. برای $k \geq 3$، این الگوریتم‌ها بر پایه‌ی یک لم رنگ‌آمیزی برای صفحه تصویری استوار هستند که می‌تواند در نوع خود جالب باشد. این روش‌ها ابزارهای ترکیبیاتی و احتمالاتی را با نتایجی در مورد تجزیه‌ی گراف و وجود طرح‌های خاص، تلفیق می‌کنند.

۲.۲ خانواده‌های تُنُک

تعریف ۱.۲.۲. [۲] خانواده‌ی $\mathcal{A} = \{A_1, \ldots, A_k\}$ از مجموعه‌ها، **تنک** [۲] نامیده می‌شود هرگاه گردایه‌ای از دوتایی‌های دوبه‌دو مجزا از اعضای $V = \cup_{i=1}^{k} A_i$ موجود باشد به‌طوری‌که هر A_i شامل حداقل یکی از دوتایی‌ها باشد. به عبارت دیگر، $\mathcal{A}$ تنک است اگر و تنها اگر تطابقی روی V موجود باشد به‌طوری‌که برای هر $A \in \mathcal{A}$، پاسخ پرسش Q_A، "مثبت" باشد.

در یک صفحه تصویری از مرتبه p، هر مجموعه با بیشتر از $\frac{(p^2+p+1)}{2}$ خط تنک نیست. چون، در غیر این صورت، تعداد نقاط این صفحه از $p^2 + p + 1$ نقطه بیشتر خواهد بود. همچنین نتایج به‌دست آمده از این بخش بیان می‌کند که هر خانواده شامل حداکثر $0/32\binom{m+2}{2}$ مجموعه، هریک از اندازه‌ی حداقل m، تنک است.

تعریف ۲.۲.۲. [۲] برای خانواده‌ی $\mathcal{F}$ از زیرمجموعه‌ها، **t-وزن** [۳] خانواده‌ی $\mathcal{F}$ را به‌صورت زیر تعریف کرده و آن را با $w_t(\mathcal{F})$ نشان می‌دهیم

$$w_t(\mathcal{F}) = \sum_{F \in \mathcal{F}} \frac{1}{\binom{|F|+t}{2}}.$$

برای سهولت، ۲-وزن یک خانواده را **وزن** نامیده و آن را با $w(\mathcal{F})$ نشان می‌دهیم.

[۲] sparse
[۳] t-weight

اکنون لم اصلی این بخش را بیان می‌کنیم.

لم ۳.۲.۲. [۲] *هر خانواده‌ی $\mathcal{F}$ از مجموعه‌ها با وزن حداکثر $\frac{۴۹}{۱۵۳}$، تنک است.*

برهان. اگر $\mathcal{F}$ شامل مجموعه‌ای از اندازه‌ی ۱ باشد، آنگاه

$$w(\mathcal{F}) \geq \frac{۱}{۳} > \frac{۴۹}{۱۵۳}.$$

بنابراین، می‌توان فرض کرد همه‌ی مجموعه‌ها در $\mathcal{F}$، از اندازه‌ی حداقل ۲ هستند. فرض کنید $M \in \mathcal{F}$، مجموعه‌ای با کوچک‌ترین اندازه باشد و $|M| = m$.

برای دو عضو متمایز p و q، خانواده‌ی $\mathcal{F}(p,q)$ را به‌صورت زیر تعریف می‌کنیم:

$$\mathcal{F}(p,q) = \{F - \{p,q\} \mid F \in \mathcal{F}, \{p,q\} \nsubseteq F\}.$$

دقت کنید که اگر دوتایی $\{p,q\}$ را به‌عنوان یالی از تطابق (که سعی داریم با ساختن آن نشان دهیم $\mathcal{F}$ تنک است) انتخاب کنیم، آنگاه $\mathcal{F}(p,q)$، آن مجموعه‌هایی هستند که با باقی‌مانده‌ی تطابق، تنک می‌شوند. با توجه به این موضوع ادعای زیر مطرح می‌شود:

ادعا: دو عضو متمایز p و q در M موجود است به‌طوری‌که $w(\mathcal{F}(p,q)) \leq w(\mathcal{F})$. برای اثبات این ادعا دو عضو p و q را به‌صورت تصادفی و یکنواخت از M انتخاب می‌کنیم، کافی است نشان دهیم

$$E(w(\mathcal{F}(p,q))) \leq w(\mathcal{F}).$$

در این صورت دو عضو متمایز p و q در M موجود است به‌طوری‌که

$$w(\mathcal{F}(p,q)) \leq w(\mathcal{F}).$$

در اینجا اعضای $\mathcal{F} \setminus \{M\}$ را با F و $|F \cap M|$ را با $\kappa(F)$ نمایش می‌دهیم. بنابراین، داریم:

$$\begin{aligned}
E(w(\mathcal{F}(p,q))) &= E\Big(\sum_{F_\circ \in \mathcal{F}(p,q)} \frac{۱}{\binom{|F_\circ|+۲}{۲}}\Big) \\
&= \sum_{F_\circ \in \mathcal{F}(p,q)} E\Big(\frac{۱}{\binom{|F_\circ|+۲}{۲}}\Big) \\
&= \sum_{F}\Bigg(\frac{\binom{m-\kappa(F)}{۲}}{\binom{m}{۲}} \cdot \frac{۱}{\binom{|F|+۲}{۲}} + \frac{\binom{\kappa(F)}{۱}\binom{m-\kappa(F)}{۱}}{\binom{m}{۲}} \cdot \frac{۱}{\binom{|F|+۱}{۲}}\Bigg)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_F \left(\frac{\binom{m}{۲} - \binom{\kappa(F)}{۱}\binom{m-\kappa(F)}{۱} - \binom{\kappa(F)}{۲}}{\binom{m}{۲}} \cdot \frac{۱}{\binom{|F|+۲}{۲}} + \frac{\kappa(F)(m-\kappa(F)}{\binom{m}{۲}} \cdot \frac{۱}{\binom{|F|+۱}{۲}} \right) \\
&= w(\mathcal{F}) - \frac{۱}{\binom{m+۲}{۲}} - \sum_F \left(\frac{\kappa(F)(m-\kappa(F))}{\binom{m}{۲}} \cdot \frac{۱}{\binom{|F|+۲}{۲}} + \frac{\binom{\kappa(F)}{۲}}{\binom{m}{۲}} \cdot \frac{۱}{\binom{|F|+۲}{۲}} \right) \\
&+ \sum_F \left(\frac{\kappa(F)(m-\kappa(F))}{\binom{m}{۲}} \cdot \frac{۱}{\binom{|F|+۱}{۲}} \right) \\
&= w(\mathcal{F}) - \frac{۱}{\binom{m+۲}{۲}} + \sum_F \left(\frac{\kappa(F)(m-\kappa(F)}{\binom{m}{۲}} \left(\frac{۱}{\binom{|F|+۱}{۲}} - \frac{۱}{\binom{|F|+۲}{۲}} \right) - \frac{\binom{\kappa(F)}{۲}}{\binom{m}{۲}} \frac{۱}{\binom{|F|+۲}{۲}} \right) \\
&= w(\mathcal{F}) - \frac{۱}{\binom{m+۲}{۲}} + \sum_{k<m} \sum_{\kappa(F)=k} \left(\frac{k(m-k)}{\binom{m}{۲}} \left(\frac{۱}{\binom{|F|+۱}{۲}} - \frac{۱}{\binom{|F|+۲}{۲}} \right) - \frac{\binom{k}{۲}}{\binom{m}{۲}} \frac{۱}{\binom{|F|+۲}{۲}} \right) \\
&= w(\mathcal{F}) - \frac{۱}{\binom{m+۲}{۲}} + \frac{۱}{\binom{m}{۲}} \sum_{k<m} \sum_{\kappa(F)=k} \left(k(m-k) \left(\frac{\binom{|F|+۲}{۲}}{\binom{|F|+۱}{۲}} - ۱ \right) - \binom{k}{۲} \right) \frac{۱}{\binom{|F|+۲}{۲}} \\
&= w(\mathcal{F}) - \frac{۱}{\binom{m+۲}{۲}} + \frac{۱}{\binom{m}{۲}} \sum_{k<m} \sum_{\kappa(F)=k} \left(\frac{۲k(m-k)}{|F|} - \binom{k}{۲} \right) \frac{۱}{\binom{|F|+۲}{۲}} \\
&\le w(\mathcal{F}) - \frac{۱}{\binom{m+۲}{۲}} + \frac{۱}{\binom{m}{۲}} \sum_{k<m} \sum_{\kappa(F)=k} \left(\frac{۲k(m-k)}{m} - \binom{k}{۲} \right) \frac{۱}{\binom{|F|+۲}{۲}} \\
&= w(\mathcal{F}) - \frac{۱}{\binom{m+۲}{۲}} + \frac{۱}{\binom{m}{۲}} \sum_{k<m} \left(\frac{۲k(m-k)}{m} - \binom{k}{۲} \right) \sum_{\kappa(F)=k} \frac{۱}{\binom{|F|+۲}{۲}} \\
&= w(\mathcal{F}) - \frac{۱}{\binom{m+۲}{۲}} + \frac{۱}{\binom{m}{۲}} \sum_{k<m} \mu(m,k) \sum_{\kappa(F)=k} \frac{۱}{\binom{|F|+۲}{۲}} \ ,
\end{aligned}
$$

که در رابطه‌ی بالا

$$\mu(m,k) = \frac{۲k(m-k)}{m} - \binom{k}{۲}.$$

به ازای هر m داریم:

$$\mu(m,۰) = ۰,$$

$$\mu(m,۱) = ۲ - (\frac{۲}{m}),$$

$$\mu(m,۲) = ۳ - (\frac{۸}{m}),$$

و برای هر $k \ge ۲$ داریم:

$$\mu(m,k) \le \mu(m,۲).$$

پس $\mu(m,k)$ در $k=۱$ یا $k=۲$ ماکزیمم می‌شود. تعریف می‌کنیم:

$$\mu(m)=\max_{k<m}\mu(m,k).$$

اکنون داریم:

$$\begin{aligned}
E(w(\mathcal{F}(p,q))) &\leq w(\mathcal{F})-\frac{۱}{\binom{m+۲}{۲}}+\frac{۱}{\binom{m}{۲}}\sum_{k<m}\mu(m,k)\sum_{\kappa(F)=k}\frac{۱}{\binom{|F|+۲}{۲}}\\
&\leq w(\mathcal{F})-\frac{۱}{\binom{m+۲}{۲}}+\frac{۱}{\binom{m}{۲}}\sum_{k<m}\mu(m)\sum_{\kappa(F)=k}\frac{۱}{\binom{|F|+۲}{۲}}\\
&= w(\mathcal{F})-\frac{۱}{\binom{m+۲}{۲}}+\frac{۱}{\binom{m}{۲}}\mu(m)\sum_{k<m}\sum_{\kappa(F)=k}\frac{۱}{\binom{|F|+۲}{۲}}\\
&= w(\mathcal{F})-\frac{۱}{\binom{m+۲}{۲}}+\frac{۱}{\binom{m}{۲}}\mu(m)\sum_{F\neq M}\frac{۱}{\binom{|F|+۲}{۲}}\\
&= w(\mathcal{F})-\frac{۱}{\binom{m+۲}{۲}}+\frac{۱}{\binom{m}{۲}}\mu(m)\left(w(\mathcal{F})-\frac{۱}{\binom{m+۲}{۲}}\right)\\
&= w(\mathcal{F})-\frac{۱}{\binom{m+۲}{۲}}+\frac{۱}{\binom{m+۲}{۲}}\frac{(m+۲)(m+۱)}{m(m-۱)}\mu(m)\left(w(\mathcal{F})-\frac{۱}{\binom{m+۲}{۲}}\right).
\end{aligned}$$

بنابراین، کافی است ثابت کنیم

$$\frac{(m+۲)(m+۱)}{m(m-۱)}\mu(m)\left(w(\mathcal{F})-\frac{۱}{\binom{m+۲}{۲}}\right)\leq ۱$$

یا به‌طور معادل

$$w(\mathcal{F})\leq\frac{m(m-۱)}{(m+۲)(m+۱)\mu(m)}+\frac{۱}{\binom{m+۲}{۲}}.$$

همان‌طور که در بالا اشاره شد، $\mu(m)$ برابر است با $\mu(m,۱)$ یا $\mu(m,۲)$. اگر

$$\mu(m)=\mu(m,۱),$$

برای $m>۱$ داریم:

$$\frac{m(m-۱)}{(m+۲)(m+۱)\mu(m)}+\frac{۱}{\binom{m+۲}{۲}}=\frac{m^۲+۴}{۲(m+۲)(m+۱)}\geq\frac{۱۳}{۴۰}$$

(تساوی در $m=۳$، برقرار است). اگر

$$\mu(m)=\mu(m,۲),$$

برای $m > 2$ داریم:

$$\frac{m(m-1)}{(m+2)(m+1)\mu(m)} + \frac{1}{\binom{m+2}{2}} = \frac{1}{3} \cdot \frac{m^2(m-1)+6m-16}{(m+2)(m+1)(m-\frac{8}{3})} \geq \frac{49}{153}$$

(تساوی در $m = 16$ برقرار است). از طرفی طبق فرض داریم:

$$w(\mathcal{F}) \leq \frac{49}{153} < \frac{13}{40},$$

لذا ادعا ثابت شد.

با اعمال این ادعا به‌طور مکرر، خانواده‌هایی با اندازه‌ی کوچک‌تر و با وزن حداکثر $\frac{49}{153}$، به‌دست می‌آید. این روش به تطابقی منتهی می‌شود که هر عضو $\mathcal{F}$ شامل یالی از این تطابق است. بنابراین، این تطابق نشان می‌دهد که $\mathcal{F}$ تنک است. □

۳.۲ کران پایین برای حالت کاملاً غیرانطباقی

در این بخش نشان داده می‌شود جواب با عاملی ثابت کران‌دار می‌شود. برای این منظور ابتدا لم ساده‌ی زیر را بیان می‌کنیم.

لم ۱.۳.۲. [۲] *فرض کنید $\mathcal{F}$ خانواده‌ای از زیرمجموعه‌های $V = \{1, 2, \ldots, n\}$ باشد به‌طوری‌که $\mathcal{F}$ مسأله‌ی تطابق روی V را حل می‌کند. برای هر دو عضو متمایز $a, b \in V$، خانواده‌ی*

$$\mathcal{F}_{a,b} = \{F - \{a,b\} \mid F \in \mathcal{F}, \{a,b\} \subseteq F\}$$

تنک نیست.

برهان. فرض کنید $a, b \in V$ وجود دارد به‌طوری‌که $\mathcal{F}_{a,b}$ تنک است. بنابراین، تطابق M روی $V - \{a,b\}$ وجود دارد به‌طوری‌که هر عضو $\mathcal{F}_{a,b}$ شامل یالی از M است. در این صورت برای هر $F \in \mathcal{F}$ پاسخ Q_F برای دو تطابق M و $M \cup \{a,b\}$ یکسان است و این با فرض که $\mathcal{F}$ مسأله‌ی تطابق را حل می‌کند تناقض دارد. □

قضیه ۲.۳.۲. [۲] *فرض کنید $n > 2$، برای هر خانواده‌ی $\mathcal{F}$ که مسأله‌ی تطابق را روی n رأس حل می‌کند*

داریم:

$$|\mathcal{F}| \geq \frac{۴۹}{۱۵۳}\binom{n}{۲}.$$

برهان. فرض کنید $\mathcal{F}$ خانواده‌ای از زیرمجموعه‌های $V = \{۱, ۲, \ldots, n\}$ است که مسأله‌ی تطابق روی V را حل می‌کند. همچنین فرض کنید a و b هر جفت از رئوس متمایز در V باشند. طبق لم ۱.۳.۲ خانواده‌ی

$$\mathcal{F}_{a,b} = \{F - \{a, b\} \mid F \in \mathcal{F}, \{a, b\} \subseteq F\}$$

تنک نیست. لذا طبق لم ۳.۲.۲ داریم:

$$\sum_{F \in \mathcal{F}, \{a,b\} \subseteq F} \frac{۱}{\binom{|F|}{۲}} = \sum_{F' \in \mathcal{F}_{a,b}} \frac{۱}{\binom{|F'|+۲}{۲}} > \frac{۴۹}{۱۵۳}. \qquad (۱.۲)$$

برای هر $F \in \mathcal{F}$، وزن $\frac{۱}{\binom{|F|}{۲}}$ را به هر جفت از اعضای متمایز $a, b \in F$، نسبت می‌دهیم. سهم هر $F \in \mathcal{F}$ در این وزن‌دهی، برابر با ۱ است. لذا وزن کلی توزیع شده به این روش برابر است با $|\mathcal{F}|$، از طرف دیگر طبق (۱.۲) وزن کلی نسبت داده شده به هر جفت $a, b \in V$ حداقل $\frac{۴۹}{۱۵۳}$ است. در نتیجه $|\mathcal{F}| \geq \frac{۴۹}{۱۵۳}\binom{n}{۲}$. □

۴.۲ کران بالا برای حالت کاملاً غیرانطباقی

در این بخش نشان داده می‌شود که خانواده‌هایی با اندازه‌ی $(\frac{۱}{۲} + o(۱))\binom{n}{۲}$ وجود دارند به‌طوری‌که می‌توانند مسأله‌ی تطابق را روی n رأس حل کنند. برای این منظور، ابتدا مسأله‌ی تطابق را تعمیم می‌دهیم. گفته می‌شود خانواده‌ی $\mathcal{F}$ از خوشه‌هایی که در G قرار دارند، مسأله‌ی تطابق را روی G حل می‌کند اگر برای هر دو تطابق متمایز در G، حداقل یک خوشه در $\mathcal{F}$ موجود باشد که یالی از یکی از این تطابق‌ها را شامل شود ولی هیچ یالی از تطابق دیگر شامل نشود. برای مثال، اگر G یک مثلث باشد آنگاه $\mathcal{F}$ باید مجموعه‌ای از یال‌ها باشد. در این صورت، مسأله‌ی تطابق روی n رأس با مسأله‌ی تطابق روی گراف K_n معادل می‌شود. اندازه‌ی کوچک‌ترین خانواده که مسأله‌ی تطابق را روی G حل می‌کند با $f(G)$ مشخص می‌کنیم.

در طول این بخش، فرض کنید $E(H)$ مجموعه یال‌های گراف H و $\gcd(H)$ بزرگ‌ترین مقسوم‌علیه مشترک درجات همه‌ی رئوس H را مشخص می‌کند.

قضیه ۲.۴.۱. [۱۵] **(ویلسون)** *برای هر گراف H عدد ثابت N وجود دارد به‌طوری‌که برای هر $n \geq N$، K_n با اجتماع $\frac{\binom{n}{2}}{|E(H)|}$ گراف دوبه‌دو یال-مجزای یکریخت با H برابر است، اگر و تنها اگر $\binom{n}{2}$ بر $|E(H)|$ و $n-1$ بر $\gcd(H)$ قابل قسمت باشد.*

دقت کنید که اگر $n-1$ بر $2|E(H)|$ قابل قسمت باشد آنگاه شرایط ویلسون برقرار است. یعنی، گراف K_n قابل تجزیه به گراف‌های دوبه‌دو یال-مجزای یکریخت با گراف H است.

نتیجه ۲.۴.۲. [۲] *برای هر گراف معین H،*

$$f(K_n) \leq \frac{f(H)}{|E(H)|}\binom{n}{2} + O(n).$$

همچنین، به ازای گراف معین H، راه حل مسأله‌ی تطابق برای K_n، سازنده است.

برهان. c رأس افزوده را به n رأس اضافه می‌کنیم به‌طوری‌که

$$n + c - 1 \equiv 0 \pmod{2|E(H)|},\ 0 \leq c \leq 2|E(H)| - 1$$

تا گراف K_{n+c} قابل تجزیه به گراف H شود. لذا طبق قضیه‌ی ویلسون K_{n+c} برابر است با اجتماع

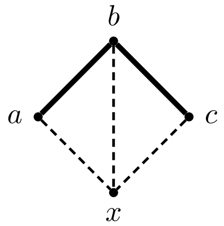
$$\frac{\binom{n+c}{2}}{|E(H)|} = \frac{\binom{n}{2}}{|E(H)|} + O(n)$$

گراف دوبه‌دو یال-مجزای یکریخت با H. مسأله‌ی تطابق را در هریک از این گراف‌ها با استفاده از خانواده‌ای از اندازه‌ی $f(H)$ حل می‌کنیم. □

تعریف ۲.۴.۳. [۲] گفته می‌شود خانواده‌ی $\mathcal{F}$ یال e را تعیین وضعیت می‌کند اگر برای هر دو تطابق M_1 و M_2 که $e \in M_1$ و $e \notin M_2$ حداقل یک خوشه در $\mathcal{F}$ باشد که این خوشه یالی از یکی از این تطابق‌ها را شامل شود ولی هیچ یالی از دیگری را شامل نشود.

طبق تعریف بالا به‌سادگی می‌توان دید خانواده‌ی $\mathcal{F}$ مسأله‌ی تطابق را حل می‌کند اگر و تنها اگر $\mathcal{F}$ تمامی یال‌ها را تعیین وضعیت کند.

لم ۴.۴.۲. [۲] **(دو-سوم)** *چهار رأس متمایز a، b، c و x را در نظر بگیرید. فرض کنید خانواده‌ی $\mathcal{F}$ یال‌های $\{a,b\}$ و $\{b,c\}$ را تعیین وضعیت می‌کند. اگر $\mathcal{F}$ شامل دو مثلث $\{a,b,x\}$ و $\{b,c,x\}$ باشد، آنگاه $\mathcal{F}$ یال‌های $\{a,x\}$ و $\{b,x\}$ و $\{c,x\}$ را نیز تعیین وضعیت می‌کند.*



برهان. ابتدا نشان می‌دهیم $\mathcal{F}$ یال $\{b,x\}$ را تعیین وضعیت می‌کند. فرض کنید M_1 و M_2 دو تطابق دلخواه باشند که $\{b,x\} \in M_1$ و $\{b,x\} \notin M_2$. باید نشان دهیم حداقل یک خوشه در $\mathcal{F}$ وجود دارد به‌طوری‌که این خوشه شامل یالی از یکی از تطابق‌های M_i می‌شود ولی هیچ یالی از دیگری را شامل نمی‌شود.

اگر هیچ یک از یال‌های مثلث $\{a,b,x\}$ در M_2 نباشد، از آنجا که $\mathcal{F}$ شامل این مثلث است، بنابراین این مثلث خوشه‌ی مورد نظر خواهد بود. در غیر این صورت، یکی از دو حالت زیر اتفاق می‌افتد:

(i) $\{a,b\} \in M_2$

(ii) $\{a,x\} \in M_2$

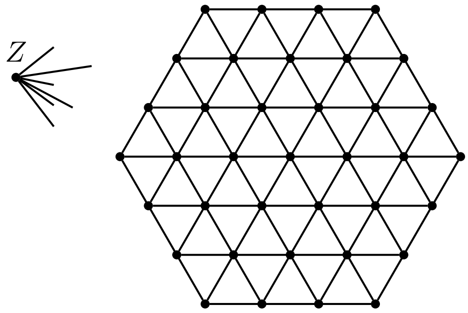
در حالت (i) چون $\{a,b\} \notin M_1$ و طبق فرض، $\mathcal{F}$ یال $\{a,b\}$ را تعیین وضعیت می‌کند لذا خوشه‌ی مورد نظر همان خوشه‌ای است در $\mathcal{F}$ که یال $\{a,b\}$ را تعیین وضعیت می‌کند.

برای حالت (ii) نیز تنها یکی از دو زیرحالت زیر اتفاق می‌افتد:

(آ) $\{b,c\} \in M_2$

(ب) $\{b,c\} \notin M_2$

در حالت (آ) چون $\{b,c\} \notin M_1$ و $\mathcal{F}$ یال $\{b,c\}$ را تعیین وضعیت می‌کند بنابراین، خوشه‌ی مورد نظر در $\mathcal{F}$ موجود است.



شکل ۱.۲: گراف HEX_s^+

در حالت (ب) چون $\{c, x\}, \{b, c\} \notin M_۲$ پس مثلث $\{b, c, x\}$ خوشه‌ی مورد نظر خواهد بود. بنابراین، $\mathcal{F}$ یال $\{b, x\}$ را تعیین وضعیت می‌کند.

حال نشان می‌دهیم $\mathcal{F}$ یال $\{a, x\}$ را تعیین وضعیت می‌کند. مانند قبل فرض کنید $M_۱$ و $M_۲$ دو تطابق دلخواه باشند که $\{a, x\} \in M_۱$ و $\{a, x\} \notin M_۲$. اگر هیچ یالی از مثلث $\{a, b, x\}$ در $M_۲$ نباشد، آنگاه این مثلث خوشه‌ی مورد نظر خواهد بود. در غیر این صورت، اگر یکی از یال‌های $\{a, b\}$ یا $\{b, x\}$ در $M_۲$ باشد با توجه به این‌که این یال در $M_۱$ نیست و $\mathcal{F}$ این یال را تعیین وضعیت می‌کند بنابراین، خوشه‌ای در $\mathcal{F}$ هست که یال $\{a, x\}$ را تعیین وضعیت کند.

به‌طور مشابه، $\mathcal{F}$ یال $\{c, x\}$ را تعیین وضعیت می‌کند. $\square$

دقت شود با به‌کار بردن لم دو-سوم می‌توان جوابی از اندازه‌ی $\frac{۲}{۳}\binom{n}{۲} + O(n)$ برای مسأله‌ی تطابق روی n رأس به‌دست آورد. اکنون گرافی را معرفی می‌کنیم که به کمک آن می‌توان کران بالای به‌دست آمده برای مسأله‌ی تطابق روی n رأس را بهبود داد.

تعریف ۵.۴.۲. [۲] فرض کنید $s \geq ۱$ عددی صحیح است. شش ضلعی با اضلاع به طول s را با مثلث‌های متساوی‌الاضلاع به ضلع واحد فرش می‌کنیم (یعنی، شش ضلعی را طوری افراز می‌کنیم که هر بخش یک مثلث متساوی‌الاضلاع به ضلع واحد باشد). رأس Z را اضافه کرده، یال‌هایی از Z به سایر رئوس رسم می‌کنیم. گراف به‌دست آمده را $\mathbf{HEX_s^+}$ می‌نامیم که در شکل ۱.۲ رسم شده است.

شش ضلعی فرش‌شده با مثلث‌ها دارای

$$v = ۳s^۲ + ۳s + ۱$$

رأس،

$$e = ۹s^۲ + ۳s$$

یال و

$$f = ۶s^۲$$

مثلث است. بنابراین، گراف HEX_s^+ دارای

$$v + e = ۱۲s^۲ + ۶s + ۱$$

یال است.

مسأله‌ی تطابق روی HEX_s^+ با گروه‌های زیر حل می‌شود:

چهار وجهی: $T \cup \{Z\}$، برای هر مثلث T در فرش.

مرز: هر یال مرزی از فرش و هر یال از Z به رأسی روی مرز.

اکنون بررسی می‌کنیم چرا این گروه‌ها کفایت می‌کنند. حالت‌های زیر را در نظر بگیرید.

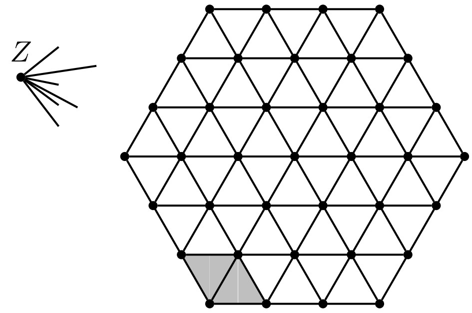
حالت ۱: برای رأس Y روی مرز فرش، $ZY \in M$.

این یال در گروه‌ها موجود است، بنابراین می‌دانیم $ZY \in M$. در این حالت، وضعیت دیگر یال‌های مجاور با Z مشخص است لذا گروه‌های چهار وجهی با مثلث‌ها متناظر می‌شوند. در نتیجه، کافی است لم دو ـ سوم را با شروع از مرز فرش و از گوشه‌ی شش ضلعی، به‌طور مکرر به‌کار ببریم تا دیگر یال‌های تطابق مورد نظر را شناسایی کنیم (شکل‌های ۲.۲ و ۳.۲).

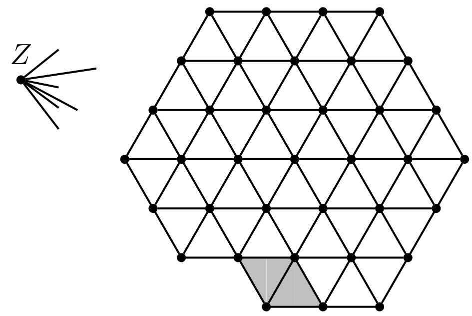
حالت ۲: برای رأس Y درون فرش، $ZY \in M$.

در این حالت شش چهاروجهی شامل Y دارای جواب "مثبت" هستند. چنین حالتی را برای Y **ویژگی ۶-مثلث** نامیده و هر شش مثلث دوبه‌دو مجاور را heg_1^+ می‌نامیم (شکل ۴.۲).

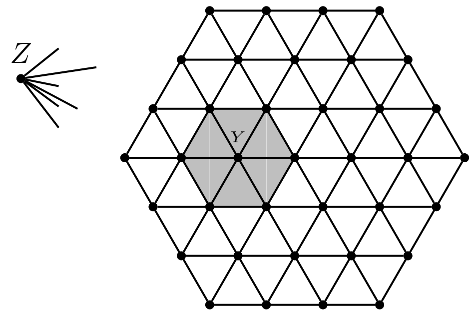
زیرحالت‌های زیر را درنظر بگیرید.



شکل ۲.۲: شروع از گوشه با استفاده از لم دو-سوم



شکل ۳.۲: طبق لم دو-سوم ادامه دهید



شکل ۴.۲: نقطه‌ی Y با ویژگی ۶-مثلث و heg_1^+

(i) تنها یک رأس دارای ویژگی ۶-مثلث باشد. در این صورت می‌دانیم این رأس در تطابق با Z مجاور است. چون، به سادگی می‌توان دید هیچ تطابقی نمی‌تواند مثلث‌های heg_1^+ را به‌طور هم‌زمان "مثبت" کند مگر این‌که یکی از رئوس این شش مثلث با Z مجاور باشد (هر تطابق روی heg_1^+، دورانی از یکی از حالت‌های شکل ۵.۲ خواهد بود اگر یال‌های هاشورخورده را یال‌های تطابق در نظر بگیریم و اگر هیچ رأسی از heg_1^+ با Z مجاور نباشد دو مثلث رنگی نمی‌توانند هیچ یالی از تطابق را اختیار کنند لذا "منفی" خواهند بود) و اگر رأس دیگری با Z مجاور باشد آنگاه آن رأس نیز باید دارای ویژگی ۶-مثلث باشد.

(ii) سه رأس متمایز دارای ویژگی ۶-مثلث باشند.

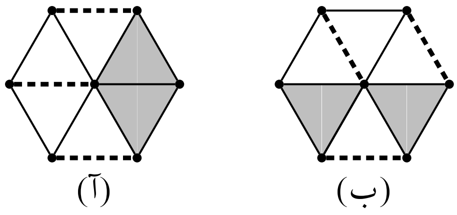
از آنجا که Z فقط با یک رأس مجاور است لذا هر heg_1^+ باید شامل این رأس (جفت Z) باشد. بنابراین، این سه رأس در یک خط مستقیم با هم مجاورند و رأس میانی با Z مجاور است (شکل ۶.۲). فرض کنید این سه رأس در یک خط مستقیم قرار نداشته باشند، طبق بالا رأس میانی باید با Z مجاور باشد (رأس Y)، هر تطابق روی دیگر یال‌ها در واقع دورانی از یکی از حالات شکل ۷.۲ خواهد بود و در آن‌ها حداقل یک مثلث از یک heg_1^+، "منفی" خواهد بود، پس این حالت اتفاق نمی‌افتد.

در نتیجه حداکثر سه رأس می‌تواند دارای ویژگی ۶-مثلث باشد، چون در غیر این صورت یا این رئوس غیر مجاورند یا در یک خط مستقیم قرار ندارند. در هر یک از این زیرحالت‌ها جستجو را مانند حالت ۱ تمام می‌کنیم.

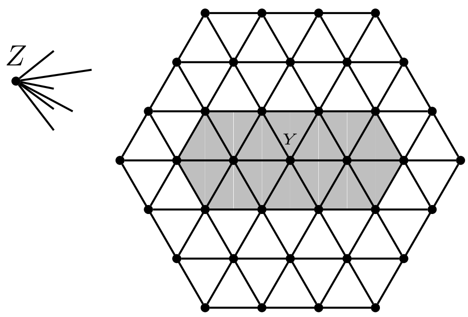
(iii) اگردقیقاً دو رأس دارای ویژگی ۶-مثلث باشد، آن دو رأس مجاورند (مشابه زیرحالت (ii)). مانند حالت ۱ جستجو را ادامه می دهیم تا به ۱۰ مثلث شامل این دو رأس برسیم. به کمک پرسش‌هایی که قبلاً انجام شده‌اند، یال‌های اطراف این ۱۰ مثلث نیز تعیین وضعیت می‌شوند (به‌وضوح، برخی از این یال‌ها در تطابق نیستند). آن رأس با ویژگی ۶-مثلث، جفت Z است که در یک مثلث با یال‌های غیر-تطابق قرار دارد.

حالت ۳: برای هر Y در فرش، $ZY \notin M$.

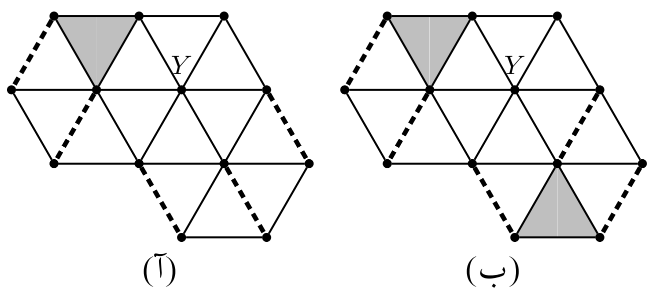
در این صورت طبق زیرحالت (ii)، هیچ نقطه‌ای ویژگی ۶-مثلث را ندارد. بنابراین، می‌دانیم Z رأسی از



شکل ۵.۲: با هر تطابق روی heg_1^+ حداقل دو مثلث، "منفی" می‌ماند



شکل ۶.۲: ۳ نقطه با ویژگی ۶-مثلث



شکل ۷.۲: ۳ نقطه با ویژگی ۶-مثلث در یک خط مستقیم مجاور نیستند

تطابق نیست. پس مانند حالت ۱ جستجو را تمام می‌کنیم.

تعداد کلی گروه‌های ساخته شده با این روش، برابر است با

$$f + ۱۲s = ۶s^۲ + ۱۲s.$$

نسبت تعداد گروه‌ها به تعداد یال‌ها برابر است با

$$\begin{aligned}\frac{۶s^۲ + ۱۲s}{۱۲s^۲ + ۶s + ۱} &= \frac{۱}{۲} + \frac{۱۸s - ۱}{۲۴s^۲ + ۱۲s + ۲}\\ &= \frac{۱}{۲} + O(\frac{۱}{s}).\end{aligned}$$

بنابراین با توجه به نتیجه‌ی ۲.۴.۲

$$f(K_n) = \left(\frac{۱}{۲} + O(\frac{۱}{s})\right)\binom{n}{۲}.$$

فرض کنید s تابع صعودی ضعیفی [۴] از n باشد، لذا داریم:

نتیجه ۶.۴.۲. [۲]

$$f(K_n) \leq \left(\frac{۱}{۲} + o(۱)\right)\binom{n}{۲}.$$

۵.۲ الگوریتم‌های احتمالاتی غیرانطباقی

در این بخش، الگوریتم تصادفی کارآمدی برای حل مسأله‌ی جستجوی تطابق ارائه می‌شود. ساده‌ترین مدل این الگوریتم، تعداد $bn\log n$ زیرمجموعه‌ی تصادفی از اندازه‌ی $c\sqrt{n}$ را مورد پرسش قرار می‌دهد. همچنین نشان داده می‌شود که با انتخاب مناسب c و b، الگوریتم با احتمال بالا، مسأله‌ی تطابق را در یک مرحله حل می‌کند. از آنجا که این الگوریتم و برخی الگوریتم‌های جایگزین آن ممکن است به‌طور عملی مورد استفاده قرار گیرند، به‌منظور بهینه کردن مقادیر ثابت به‌دست آمده در تخمین تعداد کلی تست‌ها، تلاش‌هایی انجام شده است. همچنین نتیجه می‌شود که برای بهبود این مقادیر، بهتر است گروه‌ها براساس یک صفحه تصویری تصادفی اصلاح شده طرح شوند. ابتدا به‌طور مختصر، برخی ویژگی‌های یک صفحه تصویری را بیان می‌کنیم.

[۴] slowly growing function

تعریف ۱.۵.۲. یک صفحه تصویری متشکل است از مجموعه‌ای به نام نقاط و مجموعه‌ای دیگر موسوم به مجموعه‌ی خطوط (شامل نقاط) و یک رابطه‌ی وقوع به‌طوری‌که اصول زیر برقرار باشد.

(i) از هر دو نقطه دقیقاً یک خط می‌گذرد.

(ii) هر دو خط دقیقاً در یک نقطه یکدیگر را قطع می‌کنند.

(iii) چهار نقطه موجودند که هیچ سه‌تایی از آن‌ها روی یک خط واقع نیستند.

(iv) هر گزاره‌ی درست برای صفحه تصویری، گزاره‌ی درست دیگری به نام دوگان آن را نتیجه می‌دهد که در آن نقطه و خط جابه‌جا شده‌اند.

صفحه تصویری متناهی، صفحه‌ای با تعداد نقاط متناهی است که تعداد نقاط هر خط $p+۱$ و طبق (iv) هر خط دارای $p+۱$ نقطه است. در این صورت صفحه تصویری را از مرتبه‌ی p می‌نامیم. تعداد نقاط یک صفحه تصویری متناهی از مرتبه‌ی p (و همچنین تعداد خطوط) برابر $p^۲+p+۱$ است. این صفحه برای هر عدد اول p وجود دارد و می‌توان آن را با ساختار جبری زیر ساخت.

فرض کنید B مجموعه‌ی همه‌ی بردارهای ناصفر $\bar{x}=(x_۰,x_۱,x_۲)$، به‌طول ۳، روی میدان متناهی $GF(p)$ باشد. یک رابطه‌ی هم‌ارزی روی B به‌صورت زیر تعریف می‌کنیم.

دو بردار $\bar{x}_۱$ و $\bar{x}_۲$ هم‌ارزند اگر یکی مضرب دیگری باشد. یعنی

$$\bar{x}_۱ = k\cdot\bar{x}_۲ \ ; \ k\in GF(p).$$

نقاط و خطوط صفحه تصویری با کلاس‌های هم‌ارزی این رابطه روی B نشان داده می‌شوند به‌طوری‌که نقطه‌ی $\bar{x}=(x_۰,x_۱,x_۲)$ بر روی خط $\bar{y}=(y_۰,y_۱,y_۲)$ قرار می‌گیرد اگر و تنها اگر ضرب داخلی $\bar{x}$ و $\bar{y}$ روی $GF(p)$ برابر صفر باشد. یعنی

$$\langle\bar{x},\bar{y}\rangle = x_۰y_۰+x_۱y_۱+x_۲y_۲ = ۰.$$

۱.۵.۲ روش RPP: گروه‌بندی براساس یک صفحه تصویری تصادفی

فرض کنید $n = p^۲ + p + ۱$ و p عددی اول است. گروه‌ها را براساس یک صفحه تصویری تصادفی به‌صورت زیر طرح می‌کنیم.

n رأس را، به‌طور تصادفی، جایگشت می‌دهیم و آن‌ها را به‌عنوان نقاط صفحه تصویری P از مرتبه‌ی p و سپس هر خط این صفحه را به‌عنوان یک گروه درنظر می‌گیریم.

اکنون زوج (x,y) را در نظر می‌گیریم. دقیقاً یک خط در P شامل x و y است. احتمال این‌که این خط شامل یالی از M نباشد، به جز امکان (x,y)، حداقل برابر است با این احتمال زمانی‌که تطابق کامل است و xy یال تطابق نیست. در این مورد این احتمال برابر است با

$$\frac{\binom{n-۴}{۱}\binom{n-۶}{۱}\cdots\binom{n-۴-۲(p-۲)}{۱}}{\binom{n-۲}{۱}\binom{n-۳}{۱}\cdots\binom{n-۲-(p-۲)}{۱}} = \frac{(n-۴)(n-۶)\cdots(n-۲p)}{(n-۲)(n-۳)\cdots(n-p)}.$$

در حقیقت، صورت کسر برابر است با تعداد راه‌های انتخاب یک مجموعه‌ی مرتب از $p-۱$ نقطه‌ی دیگر از خط (به جز x و y) بدون دربرگرفتن هیچ یالی از تطابق. بعد از انتخاب i نقطه (با احتساب x و y) که قبلاً انتخاب شده‌اند، $n-۲i$ امکان برای انتخاب نقطه‌ی بعدی، که باید از نقاط قبلی و جفت‌های آن‌ها متمایز باشد، وجود دارد. مخرج نیز برابر است با تعداد کلی حالات انتخاب یک مجموعه‌ی مرتب شامل $p-۱$ نقطه. عبارت آخر برابر است با

$$\frac{(۱-\frac{۴}{n})\cdots(۱-\frac{۲p}{n})}{(۱-\frac{۲}{n})\cdots(۱-\frac{p}{n})}. \tag{۲.۲}$$

با استفاده از رابطه‌ی

$$(۱-\tfrac{i}{n})^{\frac{n}{i}-۱} \geq e^{-۱}$$

صورت (۲.۲) برابر است با

$$\begin{aligned}\prod_{i=۲}^{p}(۱-\tfrac{۲i}{n}) &\geq \prod_{i=۲}^{p} e^{\frac{-۲i}{n-۲i}}\\ &= e^{\sum_{i=۲}^{p}\frac{-۲i}{n-۲i}}\\ &\geq e^{\frac{-(p^۲+p)}{n-۲p}},\end{aligned}$$

همچنین با استفاده از رابطه‌ی

$$(1 - \tfrac{i}{n})^{\frac{n}{i}} \leq e^{-1}$$

مخرج (۲.۲) برابر است با

$$\begin{aligned}\prod_{i=2}^{p}(1 - \tfrac{i}{n}) &\leq \prod_{i=2}^{p} e^{\frac{-i}{n}} \\ &= e^{\sum_{i=2}^{p}\frac{-i}{n}} \\ &= e^{\frac{-(p^2+p-2)}{2n}}.\end{aligned}$$

بنابراین، مقدار (۲.۲) حداقل برابر است

$$\begin{aligned}e^{\frac{-(p^2+p)}{n-2p}+\frac{p^2+p-2}{2n}} &= e^{-(1-o(1))\frac{p^2}{2n}} \\ &= e^{-\frac{1}{2}}(1-o(1)).\end{aligned}$$

۲.۵.۲ گروه‌بندی براساس $d\log n$ صفحه تصویری تصادفی

روش RPP را $d\log n$ مرتبه، به‌طور مستقل و هم‌زمان، به ازای عدد حقیقی d انجام می‌دهیم. احتمال این‌که هر خط شامل x و y شامل یالی از M (به غیر از امکان (x,y)) شود حداکثر برابر است با

$$\pi(d) = \Big((1 - e^{-\frac{1}{2}})^{d}(1 + o(1))\Big)^{\log n}.$$

اگر قرار دهیم

$$d = \frac{(1 + o(1))\ln 2}{\ln((1 - e^{-\frac{1}{2}})^{-1})} \approx 0{,}74,$$

آنگاه

$$\pi(d) \leq \frac{1}{n}.$$

از طرفی، اگر هر خط شامل x و y، شامل یالی از M (به جز امکان xy) باشد، در این صورت پاسخ همواره مثبت خواهد بود. لذا یال xy در این گروه‌ها شناسایی نمی‌شود. فرض کنید متغیر تصادفی X برابر باشد با

تعداد یال‌های غیرتطابق که با این گروه‌ها شناسایی نمی‌شوند. می‌توان گفت

$$X = \sum_{e \in E(K_n) \setminus M} X_e$$

به‌طوری‌که

$$X_e = \begin{cases} ۱ & \text{یال } e \text{ شناسایی نشود} \\ ۰ & o.w \end{cases}$$

طبق خاصیت خطی بودن میانگین داریم:

$$\begin{aligned} E(X) &= \sum_{e \in E(K_n) \setminus M} E(X_e) \\ &\leq (\binom{n}{۲} - |M|)\frac{۱}{n} \\ &\leq \frac{۱}{n}\binom{n}{۲} \\ &\leq \frac{n}{۲}\,. \end{aligned}$$

در نتیجه، این گروه‌ها برای شناسایی همه‌ی یال‌های غیرتطابق به جز، به‌طور میانگین، $\frac{n}{۲}$ آن‌ها کفایت می‌کنند.

همه‌ی یال‌های غیرتطابق باقی‌مانده و همه‌ی یال‌های تطابق در مرحله‌ی دوم تنها با n گروه شناسایی می‌شوند. پس این الگوریتم یک الگوریتم ۲ - مرحله‌ای است که مسأله‌ی تطابق را با ایجاد تقریباً $۰/۷۴n \log n$ گروه، بدون خطا حل می‌کند.

اگر d را تقریباً دو برابر انتخاب کنیم، یعنی $d = \frac{۲\ln ۲}{\ln(۱/(۱-e^{-\frac{۱}{۲}}))} + \epsilon$، به ازای مقدار دلخواه و کوچک ϵ (مثلاً $d \approx ۱/۴۹$)، آنگاه

$$\pi(d) = o(\frac{۱}{n^۲}).$$

در نتیجه، این گروه‌ها همه‌ی یال‌های غیرتطابق را با احتمال بالا در یک مرحله شناسایی می‌کنند و با استدلال مشابه نشان داده می‌شود که همه‌ی یال‌های تطابق نیز با احتمال بالا شناسایی می‌شوند. بنابراین یک الگوریتم ۱ - مرحله‌ای به‌دست می‌آید که مسأله‌ی تطابق را با ایجاد تقریباً $۱/۴۹n \log n$ گروه، بدون خطا حل می‌کند.

۳.۵.۲ دوبرابر کردن نقطه

ثابت‌های به‌دست آمده در روش قبل را می‌توان بهبود داد. این کار را می‌توان با به‌طور تصادفی ”دوبرابر کردن“ [5] برخی نقاط در صفحه تصویری به آسانی انجام داد.

نقطه‌ی x را با افزودن نقطه‌ی جدید x'، به هر یک از خطوطی که شامل x است، دو برابر می‌کنیم و x' را دو برابر شده‌ی x می‌نامیم.

به‌طور دقیق‌تر، فرض کنید $n = \lceil(۲\ln ۲)m\rceil$، که $m = p^۲ + p + ۱$ و p عددی اول است. با یک صفحه تصویری از مرتبه‌ی p شروع می‌کنیم و $\lceil(۲\ln ۲ - ۱)m\rceil$ نقطه را، که به‌طور تصادفی انتخاب شده‌اند، دو برابر می‌کنیم. لذا n نقطه تولید می‌شود. اما هنوز تعداد خطوط برابر است با $m \approx \frac{n}{(۲\ln ۲)}$. طبق تخمین استاندارد ۲.۱.۱، با احتمال بالا تعداد نقاط هریک از این خطوط تقریباً برابر است با

$$۲\ln ۲\sqrt{m} \approx \sqrt{(۲\ln ۲)n}.$$

اکنون یک صفحه تصویری اصلاح شده، جفتی از نقاط x و y و خطی شامل x و y در نظر بگیرید (اگر نقاط x و y دو برابر شده‌ی دیگری نباشند، این خط یکتا است، در غیر این صورت $p + ۱$ خط شامل x و y وجود دارد، لذا کافی است یکی از آن‌ها را در نظر بگیرید.) فرض کنید $t \approx \sqrt{(۲\ln ۲)n}$ تعداد نقاط روی این خط باشد. با استدلالی مشابه قبل، احتمال این‌که این خط شامل هیچ یالی از M، به جز امکان xy، نباشد حداقل برابر است با

$$\begin{aligned}
\frac{\binom{n-۴}{۱}\binom{n-۶}{۱}\cdots\binom{n-۴-۲(t-۳)}{۱}}{\binom{n-۲}{۱}\binom{n-۳}{۱}\cdots\binom{n-۲-(t-۳)}{۱}} &= \frac{(n-۴)(n-۶)\cdots(n-۲t+۲)}{(n-۲)(n-۳)\cdots(n-t+۱)} \\
&= e^{-(۱+o(۱))\frac{t^۲}{۴n}} \\
&= \frac{۱}{۲}(۱+o(۱)).
\end{aligned}$$

$d\log n$ صفحه تصویری تصادفی با نقاط دو برابر شده به‌صورت بالا، در نظر بگیرید. احتمال این‌که هر خط شامل x و y شامل یالی از M، به جز امکان xy، باشد حداکثر برابر است با

$$\pi'(d) = \Big((\frac{۱}{۲})^d(۱+o(۱))\Big)^{\log n}.$$

[5] Point doubling

بنابراین

$$\pi'(۱ + o(۱)) = \frac{۱}{n}$$

و

$$\pi'(۲ + o(۱)) = \frac{۱}{n^۲}.$$

از آنجا که هر طرح شامل تقریباً $\frac{n}{(۲\ln ۲)}$ خط است، نتیجه می‌گیریم:

قضیه ۲.۵.۲. [۲] *مسأله‌ی تطابق روی n رأس توسط الگوریتم‌های احتمالاتی با شاخص‌های زیر حل می‌شود:*

- ۲ *مرحله و* $(\frac{۱}{(۲\ln ۲)})n\log n(۱ + o(۱)) \approx ۰/۷۲ n\log n$ *پرسش؛*
- ۱ *مرحله و* $(\frac{۱}{\ln ۲})n\log n(۱ + o(۱)) \approx ۱/۴۴ n\log n$ *پرسش.*

زمانی که الگوریتم‌ها تعیین می‌کنند کدام یال‌ها، یال‌های تطابق و کدام یال‌ها غیرتطابقی هستند، هیچ خطایی وجود ندارد. با احتمال بالا همه‌ی اطلاعات با یک الگوریتم ۱-مرحله‌ای به‌دست می‌آید. اما در برخی موارد این چنین نیست و با دانستن این موضوع، مجبوریم یک مجموعه‌ی افزوده از پرسش‌ها برای شناسایی همه‌ی یال‌هایی که تعیین وضعیت نشده‌اند بسازیم. در الگوریتم‌های ۲-مرحله‌ای، همواره همه‌ی اطلاعات در مورد وضعیت یال‌ها به‌دست می‌آید.

۶.۲ الگوریتم‌های قطعی چند مرحله‌ای

در این بخش الگوریتم‌های قطعی چند مرحله‌ای ارائه می‌شود. تعداد پرسش‌هایی که این الگوریتم‌ها در هر مرحله مطرح می‌کنند حداکثر برابر است با $O(n^{۱+\frac{۱}{۲(k-۱)}}\operatorname{poly}\log n)$. در مورد خاص $k = ۲$، به عامل $\operatorname{poly}\log n$ نیازی نیست.

در مورد عدد رنگی یک گراف که شامل همه‌ی یال‌هایی است که در نیمی از خطوط یک صفحه تصویری قرار دارند، یک لم وجود دارد که کمک می‌کند مسأله‌ی متداول را به مسأله‌ی تطابق دوبخشی تبدیل کنیم. این لم، با درنظر گرفتن مقادیر ویژه‌ی ماتریس وقوع صفحه اثبات می‌شود.

لم ۲.۶.۱. [۲] **(رنگ‌آمیزی)** *فرض کنید P یک صفحه تصویری متناهی با n رأس باشد. Q را با حذف حداقل $\frac{n}{2}$ از خطوط P به‌دست می‌آوریم. فرض کنید $G=(V,E)$ که V مجموعه‌ی نقاط P و E شامل همه‌ی زوج‌های (x,y) است به‌طوری‌که خطی در Q شامل x و y وجود داشته باشد. در این صورت G، $\sqrt{n}\ln n(1+o(1))$ رنگ‌پذیر است و اندازه‌ی هر کلاس رنگی کم‌تر از $\sqrt{n}$ است. همچنین، این رنگ‌آمیزی را می‌توان در زمان چند جمله‌ای برحسب n پیدا کرد.*

برهان. مجموعه‌ی B از نقاط را در نظر بگیرید. برای خط l، متغیر تصادفی X_l را برابر با تعداد نقاط B که در خط l قرار دارند، تعریف می‌کنیم. لذا می‌توان گفت

$$X_l=\sum_{v\in B}X_v$$

که در آن

$$X_v=\begin{cases}1 & v\in l\\ 0 & o.w\end{cases}$$

درنتیجه

$$E(X_l)=\sum_{v\in B}E(X_v)=\frac{|B|}{\sqrt{n}}.$$

یعنی، به‌طور میانگین، یک خط در حدود $\frac{|B|}{\sqrt{n}}$ نقطه از B خواهد داشت. نشان می‌دهیم اگر B خیلی کوچک نباشد آنگاه اکثر خطوط حداقل نیمی از نقاط را شامل می‌شوند. این موضوع به ما اجازه خواهد داد کلاس‌های رنگی را از میان خطوطی که از P حذف شده‌اند انتخاب کنیم. برای این منظور به لم زیر نیازمندیم.

لم ۲.۶.۲. [۳] *فرض کنید $G=(U,V;E)$ یک گراف d- منتظم دوبخشی با بخش‌های U و V که*

$$|U|=|V|=n$$

و $A=(A_{u,v}:u\in U,v\in V)$ ماتریس مجاورت (دوبخشی) G باشد به این صورت که $A_{u,v}=1$ اگر و تنها اگر $uv\in E$ و در غیر این صورت $A_{u,v}=0$. همچنین، فرض کنید هر مقدار ویژه‌ی A^tA به جز بزرگ‌ترین (که برابر است با d^2) حداکثر برابر است با λ^2. در این صورت برای هر $B\subseteq V$،

$$\sum_{u\in U}\Big(|N(u)\cap B|-d\frac{|B|}{n}\Big)^2\le\lambda^2|B|\Big(1-\frac{|B|}{n}\Big).\qquad(۲.۳)$$

برهان. به ازای هر $u \in U$

$$|N(u) \cap B| = \sum_{v \in N(u)} X_v$$

به‌طوری‌که

$$X_v = \begin{cases} ۱ & v \in B \\ ۰ & o.w \end{cases}$$

لذا طبق خاصیت خطی بودن میانگین داریم:

$$E(|N(u) \cap B|) = \sum_{v \in N(u)} E(X_v) = d \cdot \frac{|B|}{n}.$$

قرار می‌دهیم

$$b = \frac{|B|}{n}.$$

تابع $f : V \rightarrow R$ را تعریف می‌کنیم:

$$f(v) = \begin{cases} ۱ - b & v \in B \\ -b & o.w \end{cases}$$

داریم:

$$\sum_{v \in V} f(v) = bn(۱ - b) + (n - bn)(-b) = ۰$$

یعنی f بر بردار ویژه‌ی بزرگ‌ترین مقدار ویژه‌ی $A^t A$ عمود است. بنابراین

$$(Af, Af) \leq \lambda^۲ (f, f).$$

از طرفی داریم:

$$\lambda^۲ (f, f) = \lambda^۲ \Big(|B|(۱ - b)^۲ + (n - |B|) b^۲ \Big)$$

$$= \lambda^۲ b(۱ - b) n,$$

و همچنین

$$(Af, Af) = \sum_{u \in U} \left(|N(u) \cap B| - d \cdot \frac{|B|}{n} \right)^۲,$$

پس رابطه‌ی (۳.۲) اثبات می‌شود. □

فرض کنید P یک صفحه تصویری از مرتبه p باشد. بنابراین P دارای $n = p^۲ + p + ۱$ نقطه است. فرض کنید G گراف وقوع P باشد، یعنی، گرافی دوبخشی با بخش‌های U و V که $|U| = |V| = n$، به‌طوری‌که V مجموعه‌ی نقاط و U مجموعه‌ی خطوط است و uv یک یال در G است اگر و تنها اگر خط u شامل نقطه‌ی v باشد. اگر A ماتریس مجاورت G باشد آنگاه A^tA ماتریسی است که همه‌ی درایه‌های قطری آن $p+۱$ و دیگر درایه‌های آن ۱ است. در نتیجه، بزرگ‌ترین مقدار ویژه‌ی A^tA برابر است با $(p+۱)^۲$ و دیگر مقادیر ویژه‌ی آن برابر است با p. بنابراین، برای هر مجموعه از نقاط $B \subset V$، می‌توان تعداد خطوط u را، که شامل کم‌تر از $\frac{d|B|}{۲n}$ نقطه از B هستند، با استفاده از لم بالا کران‌دار کرد. یعنی اگر قرار دهیم

$$L = \left\{ u \in U : |N(u) \cap B| < \frac{d}{۲} \frac{|B|}{n} \right\}$$

آنگاه داریم:

$$\begin{aligned}
\lambda^۲ |B| \Big(۱ - \frac{|B|}{n} \Big) &\geq \sum_{u \in U} \Big(|N(u) \cap B| - d \frac{|B|}{n} \Big)^۲ \\
&\geq \sum_{u \in L} \Big(|N(u) \cap B| - d \frac{|B|}{n} \Big)^۲ \\
&> \Big(\frac{d|B|}{۲n} \Big)^۲ |L|.
\end{aligned}$$

بنابراین

$$\begin{aligned}
\Big| \Big\{ u \in U : |N(u) \cap B| < \frac{d}{۲} \frac{|B|}{n} \Big\} \Big| &< \frac{۴\lambda^۲}{d^۲} \frac{n^۲}{|B|} \\
&= \frac{۴p}{(p+۱)^۲} \frac{n^۲}{|B|} \\
&\leq \frac{۴n^{\frac{۳}{۲}}}{|B|}.
\end{aligned}$$

اگر $|B| > ۱۰\sqrt{n}$، این مقدار از $۰/۴n$ کم‌تر است. به‌عبارت دیگر، اگر $|B| > ۱۰\sqrt{n}$ آنگاه هر مجموعه با $۰/۴n$ خط، حتماً شامل خطی با حداقل $\frac{|B|}{۲\sqrt{n}} = \frac{\sqrt{n}}{۲} \frac{|B|}{n}$ عضو از B خواهد بود.

نتیجه ۲.۶.۳. [۲] *هر مجموعه‌ی S با حداقل $۰/۴n$ خط، دارای زیرمجموعه‌ای با حداکثر $\sqrt{n} \ln n$ خط است که همه‌ی نقاط به جز حداکثر $۱۰\sqrt{n}$ نقطه را می‌پوشاند.*

برهان. ابتدا قرار می‌دهیم $B = V$. تا زمانی‌که $|B| > ۱۰\sqrt{n}$، می‌توان خطی در S شامل حداقل $\frac{|B|}{۲\sqrt{n}}$ نقطه از B انتخاب و سپس این نقاط را از B خارج کرد. اندازه‌ی B بعد از مرحله‌ی i-ام را با B_i نشان می‌دهیم و $B_۰ = B$. فرض کنید پس از تکرار k مرحله داریم:

$$|B_k| \leq ۱۰\sqrt{n}.$$

از طرفی طبق رابطه‌ی بازگشتی داریم:

$$\begin{aligned} |B_k| &\leq |B_{k-۱}| - \frac{|B_{k-۱}|}{۲\sqrt{n}} \\ &= |B_{k-۱}|(۱ - \frac{۱}{۲\sqrt{n}}) \\ &\leq |B_۰|(۱ - \frac{۱}{۲\sqrt{n}})^k \\ &= n(۱ - \frac{۱}{۲\sqrt{n}})^k. \end{aligned}$$

بنابراین

$$k \leq \frac{\ln \frac{۱۰\sqrt{n}}{n}}{\ln(۱ - \frac{۱}{۲\sqrt{n}})}.$$

طبق رابطه‌ی

$$(۱ - x) \leq e^{-x}$$

داریم:

$$k \leq ۲\sqrt{n}\ln\left(\frac{n}{۱۰\sqrt{n}}\right).$$

یعنی، پس از تکرار حداکثر

$$۲\sqrt{n}\ln\left(\frac{n}{۱۰\sqrt{n}}\right) < \sqrt{n}\ln n$$

مرحله، اندازه‌ی B حداکثر برابر $۱۰\sqrt{n}$ خواهد بود. □

برای تکمیل اثبات لم رنگ‌آمیزی، کافی است مجموعه‌ی S را که در نتیجه‌ی قبل مطرح شد، مجموعه‌ی خطوطی درنظر بگیریم که از P حذف شده‌اند. در این صورت کلاس‌های رنگی، $\sqrt{n}\ln n$ خطی هستند

که وجود آن‌ها در این نتیجه اثبات شد و همچنین $۱۰\sqrt{n}$ نقطه‌ای که با این خطوط پوشانده نشده‌اند. اگر نقطه‌ای به بیشتر از یکی از این خطوط متعلق بود، آنگاه کلاس رنگی آن را به دلخواه از میان یکی از این خطوط انتخاب می‌کنیم. به‌وضوح این رنگ‌آمیزی یک رنگ‌آمیزی مجاز را برای G با کلاس‌های رنگی، از اندازه‌ی کم‌تر از $\sqrt{n}$، نتیجه می‌دهد. □

لم ۴.۶.۲. [۲]**(اولین دوبخشی)** *فرض کنید M تطابقی ناتهی روی V و V اجتماع دو مجموعه‌ی مجزای L و R باشد به‌طوری‌که $|R| \geq ۲$. همچنین فرض کنید می‌دانیم که L و R شامل هیچ یالی از M نیستند. می‌توان M را با یک الگوریتم k- مرحله‌ای جستجو کرد که در هر مرحله حداکثر $|M||L|^{\frac{۱}{k}} \log |R|$ پرسش طرح می‌کند.*

برهان. با استقرا روی k ثابت می‌کنیم. اگر $k = ۱$، آنگاه می‌توان یک جستجوی دودویی هم‌زمان [۶] را برای هر عضو L انجام داد تا جفت آن، در صورت وجود، در R پیدا شود. به این صورت که به رئوس R، $(۰, ۱)$- بردارهای متمایز و ناصفر به طول $r = \lceil \log |R| \rceil$ نسبت می‌دهیم. یعنی

$$y = (y_۱, \ldots, y_r)\ ;\ y_i \in \{۰, ۱\}.$$

برای هر $x \in L$ و هر $۱ \leq i \leq r$ گروه‌های زیر را می‌سازیم

$$\{x\} \cup S_i$$

که در آن

$$S_i = \{y \in R \mid y_i = ۱\}.$$

نتیجه‌ی یک گروه "مثبت" است اگر و تنها اگر جفت x در آن گروه باشد. بنابراین، نتیجه‌ی این گروه‌ها مختصات جفت x را در R تعیین می‌کند.

تعداد کلی گروه‌هایی که به این شکل ساخته می‌شوند برابر است با $|L|\lceil \log |R| \rceil$.

حال فرض کنید $k \geq ۲$ و الگوریتمی $(k - ۱)$- مرحله‌ای با حداکثر

$$|M||L|^{\frac{۱}{(k-۱)}} \log |R|$$

[۶] parallel binary search

پرسش در هر مرحله، موجود است. قرار می‌دهیم $t = |L|^{\frac{1}{k}}$. سپس L را به t قسمت $L_1, \dots, L_t$ از اندازه‌ی $\frac{|L|}{t}$ افراز می‌کنیم. در مرحله‌ی ۱، برای هر i، $L_i \cup R$ را مورد پرسش قرار می‌دهیم. حداکثر $|M|$ تا از این مجموعه‌ها شامل یالی از M خواهند بود، به‌عنوان مثال

$$L_{i_1} \cup R, \dots, L_{i_m} \cup R\ ;\ m \leq |M|.$$

فرض استقرا را برای هریک از این m مجموعه به‌کار می‌بریم. فرض کنید e_j تعداد یال‌ها در $L_{i_j} \cup R$ باشد. تعداد پرسش‌ها در هر مرحله حداکثر برابر است با

$$\sum_j e_j (\frac{|L|}{t})^{\frac{1}{(k-1)}} \log |R| = |M||L|^{\frac{1}{k}} \log |R|.$$

□

لم ۵.۶.۲. [۲]**(دومین دوبخشی)** *فرض کنید M تطابقی ناتهی روی V و V اجتماع دو مجموعه‌ی مجزای L و R باشد، به‌طوری‌که می‌دانیم L و R شامل هیچ یالی از M نیستند. فرض کنید c عددی حقیقی است که $0 < c < 1$ و $c|L|^{\frac{1}{k}} \geq 1$ و $k \geq 2$. می‌توان جستجوی M را با الگوریتمی k-مرحله‌ای انجام داد، که حداکثر $c|L|^{\frac{1}{k}}$ پرسش در مرحله‌ی اول و حداکثر $|M||L|^{\frac{1}{k}} \log |R| \cdot \frac{1}{c^{\frac{1}{k-1}}}$ پرسش در مراحل بعد، مطرح می‌کند.*

برهان. قرار دهید $t = c|L|^{\frac{1}{k}}$. L را به t قسمت $L_1, \dots, L_t$ از اندازه‌ی $\frac{|L|}{t}$ افراز می‌کنیم. در مرحله‌ی اول، برای هر i، $L_i \cup R$ را مورد پرسش قرار می‌دهیم. حداکثر $|M|$ تا از این مجموعه‌ها شامل یالی از M خواهند بود، برای مثال

$$L_{i_1} \cup R, \dots, L_{i_m} \cup R\ ;\ m \leq |M|$$

لم اولین دوبخشی را روی این m مجموعه به‌کار می‌بریم. فرض کنید e_j تعداد یال‌ها در $L_{i_j} \cup R$ باشد. تعداد پرسش‌های انجام شده در مرحله‌ی اول برابر است با $t = c|L|^{\frac{1}{k}}$ و در مراحل بعد حداکثر برابر است با

$$\sum_j e_j (\frac{|L|}{t})^{\frac{1}{(k-1)}} \log |R| = |M||L|^{\frac{1}{k}} \log |R| \cdot \frac{1}{c^{\frac{1}{(k-1)}}}.$$

□

قضیه ۶.۶.۲. [۲] *به ازای* $3 \leq k \leq \log n$*، الگوریتم قطعی* k*- مرحله‌ای برای مسأله‌ی تطابق وجود دارد که در هر مرحله تعداد* $O\Big(n^{1+\frac{1}{2(k-1)}}(\log n)^{1+\frac{1}{(k-1)}}\Big)$ *پرسش، مطرح می‌کند.*

برهان. با اضافه کردن $o(n)$ رأس افزوده، می‌توان فرض کرد $n = p^2 + p + 1$ که p عددی اول است. این رأس‌های افزوده با انجام پرسش‌ها حذف خواهند شد. در مرحله‌ی اول، یک صفحه تصویری با n نقطه می‌سازیم و هر خط را به‌عنوان یک گروه درنظر می‌گیریم. هر خطی که شامل هیچ یالی از تطابق نیست را حذف می‌کنیم. در نتیجه حداقل $\frac{n}{2}$ خط حذف خواهد شد. G و کلاس‌های رنگی آن را همان‌طور که در لم رنگ‌آمیزی گفته شد می‌سازیم. اگر $(x, y) \in M$ آنگاه x و y در دو کلاس رنگی متمایز از G قرار خواهند داشت. برای هر جفت از کلاس‌های رنگی، لم دومین دوبخشی را به ازای

$$c = \frac{(\log n)^{\frac{1}{(k-1)}}}{\log n}$$

و تعداد مراحل برابر با $(k-1)$ به‌کار می‌بریم. تعداد پرسش‌ها در مرحله‌ی دوم حداکثر برابر است با

$$O\Big(\binom{\sqrt{n}\log n}{2} c(\sqrt{n})^{\frac{1}{(k-1)}}\Big) = O\Big(n^{1+\frac{1}{2(k-1)}}(\log n)^{1+\frac{1}{(k-1)}}\Big).$$

تعداد پرسش‌ها در هر یک از مراحل ۳ تا k حداکثر برابر است با

$$O\Big((\frac{n}{2})(\sqrt{n})^{\frac{1}{(k-1)}}\log\sqrt{n}\cdot\frac{1}{c^{\frac{1}{(k-2)}}}\Big) = O\Big(n^{1+\frac{1}{2(k-1)}}(\log n)^{1+\frac{1}{(k-1)}}\Big).$$

□

اکنون یک الگوریتم قطعی ۲-مرحله‌ای را شرح می‌دهیم که در آن از فضای تصویری متناهی با بعد ۴ استفاده می‌شود. مزیت این الگوریتم این است که از گروه‌هایی با اندازه‌ی تقریبی $n^{\frac{1}{4}}$ یا کم‌تر استفاده می‌کند.

قضیه ۷.۶.۲. [۲] *الگوریتم قطعی ۲- مرحله‌ای برای مسأله‌ی تطابق وجود دارد که* $\frac{5}{4}n^{\frac{3}{2}}(1+o(1))$ *گروه، از اندازه‌ی حداکثر* $n^{\frac{1}{4}}$*، می‌سازد.*

برهان. قرار دهید $m \approx n^{\frac{1}{4}}$ به‌طوری‌که K_n برابر است با اجتماع مجزای تقریباً $n^{\frac{3}{2}}$ کپی از K_m. (می‌توان از یک فضای تصویری با بعد ۴ که هر خط آن به طول m است، استفاده کرد. یعنی، یک صفحه تصویری

با n نقطه و n خط، که هر خط خود یک صفحه تصویری شامل تقریباً $\sqrt{n}$ خط است، با طول تقریبی $n^{\frac{1}{4}}$.)

مرحله ۱: هر کپی از K_m را به‌عنوان یک گروه در نظر می‌گیریم. در این مرحله تقریباً

$$n\sqrt{n} = n^{\frac{3}{2}}$$

گروه ساخته می‌شود که حداکثر $\frac{n}{2}$ این گروه‌ها شامل یالی از تطابق هستند.

مرحله ۲: هریک از یال‌های باقی‌مانده را مورد پرسش قرار می‌دهیم. در این مرحله تعداد گروه‌ها حداکثر برابر است با

$$\frac{n}{2}\binom{n^{\frac{1}{4}}}{2} \approx \frac{1}{4}n^{\frac{3}{2}}.$$

بنابراین در مجموع

$$n^{\frac{3}{2}} + \frac{1}{4}n^{\frac{3}{2}} = \frac{5}{4}n^{\frac{3}{2}}$$

گروه از اندازه‌ی حداکثر $n^{\frac{1}{4}}$ داریم. □

فصل ۳

جستجو در خانواده‌ای خاص

۱.۳ مقدمه

در این فصل، روش‌هایی که در فصل گذشته شرح داده شد، تعمیم داده می‌شود. آلن و همکارانش در [۲] مسأله‌ی تطابق را به مسأله‌ی جستجوی عضوی مخفی از یک خانواده‌ی داده شده از گراف‌ها، تعمیم دادند و مدل پرسش ۲.۱ را مطرح کردند. آن‌ها در این مقاله، علاوه بر خانواده‌ی تطابق‌ها، تنها یک خانواده‌ی خاص از گراف‌ها را بررسی کردند و یک کران پایین برای تعداد پرسش‌های مورد نیاز برای حل این مسأله ارائه دادند. آن‌ها در محاسبات از نامساوی مارکوف استفاده کردند.

گزاره ۱.۱.۳. [۱۲]**(*نامساوی مارکوف*)** *فرض کنید X یک متغیر تصادفی نامنفی و λ یک عدد حقیقی باشد در این صورت*

$$Pr(X \geq \lambda) \leq \frac{E(X)}{\lambda}.$$

این نتایج کران پایین $\Omega(n^{2})$ را برای مسأله‌ی جستجوی تطابق کامل یا دور همیلتونی یا به‌طور کلی، مسأله‌ی جستجوی یک کپی مخفی از هر گراف با درجه‌ی کران‌دار و ثابت با $\Omega(n)$ یال، نتیجه می‌دهد. همچنین می‌توان کران پایین $\Omega(n^{2})$ را برای مسأله‌ی جستجوی یک کپی مخفی از اجتماع رأس- مجزای یک خوشه از اندازه‌ی $n-3$، یک یال و یک رأس منفرد به‌دست آورد. اما این کران برای مسأله‌ی جستجوی یک کپی مخفی از اجتماع رأس- مجزای یک خوشه از اندازه‌ی $n-2$ و یک یال مناسب نیست. در حقیقت به‌سادگی می‌توان دید که $O(n)$ پرسش برای مسأله‌ی آخر کافی است. در ادامه، خانواده‌ای از گراف‌ها، شامل

همه‌ی گراف‌های روی V و یکریخت با گراف G، در نظر گرفته شده است. آلن و ازُدی در [۱] کران پایینی برای این مسأله ارائه دادند که این کران به عدد استقلال گراف G وابسته است. سپس با استفاده از این کران، کران پایین $\Omega(\frac{n^۲}{\log^۲ n})$ را برای گراف تصادفی $G(n, \frac{۱}{۲})$، به‌دست آوردند.

تعریف ۳.۱.۲. [۳] فرض کنید n عدد صحیح مثبت است و $۰ \leq p \leq ۱$. **گراف تصادفی** $G(n,p)$، عبارت است از فضای احتمال روی مجموعه‌ی گراف‌هایی با مجموعه رئوس $\{۱, \ldots, n\}$، به‌طوری‌که

$$Pr(\{i,j\} \in E(G)) = p.$$

۲.۳ خانواده‌ی p-تُنُک

تعریف ۳.۲.۱. [۲] خانواده‌ی $\mathcal{F}$ از زیرمجموعه‌های V، **p-تنک** [۱] نامیده می‌شود هرگاه گردایه‌ای از حداکثر pn دوتایی دوبه‌دو مجزا از اعضای V موجود باشد به طوری که هر $F \in \mathcal{F}$، شامل حداقل یکی از این دوتایی‌ها باشد. به عبارت دیگر $\mathcal{F}$، p-تنک است اگر و تنها اگر تطابقی با حداکثر pn یال روی V موجود باشد به‌طوری‌که برای هر $F \in \mathcal{F}$، پاسخ پرسش Q_F "مثبت" باشد.

لم ۳.۲.۲. [۲] *عدد ثابت $c_۱ > ۰$ موجود است به‌طوری‌که هر خانواده‌ی $\mathcal{F}$ از زیرمجموعه‌های V با وزن حداکثر $c_۱ p^۲$، p-تنک است.*

برهان. زیرمجموعه‌ی تصادفی $V_۱$ از V را به‌صورت زیر می‌سازیم. هر عضو $v \in V$، به‌طور تصادفی و مستقل، با احتمال p در $V_۱$ قرار می‌گیرد. داریم:

$$E(|V_۱|) = pn.$$

در نتیجه طبق نامساوی مارکوف (۱.۱.۳) داریم:

$$Pr(|V_۱| > ۲pn) \leq \frac{۱}{۲},$$

بنابراین

$$Pr(|V_۱| \leq ۲pn) \geq \frac{۱}{۲}. \quad (۱.۳)$$

[۱] p-sparse

برای هر $F \in \mathcal{F}$، میانگین اندازه‌ی $F \cap V_1$ برابر است با $p|F|$، چون این اندازه نیز یک متغیر تصادفی دوجمله‌ای است، طبق تخمین استاندارد برای توزیع‌های دوجمله‌ای (۱.۱.۲)، احتمال این‌که $|F \cap V_1|$ کم‌تر از $\frac{p|F|}{2}$ باشد، از $e^{-\frac{p|F|}{8}}$، کم‌تر است. یعنی

$$Pr(|F \cap V_1| < \frac{p|F|}{2}) < e^{-\frac{p|F|}{8}}.$$

چون طبق فرض، وزن $\mathcal{F}$ حداکثر $c_1 p^2$ است (به ازای مقدار کوچک c_1)، نتیجه می‌گیریم هر مجموعه در $\mathcal{F}$ حداقل از اندازه‌ی، مثلاً، $\frac{100}{p}$ است. همچنین برای هر $x > \frac{100}{p}$ داریم:

$$e^{-px/8} < \frac{1}{p^2\binom{x+2}{2}}.$$

بنابراین احتمال این‌که مجموعه‌ی $F \in \mathcal{F}$ وجود داشته باشد به‌طوری‌که $|F \cap V_1| < \frac{p|F|}{2}$، حداکثر برابر است با

$$\begin{aligned}
\sum_{F \in \mathcal{F}} Pr(|F \cap V_1| < \frac{p|F|}{2}) &< \sum_{F \in \mathcal{F}} e^{-\frac{p|F|}{8}} \\
&< \frac{1}{p^2} \sum_{F \in \mathcal{F}} \frac{1}{\binom{|F|+2}{2}} \\
&= \frac{w(\mathcal{F})}{p^2} \\
&\leq c_1 \\
&< \frac{1}{2}.
\end{aligned}$$

در نتیجه

$$Pr(\forall F \in \mathcal{F}\ ;\ |F \cap V_1| \geq \frac{p|F|}{2}) > \frac{1}{2}. \tag{۲.۳}$$

بنابراین طبق رابطه‌های (۱.۳) و (۲.۳)، مجموعه‌ی $V_1 \subset V$ از اندازه‌ی حداکثر $2pn$ وجود دارد به‌طوری‌که برای هر $F \in \mathcal{F}$ داریم $|F \cap V_1| \geq \frac{p|F|}{2}$.

خانواده‌ی $\mathcal{F}_1$ را به‌صورت زیر تعریف می‌کنیم:

$$\mathcal{F}_1 = \{F \cap V_1 \mid F \in \mathcal{F}\}.$$

چون $w(\mathcal{F}) \leq c_1p^2$، داریم:

$$\begin{aligned} w(\mathcal{F}_1) &= \sum_{F\in\mathcal{F}} \frac{1}{\binom{|F\cap V_1|+2}{2}} \\ &\leq \sum_{F\in\mathcal{F}} \frac{1}{\binom{(\frac{p|F|}{2})+2}{2}} \\ &\leq \frac{8}{p^2}w(\mathcal{F}) \\ &\leq 8c_1 \\ &< \frac{49}{153} \end{aligned}$$

(با این فرض که c_1 به اندازه‌ی کافی کوچک باشد). طبق لم ۳.۲.۲، $\mathcal{F}_1$ تنک است و چون $|V_1| \leq 2pn$، پس $\mathcal{F}$، p-تنک است. □

لم ۳.۲.۳. [۲] *فرض کنید $\mathcal{H}$ خانواده‌ای از گراف‌ها روی $V = \{1, 2, \ldots, n\}$ است که نسبت به یکریختی بسته است. فرض کنید دو گراف متمایز $H_1, H_2 \in \mathcal{H}$ و زیرمجموعه‌ی $D \subset V$ موجود هستند به‌طوری‌که $|D| = d$ و:*

i *گراف‌های به‌دست آمده از H_1 و H_2، با حذف همه‌ی یال‌هایی که شامل دو رأس از D هستند، با هم برابرند.*

ii *تطابقی با حداقل pn یال در H_1 موجود است به‌طوری‌که شامل هیچ رأسی از D نیست (به‌وضوح این تطابق، یک تطابق در H_2 نیز خواهد بود).*

همچنین، فرض کنید $d > \frac{1}{p}$ و $\mathcal{F}$ خانواده‌ای از زیرمجموعه‌های V باشد که $\mathcal{H}$-مسأله را حل می‌کند. پس برای هر زیرمجموعه‌ی $D' \subset V$ که $|D'| = d$، خانواده‌ی

$$\mathcal{F}_{D'} = \{F - D' \mid F \in \mathcal{F}, |F \cap D'| \geq 2\}$$

p-تنک نیست.

برهان. فرض کنید $D' \subset V$ وجود دارد به‌طوری‌که $|D'| = d$ و $\mathcal{F}_{D'}$، p-تنک است. در نتیجه تطابق M از اندازه‌ی حداکثر pn روی $V \setminus D'$، موجود است به‌طوری‌که هر عضو $\mathcal{F}_{D'}$ شامل یالی از M می‌شود. طبق ii تطابق M می‌تواند به گرافی توسیع یابد که هیچ یالی در D' ندارد و با گراف به‌دست آمده از $H_۱$ (یا $H_۲$) با حذف یال‌های درون D، یکریخت است، این یکریختی، D را به روی D' می‌نگارد. اکنون می‌توان تنها با اضافه کردن یال‌هایی درون D'، این گراف را به یک کپی از $H_۱$، گسترش داد. گراف حاصل را $H'_۱$ می‌نامیم. به همین ترتیب، گراف $H'_۲$ ساخته می‌شود که یک کپی از $H_۲$ است. از آنجا که $H_۱$ و $H_۲$ دو عضو متمایز در $\mathcal{H}$ هستند، این دو کپی نیز متمایزند و چون $\mathcal{H}$ نسبت به یکریختی بسته است، این دو کپی در $\mathcal{H}$ قرار دارند. اما در این صورت برای هر $F \in \mathcal{F}$، پاسخ پرسش Q_F برای این دو عضو متمایز از $\mathcal{H}$ یکسان است و با این فرض که $\mathcal{F}$، $\mathcal{H}$-مسأله را حل می‌کند تناقض دارد. □

اکنون به اثبات لم اصلی این بخش می‌پردازیم. در این لم، یک کران پایین برای اندازه‌ی خانواده‌ای از گراف‌ها که در شرایط لم ۳.۲.۳ صدق می‌کند بیان می‌شود. البته، در اینجا، هیچ تلاشی برای بهبود مقادیر ثابت این تخمین انجام نشده است.

لم ۴.۲.۳. [۲] *فرض کنید $\mathcal{H}$، V، n، d و p مفروضات موجود در لم ۳.۲.۳ باشند و $c > ۰$ عددی ثابت باشد. برای هر خانواده‌ی $\mathcal{F}$، که $\mathcal{H}$-مسأله را حل می‌کند داریم:*

$$|\mathcal{F}| \geq c \cdot \frac{p^۲}{d^۲} \cdot \binom{n}{۲}.$$

برهان. فرض کنید $\mathcal{F}$ خانواده‌ای از زیرمجموعه‌های V باشد که $\mathcal{H}$-مسأله را حل می‌کند. طبق لم ۱.۲.۳ و لم ۳.۲.۳، برای هر مجموعه‌ی D شامل d رأس از V، وزن خانواده‌ی

$$\mathcal{F}_D = \{F - D \mid F \in \mathcal{F}, |F \cap D| \geq ۲\}$$

حداقل $c_۱ p^۲$ است. ادعا می‌کنیم برای هر D به صورت بالا و با انتخاب مناسب $c_۲ > ۰$ داریم:

$$\sum_{F \in \mathcal{F}, |F \cap D| \geq ۲} \frac{۱}{\binom{|F|}{۲}} \geq c_۲ p^۲. \quad (۳.۳)$$

اگر $F \in \mathcal{F}$ از اندازه‌ی حداکثر، مثلاً، $\frac{۱۰}{p}$ و $F \cap D$ حداقل دو عضو داشته باشد، حکم فوراً نتیجه می‌شود.

درغیر این صورت، طبق فرض $d < \frac{۱}{p}$،

$$\frac{۱}{\binom{|F|-d+۲}{۲}} \leq ۲\frac{۱}{\binom{|F|}{۲}},$$

داریم:

$$\begin{aligned}\sum_{F\in\mathcal{F},|F\cap D|\geq ۲} \frac{۱}{\binom{|F|}{۲}} &\geq \frac{۱}{۲} \sum_{F\in\mathcal{F},|F\cap D|\geq ۲} \frac{۱}{\binom{|F|-d+۲}{۲}}\\ &\geq \frac{۱}{۲}\cdot w(\mathcal{F}_D)\\ &\geq \frac{۱}{۲}\cdot c_۱ p^۲.\end{aligned}$$

اگر D یک زیرمجموعه‌ی تصادفی از V با d رأس باشد، برای هر $F \in \mathcal{F}$ داریم:

$$\begin{aligned}Pr(|D\cap F| \geq ۲) &\leq \sum_{a,b\in F} Pr(a,b\in D)\\ &= \binom{|F|}{۲}\cdot \frac{\binom{n-۲}{d-۲}}{\binom{n}{d}}\\ &= \binom{|F|}{۲}\cdot \frac{\binom{d}{۲}}{\binom{n}{۲}}.\end{aligned}$$

متغیر تصادفی X_D را به‌صورت زیر تعریف می‌کنیم:

$$X_D = \sum_{F\in\mathcal{F},|F\cap D|\geq ۲} \frac{۱}{\binom{|F|}{۲}}.$$

به دلیل خطی بودن میانگین داریم:

$$\begin{aligned}E(X_D) &= \sum_{F\in\mathcal{F},|F\cap D|\geq ۲} E\left(\frac{۱}{\binom{|F|}{۲}}\right)\\ &\leq \sum_{F\in\mathcal{F},|F\cap D|\geq ۲} \frac{۱}{\binom{|F|}{۲}}\cdot\binom{|F|}{۲}\cdot\frac{\binom{d}{۲}}{\binom{n}{۲}}\\ &\leq \frac{\binom{d}{۲}}{\binom{n}{۲}}\cdot|\mathcal{F}|.\end{aligned}$$

از طرفی طبق (۳.۳) می‌دانیم این متغیر تصادفی حداقل $c_۲p^۲$ است. بنابراین

$$|\mathcal{F}| \geq c\cdot\frac{p^۲}{d^۲}\cdot\binom{n}{۲}.$$

□

۳.۳ خانواده‌ی $\mathcal{H}_G$

در این بخش خانواده‌ای از گراف‌ها، شامل همه‌ی گراف‌های روی V و یکریخت با گراف G، در نظر گرفته شده است. این خانواده را $\mathcal{H}_G$ نامیده و کران پایینی برای اندازه‌ی خانواده‌ای که این مسأله را حل می‌کند، ارائه می‌شود. این کران به عدد استقلال گراف G وابسته است. سپس با استفاده از این کران، کران پایین $\Omega(\frac{n^2}{\log^2 n})$ برای گراف تصادفی $G(n, \frac{1}{2})$، به‌دست می‌آید.

قضیه ۱.۳.۳. [۱] *فرض کنید $G = (V, E)$ گرافی روی n رأس است. همچنین فرض کنید سه رأس $u, v, w \in V$ وجود دارد به‌طوری‌که برای هر دو تا از آن‌ها، مجموعه‌های همسایه‌های آن‌ها به جز خود آن رئوس، متمایزند. یعنی،*

$$N(u) \setminus \{v\} \neq N(v) \setminus \{u\}$$

و

$$N(u) \setminus \{w\} \neq N(w) \setminus \{u\}$$

و

$$N(v) \setminus \{w\} \neq N(w) \setminus \{v\}.$$

در این صورت، اندازه‌ی هر خانواده که $\mathcal{H}_G$- مسأله را حل می‌کند حداقل برابر $\Omega(\frac{n^2}{\alpha^2(G)})$ است.

برهان. برای هر دو رأس $x, y \in V$، تعریف می‌کنیم $A(x, y)$ عبارت است از مجموعه‌ی رئوس $z \in V \setminus \{x, y\}$ به‌طوری‌که z یا همسایه‌ی هر دو رأس x و y است یا همسایه‌ی هیچ کدام از آن‌ها نیست. نشان می‌دهیم که دو رأس در میان u، v و w هست که اندازه‌ی این مجموعه حداقل برابر $\frac{1}{3}n - 1$ است. فرض کنید $|A(u, v)| < \frac{1}{3}n - 1$. آنگاه

$$|V \setminus (A(u, v) \cup \{u, v, w\})| > \frac{2}{3}n - 2$$

و هریک از رئوس $V \setminus (A(u, v) \cup \{u, v, w\})$ همسایه‌ی دقیقاً یکی از دو رأس u و v است. بنابراین، هریک از این رئوس در $A(u, w)$ یا $A(v, w)$ است و در نتیجه حداقل یکی از این مجموعه‌ها از اندازه‌ی

حداقل $\frac{1}{۳}n - ۱$ است. بدون از دست دادن کلیت، فرض کنید $|A(u,v)| \geq \frac{1}{۳}n - ۱$.

فرض کنید $\mathcal{F}$ خانواده‌ای باشد که $\mathcal{H}_G$-مسأله را حل می‌کند و $|\mathcal{F}| < \frac{n^۲}{۱۲\alpha^۲}$، که $\alpha = \alpha(G)$. هر مجموعه‌ی $F \in \mathcal{F}$ از اندازه‌ی حداکثر α است، چون در غیر این صورت پاسخ پرسش Q_F "مثبت" است و این نتیجه بدون پرسش نیز مشخص است. برای هر $x \in V$، تعداد مجموعه‌های $F \in \mathcal{F}$ که $x \in F$ را با $f(x)$ مشخص می‌کنیم.

$$\sum_{x \in V} f(x) = \sum_{F \in \mathcal{F}} |F| \leq \alpha |\mathcal{F}| < \frac{n^۲}{۱۲\alpha}. \tag{۴.۳}$$

قرار می‌دهیم

$$V' = \Big\{ x \in V \mid f(x) < \frac{n}{۶\alpha} \Big\}.$$

$|V'| \geq \frac{n}{۲}$، چون در غیر این صورت

$$\sum_{x \in V} f(x) \geq \sum_{x \in (V \setminus V')} f(x) > \frac{n}{۲} \cdot \frac{n}{۶\alpha} = \frac{n^۲}{۱۲\alpha}$$

که با (۴.۳) تناقض دارد. برای $x \in V'$، تعداد رئوس $z \in V$ به‌طوری‌که مجموعه‌ای در $\mathcal{F}$ شامل هر دو رأس x و z وجود داشته باشد، حداکثر برابر است با

$$\sum_{F: x \in F} |F| \leq f(x)\alpha < \frac{n}{۶}.$$

دو رأس دلخواه $x, y \in V'$ را در نظر می‌گیریم و مجموعه‌ی همه‌ی رئوس $z \in V$ به‌طوری‌که یک مجموعه‌ی $F \in \mathcal{F}$ شامل x و z یا y و z وجود داشته باشد را A می‌نامیم.

$$|A| \leq \sum_{F: x \in F} |F| + \sum_{F: y \in F} |F| < \frac{n}{۳}.$$

گراف یکریخت با G را که u به x، v به y، و فقط رئوسی از $A(u,v)$ به A نگاشته می‌شود، با $G_۱$ نمایش می‌دهیم. فرض کنید $G_۲$ گرافی باشد که u به y و v به x نگاشته می‌شود و باقی‌مانده‌ی آن همانند $G_۱$ است. تنها پرسش‌هایی که می‌توانند بین $G_۱$ و $G_۲$ تمایز قائل شوند پرسش‌های Q_F است که F شامل x یا y است، اما همه‌ی رأس‌های دیگر F در $A(u,v)$ است. بنابراین، پاسخ Q_F برای $G_۱$ و $G_۲$ یکسان است. در نتیجه $\mathcal{F}$ نمی‌تواند بین $G_۱$ و $G_۲$ تمایز قائل شود. پس با این فرض که این خانواده $\mathcal{H}_G$-مسأله را حل می‌کند تناقض دارد. $\square$

نتیجه ۳.۳.۲. *فرض کنید $G = G(n, \frac{1}{2})$ گراف تصادفی روی n رأس باشد. با احتمال بالا، می‌توان گفت هر خانواده که $\mathcal{H}_G$- مسأله را حل کند از اندازه‌ی حداقل $\Omega(\frac{n^2}{\log^2 n})$ است.*

برهان. این نتیجه از قضیه‌ی ۳.۳.۱ به‌دست می‌آید. چون، با احتمال بالا، $\alpha(G) = O(\log n)$ و با احتمال بالا، سه رأس u، v و w با مجموعه‌هایی متمایز از همسایه‌ها، همان‌طور که در قضیه تعریف شد، وجود دارد. □

فصل ۴
جستجوی ستاره

۱.۴ مقدمه

گرافی که در این فصل جستجو می‌شود گرافی است روی $V = \{1, \ldots, n\}$ که یک مؤلفه‌ی همبندی آن ستاره و دیگر مؤلفه‌های آن رئوس منفرد است. این مسأله، هم در حالتی که اندازه‌ی ستاره نامعلوم است و هم در حالتی که اندازه‌ی آن معلوم است بررسی می‌شود که در اینجا، منظور از اندازه‌ی ستاره تعداد یال‌های آن است.

خانواده‌ی همه‌ی گراف‌های روی $V = \{1, 2, \ldots, n\}$ که شامل یک کپی از $K_{1,k}$ و $n - k - 1$ رأس منفرد است را با $\mathcal{S}_k$ نشان داده و قرار می‌دهیم

$$\mathcal{S} = \cup_{k=1}^{n-1} \mathcal{S}_k.$$

سپس به ازای مقادیر مختلف k، کران‌های بالا و کران‌های پایینی برای کم‌ترین تعداد پرسش مورد نیاز برای جستجو در این خانواده‌ها، ارائه می‌شود.

۲.۴ جستجوی ستاره‌ای با اندازه‌ی نامعلوم

ابتدا حالتی را در نظر می‌گیریم که از اندازه‌ی ستاره اطلاعی نداریم. بنابراین، جستجو را در خانواده‌ی $\mathcal{S}$ انجام می‌دهیم.

قضیه ۱.۲.۴. [۱] *کوچک‌ترین اندازه‌ی خانواده‌ی $\mathcal{F}$ که $\mathcal{S}$- مسأله را حل می‌کند دقیقاً برابر است با* $\binom{n}{2}$.

برهان. نشان می‌دهیم هر خانواده‌ی $\mathcal{F}$ که $\mathcal{S}$- مسأله را حل می‌کند باید شامل تمامی جفت رأس‌ها باشد. دو رأس متمایز u و v را در V در نظر بگیرید. فرض کنید $S_۱$ ستاره‌ای است به مرکز u و برگ‌های آن همه‌ی رئوس دیگر در V هستند و $S_۲$ ستاره‌ای است به مرکز u و برگ‌های آن همه‌ی رئوس دیگر V به جز v هستند. بدیهی است پاسخ پرسش Q_F که F شامل u نیست هم برای $S_۱$ و هم برای $S_۲$، "منفی" است. همچنین پاسخ پرسش Q_F که F شامل u و رأس دیگری جز v است برای هر دو ستاره "مثبت" است. بنابراین $\mathcal{F}$ باید شامل گروه $\{u,v\}$ باشد، چون در غیر این صورت $\mathcal{F}$ نمی‌تواند میان $S_۱$ و $S_۲$ تمایز قائل شود. $\square$

اثبات بالا دقیقاً نشان می‌دهد که خانواده‌ای که $\mathcal{S}_{n-۱} \cup \mathcal{S}_{n-۲}$- مسأله را حل می‌کند نیز به $\binom{n}{۲}$ پرسش نیاز دارد.

۳.۴ جستجوی ستاره‌ای با اندازه‌ی معلوم

در این بخش حالتی که اندازه‌ی ستاره معلوم است در نظر گرفته شده و قضیه‌ی زیر اثبات می‌شود. این قضیه کران‌های بالا و کران‌های پایینی برای کوچک‌ترین اندازه‌ی خانواده‌ای که $\mathcal{S}_k$- مسأله را حل می‌کند، ارائه می‌دهد.

قضیه ۱.۳.۴. [۱] *برای هر $n > ۲$ و هر $k \le n - ۲$، خانواده‌ای از اندازه‌ی*

$$\min\Big(\Big\lceil\frac{n(n-۲)}{۲}\Big\rceil, O(k^۳ \log n)\Big)$$

وجود دارد که $\mathcal{S}_k$- مسأله را حل می‌کند. همچنین هر خانواده‌ای که $\mathcal{S}_k$- مسأله را حل می‌کند یا شامل $(۱-o(۱))\binom{n}{۲}$ جفت است یا از اندازه‌ی حداقل $\Omega(\frac{k^۳}{\log^۲ n})$ است. علاوه بر این، اگر $k \le \sqrt{n}$، آنگاه اندازه‌ی هر خانواده‌ای که $\mathcal{S}_k$- مسأله را حل می‌کند حداقل برابر است با $\Omega(\frac{k^۳ \log n}{\log k})$. به ازای $k = n - ۱$، کوچک‌ترین اندازه‌ی چنین خانواده‌ای دقیقاً برابر است با $\lceil \log n \rceil$.

بهترین کران‌های به‌دست آمده، برای مقادیر مختلف k، در جدول ۱.۴ خلاصه شده است. در ادامه، اثبات این نتایج بررسی می‌شود.

جدول ۱.۴: کران‌هایی برای اندازه‌ی خانواده‌ای که $\mathcal{S}_k$- مسأله را حل می‌کند.

کران بالا	کران پایین	k
$O(k^{۳}\log n)$	$\Omega(\frac{k^{۳}\log n}{\log k})$	$k \le \sqrt{n}$
$O(k^{۳}\log n)$	$\Omega(\frac{k^{۳}}{\log^{۲} n})$	$\sqrt{n} < k < \frac{n^{\frac{۲}{۳}}}{(۲\log n)^{\frac{۱}{۳}}}$
$\lceil\frac{n(n-۲)}{۲}\rceil$	$\Omega(\frac{k^{۳}}{\log^{۲} n})$	$\frac{n^{\frac{۲}{۳}}}{(۲\log n)^{\frac{۱}{۳}}} \le k \le O(n^{\frac{۲}{۳}}\log^{\frac{۲}{۳}} n)$
$\lceil\frac{n(n-۲)}{۲}\rceil$	$(۱-o(۱))\binom{n}{۲}$	$k=\omega(n^{\frac{۲}{۳}}\log^{\frac{۲}{۳}} n),\ k \le n-۳$
$\lceil\frac{n(n-۲)}{۲}\rceil$	$\lceil\frac{n(n-۲)}{۲}\rceil$	$k=n-۲$
$\lceil\log n\rceil$	$\lceil\log n\rceil$	$k=n-۱$

گزاره ۲.۳.۴. [۱] *برای هر $n > ۲$، کوچک‌ترین اندازه‌ی خانواده‌ای که $\mathcal{S}_{n-۱}$- مسأله را حل می‌کند، دقیقاً برابر است با $\lceil\log n\rceil$.*

برهان. اندازه‌ی خانواده‌ی $\mathcal{S}_{n-۱}$ برابر است با n. بنابراین، طبق کران پایین اطلاعات، حداقل $\lceil\log n\rceil$ پرسش مورد نیاز است. برای این‌که ثابت کنیم $\lceil\log n\rceil$ پرسش کافی است، خانواده‌ی زیر را که $\mathcal{S}_{n-۱}$- مسأله را حل می‌کند می‌سازیم. دقت شود که کافی است تنها رأس ستاره را شناسایی کنیم.

$(۰,۱)$- بردارهای متمایزی به طول $\lceil\log n\rceil$، به رئوس V نسبت می‌دهیم. برای هر $۱ \le i \le \lceil\log n\rceil$، قرار می‌دهیم F_i مجموعه‌ی همه رئوسی باشد که i- امین مؤلفه‌ی آن ۱ است. همچنین تعریف می‌کنیم:

$$\mathcal{F} = \{F_i \mid ۱ \le i \le \lceil\log n\rceil\}.$$

پاسخ پرسش Q_{F_i} "مثبت" است اگر و تنها اگر F_i شامل مرکز باشد. بنابراین، می‌توان از پاسخ پرسش‌ها، مرکز ستاره را شناسایی کرد. □

گزاره ۳.۳.۴. [۱] *کوچک‌ترین اندازه‌ی خانواده‌ی $\mathcal{F}$ که $\mathcal{S}_{n-۲}$- مسأله را حل می‌کند، به ازای $n=۳$، برابر است با ۲، به ازای $n=۴$، برابر است با ۵ و برای هر $n \ge ۵$، برابر است با $\lceil\frac{n(n-۲)}{۲}\rceil$.*

برهان. فرض کنید $\mathcal{F}$ خانواده‌ای است که $\mathcal{S}_{n-۲}$- مسأله را حل می‌کند. سه رأس متمایز u، v و w را از V

در نظر بگیرید. ستاره‌ای به مرکز u و با رأس منفرد v را $S_۱$ و ستاره‌ای به مرکز u و با رأس منفرد w را $S_۲$ می‌نامیم. تنها مجموعه‌هایی که بین $S_۱$ و $S_۲$ تمایز قائل می‌شوند $\{u,v\}$ و $\{u,w\}$ هستند. بنابراین $\mathcal{F}$ باید شامل حداقل یکی از آن‌ها باشد. لذا $\mathcal{F}$ باید شامل همه‌ی جفت رأس‌ها به جز یک تطابق باشد. در نتیجه

$$|\mathcal{F}| \geq \left\lceil \frac{n(n-۲)}{۲} \right\rceil.$$

همچنین، می‌توان به‌طور مستقیم بررسی کرد که برای $n=۳$، ۲ جفت و برای $n=۴$، ۵ جفت لازم و کافی است.

از طرف دیگر، فرض کنید $n \geq ۵$ و M یک تطابق ماکزیمم روی V باشد. خانواده‌ی همه‌ی دوتایی‌ها از رئوس به جز آن‌هایی که در M قرار دارند را $\mathcal{F}$ می‌نامیم. چون

$$|M| = \lfloor \frac{n}{۲} \rfloor$$

بنابراین

$$|\mathcal{F}| = \left\lceil \frac{n(n-۲)}{۲} \right\rceil.$$

وضعیت همه‌ی یال‌ها به جز آن‌هایی که در M قرار دارند، به‌طور مستقیم از پرسش‌ها به‌دست می‌آید. چون $n \geq ۵$ پس هر ستاره در $S_{n-۲}$ حداقل ۳ برگ دارد که حداکثر یکی از آن‌ها در M قرار دارد. در نتیجه یال‌هایی که به‌طور مستقیم از پرسش‌ها به‌دست آمده‌اند رأس مرکزی را نیز مشخص می‌کنند. در این صورت تنها یالی که ممکن است در گراف باشد ولی شناسایی نشده باشد، یالی است در M که با مرکز مجاور است، اگر چنین یالی در M موجود باشد. با توجه به اندازه‌ی ستاره‌ای که تاکنون به‌دست آمده است می‌توان تصمیم گرفت که آیا این یال متعلق به ستاره هست یا خیر. □

دقت شود کران بالای مذکور برای هر S_k، $۳ \leq k \leq n-۲$، برقرار است.

اکنون برخی کران‌های کلی برای اندازه‌ی کوچک‌ترین خانواده‌ای که S_k-مسأله را حل می‌کند ارائه می‌شود.

گزاره ۴.۳.۴. [۱] *برای هر k، خانواده‌ی $\mathcal{F}$ از اندازه‌ی $O(k^{۳}\log n)$ وجود دارد که $\mathcal{S}_k$- مسأله را حل می‌کند.*

برهان. قرار دهید

$$m = ck^{۳}\log n$$

که c عددی ثابت و مثبت است. فرض کنید $F_{۱}, F_{۲}, \ldots F_m$، m زیرمجموعه‌ی تصادفی از V باشند که به‌طور مستقل، به‌صورت زیر انتخاب شده‌اند. برای هر F_i، هر $v \in V$ به‌طور مستقل و با احتمال $\frac{۱}{k}$ در F_i قرار می‌گیرد. برای هر دو ستاره‌ی $S, S' \in \mathcal{S}_k$ قرار دهید $A_{S,S'}$ این پیشامد باشد که "هیچ یک از F_i ها بین S و S' تمایز قائل نشود". $S_{۱}$ و $S_{۲}$ را دو ستاره از اندازه‌ی k در نظر بگیرید به‌طوری‌که

$$|E(S_{۱}) \setminus E(S_{۲})| = |E(S_{۲}) \setminus E(S_{۱})| = ۱.$$

فرض کنید u مرکز $S_{۱}$ و $S_{۲}$ ، $u_{۱}, \ldots, u_{k-۱}$ دیگر رأس‌های مشترک این دو ستاره، v رأس مکمل $S_{۱}$ و w رأس مکمل $S_{۲}$ باشد. F_i بین $S_{۱}$ و $S_{۲}$ تمایز قائل می‌شود اگر و تنها اگر $u \in F_i$، برای هر $۱ \leq j \leq k-۱$، $u_j \notin F_i$ و دقیقاً یکی از رئوس v و w در F_i باشد. بنابراین احتمال این‌که F_i بین $S_{۱}$ و $S_{۲}$ تمایز قائل شود برابر است با

$$\frac{۲}{k^{۲}}\Big(۱-\frac{۱}{k}\Big)^{k} = \Omega\Big(\frac{۱}{k^{۲}}\Big).$$

هرچقدر دو ستاره در تعداد بیشتری یال متفاوت باشند، این احتمال بیشتر است. به‌عنوان مثال، اگر دو ستاره در r یال متمایز باشند این احتمال از

$$\frac{۲r}{k^{۲}}\Big(۱-\frac{۱}{k}\Big)^{k} = \Omega\Big(\frac{۱}{k^{۲}}\Big)$$

بیشتر است. بنابراین، احتمال این‌که هیچ یک از F_i ها بین $S_{۱}$ و $S_{۲}$ تمایز قائل نشود برابر است با

$$Pr(A_{S_{۱},S_{۲}}) = \Big[۱-\Omega\Big(\frac{۱}{k^{۲}}\Big)\Big]^{m} < n^{-(۲k+۲)}$$

به شرط آن‌که c به اندازه‌ی کافی بزرگ باشد. برای دو ستاره که در تعداد بیشتری یال متفاوت باشند، این احتمال کوچک‌تر است.

از طرفی می‌دانیم

$$|\mathcal{S}_k| = (k+۱)\binom{n}{k+۱} = n\binom{n-۱}{k}.$$

بنابراین، تعداد جفت ستاره‌ها برابر است با

$$\begin{aligned}\binom{|\mathcal{S}_k|}{۲} &= \frac{۱}{۲}n\binom{n-۱}{k}\left(n\binom{n-۱}{k}-۱\right)\\ &< n^۲\binom{n-۱}{k}^۲\\ &< n^۲\cdot n^{۲k}\\ &= n^{۲k+۲}.\end{aligned}$$

بنابراین،

$$Pr(\bigcup_{S,S'\in\mathcal{S}_k} A_{S,S'}) \leq \sum_{S,S'\in\mathcal{S}_k} Pr(A_{S,S'}) < ۱,$$

در نتیجه

$$Pr(\bigcap_{S,S'\in\mathcal{S}_k} \bar{A}_{S,S'}) > ۰.$$

و بنابراین، خانواده‌ی $\mathcal{F} = \{F_۱, F_۲, \dots F_m\}$ وجود دارد که $\mathcal{S}_k$- مسأله را حل می‌کند. $\square$

گزاره ۵.۳.۴. [۱] *برای هر $k \leq n-۲$، اگر $\mathcal{F}$ خانواده‌ای باشد که $\mathcal{S}_k$- مسأله را حل می‌کند آنگاه $\mathcal{F}$ یا شامل $(۱-o(۱))\binom{n}{۲}$ جفت است، یا از اندازه‌ی حداقل $\Omega(\frac{k^۳}{\log^۲ n})$ است.*

برهان. فرض کنید $\mathcal{F}$ خانواده‌ای باشد که $\mathcal{S}_k$- مسأله را حل می‌کند. بنابراین، برای هر $u \in V$ و هر $A, B \subseteq V \setminus \{u\}$ به‌طوری‌که $|A| = ۲$، $|B| = k-۱$ و $A \cap B = \emptyset$، مجموعه‌ی $F \in \mathcal{F}$ وجود دارد که $u \in F$، $|F \cap A| = ۱$ و $F \cap B = \emptyset$. در غیر این صورت، $\mathcal{F}$ نمی‌تواند بین دو ستاره که مرکز آن‌ها u است، در رئوس B اشتراک دارند و رأس مکمل یک ستاره یکی از رئوس A و رأس مکمل ستاره‌ی دیگر، رأس دیگر A است، تمایز قائل شود. خانواده‌ی همه‌ی مجموعه‌های $F \in \mathcal{F}$ از اندازه‌ی ۲ را $\mathcal{F}_۰$ می‌نامیم.

قرار می‌دهیم

$$m = c \cdot \frac{n\log n}{k}$$

و تعریف می‌کنیم:

$$\mathcal{F}_{۱} = \{F \in \mathcal{F} \mid ۲ < |F| \leq m\}$$

و

$$\mathcal{F}_{۲} = \mathcal{F} \setminus (\mathcal{F}_{۰} \cup \mathcal{F}_{۱}).$$

نشان می‌دهیم برای هر ثابت $\epsilon > ۰$، اگر

$$|\mathcal{F}_{۰}| \leq (۱-\epsilon)\tbinom{n}{۲}$$

آنگاه به ازای مقدار ثابت $c_{۱}$ که فقط به c وابسته است، داریم:

$$|\mathcal{F}_{۱} \cup \mathcal{F}_{۲}| > c_{۱}\epsilon^{۳} \cdot \frac{k^{۳}}{\log^{۲} n}.$$

فرض کنید

$$|\mathcal{F}_{۰}| \leq (۱-\epsilon)\tbinom{n}{۲}$$

و

$$|\mathcal{F}_{۱} \cup \mathcal{F}_{۲}| \leq c_{۱}\epsilon^{۳} \cdot \frac{k^{۳}}{\log^{۲} n}.$$

برای هر $u \in V$، V_u را مجموعه‌ی رئوس $v \in V \setminus \{u\}$ در نظر می‌گیریم به‌طوری‌که $\{u, v\} \notin \mathcal{F}_{۰}$. تعریف می‌کنیم:

$$V' = \{u \in V \mid |V_u| \geq \frac{\epsilon}{۲}(n-۱)\}.$$

چون $|\mathcal{F}_{۰}| \leq (۱-\epsilon)\tbinom{n}{۲}$، پس $|V'| \geq \frac{\epsilon}{۲}n$. زیرا در غیر این صورت، چون جفت رئوسی که در $\mathcal{F}_{۰}$ نیستند، دوتایی‌های $\{u, v\}$ هستند که $v \in V_u$ و از آنجایی که $v \in V_u$ اگر و تنها اگر $u \in V_v$، داریم:

$$\begin{aligned}
|\mathcal{F}_{۰}| &= \binom{n}{۲} - \frac{۱}{۲}\sum_{u \in V} |V_u| \\
&= \binom{n}{۲} - \frac{۱}{۲}\Big(\sum_{u \in V'} |V_u| + \sum_{u \in V \setminus V'} |V_u|\Big) \\
&> \binom{n}{۲} - \frac{۱}{۲}\Big[|V'|(n-۱) + |V \setminus V'|\frac{\epsilon}{۲}(n-۱)\Big]
\end{aligned}$$

$$\begin{aligned} &> \binom{n}{۲} - \frac{۱}{۲}\Big[\frac{\epsilon}{۲}n(n-۱) + n\frac{\epsilon}{۲}(n-۱)\Big] \\ &= (۱-\epsilon)\binom{n}{۲}. \end{aligned}$$

که به تناقض منجر می‌شود.

به‌طور یکنواخت، رأس $u \in V'$ و سپس به‌طور یکنواخت زیرمجموعه‌ی $A = \{v, w\} \subseteq V_u$ را انتخاب می‌کنیم. تعریف می‌کنیم:

$$\mathcal{F}'_۱ = \{F \in \mathcal{F}_۱ \mid u \in F, |F \cap A| = ۱\}.$$

برای هر $F \in \mathcal{F}_۱$

$$Pr(F \in \mathcal{F}'_۱) \leq \frac{۲|F|(|F|-۱)(n-|F|)}{\frac{\epsilon}{۲}n\binom{\frac{\epsilon}{۲}(n-۱)}{۲}} \leq \frac{۳۲}{\epsilon^۳} \cdot \frac{|F|^۲}{n^۲}.$$

بنابراین

$$E(|\mathcal{F}'_۱|) \leq \frac{۳۲}{\epsilon^۳} \sum_{F \in \mathcal{F}_۱} \frac{|F|^۲}{n^۲} \leq \frac{۳۲}{\epsilon^۳} \cdot |\mathcal{F}_۱| \cdot \frac{m^۲}{n^۲}.$$

در نتیجه، u و A مناسبی وجود دارد که

$$\begin{aligned} |\mathcal{F}'_۱| &\leq \frac{۳۲}{\epsilon^۳} \cdot |\mathcal{F}_۱| \cdot \frac{m^۲}{n^۲} \\ &\leq ۳۲ c_۱ c^۲ \cdot \frac{k^۳}{\log^۲ n} \cdot \frac{n^۲ \log^۲ n}{k^۲ n^۲} \\ &\leq \frac{k}{۲} - ۱ \end{aligned}$$

به شرط آن‌که $c_۱ c^۲$ به اندازه‌ی کافی کوچک باشد. بنابراین، زیرمجموعه‌ی $B_۱ \subseteq V \setminus (\{u\} \cup A)$ از اندازه‌ی $\frac{k}{۲} - ۱$ وجود دارد که با هر $F \in \mathcal{F}'_۱$ اشتراک دارد. زیرمجموعه‌ی تصادفی $B_۲ \subseteq V$ را از اندازه‌ی $\frac{k}{۲}$ انتخاب می‌کنیم. برای هر $F \in \mathcal{F}_۲$ و به ازای مقدار ثابت $c_۲ = \theta(c)$ داریم:

$$\begin{aligned} Pr(F \cap B_۲ = \emptyset) &= \frac{\binom{n-|F|}{\frac{k}{۲}}}{\binom{n}{\frac{k}{۲}}} \\ &\leq \Big(۱ - \frac{|F|}{n}\Big)^{\frac{k}{۲}} \\ &< e^{-\frac{km}{۲n}} \end{aligned}$$

$$\leq n^{-c_۲}.$$

بنابراین، اگر c به اندازه‌ی کافی بزرگ باشد، با احتمال بالا، $u \notin B_۲$، $A \cap B_۲ = \emptyset$ و برای هر $F \in \mathcal{F}_۲$ داریم $F \cap B_۲ \neq \emptyset$.

قرار می‌دهیم

$$B' = B_۱ \cup B_۲.$$

$B' \subseteq V \setminus (\{u\} \cup A)$ و $|B'| \leq k - ۱$. فرض کنید B یک توسیع دلخواه از B' به یک زیرمجموعه از $V \setminus (\{u\} \cup A)$، از اندازه‌ی $k - ۱$ باشد. دو ستاره‌ی $S_۱$ و $S_۲$ را به‌صورت زیر در نظر می‌گیریم.

u مرکز $S_۱$ و $S_۲$ است، این دو ستاره در رأس‌های B مشترکند، v رأس مکمل $S_۱$ و w رأس مکمل $S_۲$ است. چون A از V_u انتخاب شده بود، دوتایی های $\{u, v\}$ و $\{u, w\}$ در $\mathcal{F}_۰$ نیستند و بنابراین، هیچ مجموعه‌ای در $\mathcal{F}_۰$ نمی‌تواند بین $S_۱$ و $S_۲$ تمایز قائل شود. از بین مجموعه‌های موجود در $\mathcal{F}_۱$ هیچ کدام از آن‌هایی که شامل u نیستند و هیچ کدام از آن‌هایی که اشتراک آن‌ها با A از اندازه‌ی ۱ نیست، نمی‌توانند این دو ستاره را متمایز کنند. اعضای دیگر $\mathcal{F}_۱$، یعنی مجموعه‌های $F \in \mathcal{F}_۱$ به‌طوری‌که $u \in F$ و $|F \cap A| = ۱$ و همه‌ی اعضای $\mathcal{F}_۲$ شامل رأسی از B هستند، لذا هیچ کدام نمی‌توانند این دو ستاره را تشخیص دهند. در نتیجه $\mathcal{F}$ نمی‌تواند بین $S_۱$ و $S_۲$ تمایز قائل شود. پس با این فرض که $\mathcal{F}$، $\mathcal{S}_k$-مسأله را حل می‌کند در تناقض است. □

اکنون کران پایین بهتری برای $k \leq \sqrt{n}$ ارائه می‌کنیم. این کران به عاملی از $\log k$ محدود می‌شود. قبل از اثبات این کران، ابتدا مفهوم r-پوشش-آزاد [۱] را تعریف کرده، سپس اندازه‌ی مجموعه‌ی زمینه‌ی [۲] یک خانواده‌ی r-پوشش-آزاد از اندازه‌ی n را کران‌دار می‌کنیم.

تعریف ۶.۳.۴. [۱] فرض کنید $\mathcal{A}$ خانواده‌ای از زیرمجموعه‌های مجموعه‌ی S باشد. می‌گوییم $\mathcal{A}$، **r-پوشش-آزاد** است اگر هیچ مجموعه‌ای در $\mathcal{A}$ در اجتماع هر r مجموعه‌ی دیگر در $\mathcal{A}$ قرار نگیرد.

[۱] r-cover-free
[۲] ground set

لم ۰.۷.۳.۴ [۱] *فرض کنید S یک مجموعه از اندازه‌ی m و $\mathcal{A}$ یک خانواده‌ی r-پوشش-آزاد شامل n زیرمجموعه از S باشد که $r \leq ۲\sqrt{n}$. در این صورت داریم:*

$$m > \frac{r^{۲} \log(n - \frac{r}{۲})}{۱۰ \log r}.$$

برهان. فرض کنید

$$m \leq \frac{r^{۲} \log(n - \frac{r}{۲})}{۱۰ \log r}.$$

تا زمانی‌که $\mathcal{A}$ شامل مجموعه‌ی A از اندازه‌ی بزرگ‌تر از $\frac{۲m}{r}$ باشد، A را از $\mathcal{A}$ و اعضایش را از S و از هر مجموعه‌ی دیگر در $\mathcal{A}$ خارج می‌کنیم. چون $|S| = m$، این روند بعد از حداکثر $\frac{r}{۲}$ مرحله متوقف می‌شود. بنابراین، اکنون زیرمجموعه‌ی S' از S، و خانواده‌ی $\mathcal{A}'$ از زیرمجموعه‌های S'، که هر زیرمجموعه از اندازه‌ی حداکثر $\frac{۲m}{r}$ است، در اختیار داریم. اندازه‌ی S' را با m' و اندازه‌ی $\mathcal{A}'$ را با n' مشخص می‌کنیم. به‌وضوح،

$$n' \geq n - \frac{r}{۲}.$$

هیچ مجموعه‌ی $A \in \mathcal{A}'$ در اجتماع $\frac{r}{۲}$ مجموعه‌ی دیگر قرار نمی‌گیرد. زیرا در غیر این صورت، مجموعه‌ی اصلی که A از آن به‌دست آمده است در اجتماع این $\frac{r}{۲}$ مجموعه و مجموعه‌هایی که خارج شده‌اند قرار می‌گیرد که این نتیجه با فرض $\mathcal{A}$، r-پوشش-آزاد است در تناقض است. پس هر مجموعه در $\mathcal{A}'$ زیرمجموعه‌ای از اندازه‌ی حداکثر $\lceil \frac{۴m}{r^{۲}} \rceil$ دارد که در هیچ مجموعه‌ی دیگر $\mathcal{A}'$ قرار نمی‌گیرد. زیرا در غیر این صورت، اگر یک مجموعه‌ی $A \in \mathcal{A}'$ باشد به‌طوری‌که در این حالت صدق نکند، یعنی، هر زیرمجموعه‌ی A از اندازه‌ی حداکثر $\lceil \frac{۴m}{r^{۲}} \rceil$ در مجموعه‌ی دیگری در $\mathcal{A}'$ قرار می‌گیرد، آنگاه چون $|A| \leq \frac{۲m}{r}$، مجموعه‌ی A توسط $\frac{r}{۲}$ مجموعه‌ی دیگر در $\mathcal{A}'$ پوشیده خواهد شد که این غیر ممکن است. در نتیجه، n' مجموعه‌ی متمایز از اندازه‌ی حداکثر $\lceil \frac{۴m}{r^{۲}} \rceil$ وجود دارد. بنابراین،

$$n' \leq \binom{m'}{\lceil \frac{۴m}{r^{۲}} \rceil}.$$

اگر $۱ > \frac{۴m}{r^{۲}}$، آنگاه طرف راست نامساوی بالا برابر است با m' و بنابراین داریم $n' \leq m'$ و در نتیجه

$n \le m$، که این مطلب با فرض که

$$m \le \frac{r^۲ \log(n - \frac{r}{۲})}{۱۰ \log r} \le \frac{۴n}{۵}(۱ + o(۱))$$

متناقض است. بنابراین $۱ \ge \frac{۴m}{r^۲}$ و داریم:

$$n - \frac{r}{۲} \le n' \le \binom{m'}{\lceil \frac{۴m}{r^۲} \rceil} \le \binom{m}{\lceil \frac{۴m}{r^۲} \rceil} < ۲^{\frac{۱۰ m \log r}{r^۲}}.$$

در نتیجه،

$$m > \frac{r^۲ \log(n - \frac{r}{۲})}{۱۰ \log r},$$

که با فرض تناقض دارد. □

اکنون لم بالا را برای اثبات کران پایین برای حالت $k \le \sqrt{n}$ به‌کار می‌بریم.

گزاره ۸.۳.۴ [۱] *برای هر $k \le \sqrt{n}$، اگر $\mathcal{F}$ خانواده‌ای باشد که $\mathcal{S}_k$-مسأله را حل می‌کند، آنگاه*

$$|\mathcal{F}| = \Omega(\frac{k^۳ \log n}{\log k}).$$

برهان. فرض کنید $\mathcal{F}$ خانواده‌ای باشد که $\mathcal{S}_k$-مسأله را حل می‌کند. $A, B \subseteq V$ را به‌صورت تصادفی انتخاب می‌کنیم به‌طوری‌که $|A| = ۲$، $|B| = \frac{k}{۲} - ۱$ و $A \cap B = \emptyset$. تعریف می‌کنیم:

$$\mathcal{G} = \{F \in \mathcal{F} \mid |F \cap A| = ۱, F \cap B = \emptyset\}.$$

به‌وضوح

$$\begin{aligned} Pr(F \in \mathcal{G}) &= \frac{|F|(n - |F|)}{\binom{n}{۲}} \cdot \frac{\binom{n-|F|-۱}{\frac{k}{۲}-۱}}{\binom{n-۲}{\frac{k}{۲}-۱}} \\ &= \frac{۲|F|}{n} \cdot \frac{\binom{n-|F|}{\frac{k}{۲}}}{\binom{n-۱}{\frac{k}{۲}}} \\ &\le \frac{۲|F|}{n} \cdot \left(۱ - \frac{|F| - ۱}{n - ۱}\right)^{\frac{k}{۲}} \\ &\le \frac{۲|F|}{n} \cdot e^{-\frac{k|F|}{۴n}}. \end{aligned}$$

اگر $|F| \le \frac{۴n}{k}$، آنگاه

$$Pr(F \in \mathcal{G}) \le \frac{۸}{k}.$$

اگر $|F| > \frac{۴n}{k}$، قرار می‌دهیم

$$x = \frac{k|F|}{۴n}.$$

چون $x > ۱$ داریم:

$$Pr(F \in \mathcal{G}) \leq \frac{۸}{k} \cdot x \cdot e^{-x} < \frac{۸}{ek}.$$

پس برای هر F و به ازای مقدار ثابت c

$$Pr(F \in \mathcal{G}) \leq \frac{c}{k}$$

و بنابراین اندازه‌ی مورد انتظار $\mathcal{G}$ برابر است با $c \cdot \frac{|\mathcal{F}|}{k}$. در نتیجه، انتخاب مناسبی از A و B وجود دارد به‌طوری‌که

$$|\mathcal{G}| \leq c \cdot \frac{|\mathcal{F}|}{k}.$$

قرار می‌دهیم

$$V' = V \setminus (A \cup B),$$

و تعریف می‌کنیم:

$$\mathcal{G}' = \{F \cap V' \mid F \in \mathcal{G}\}.$$

چون $\mathcal{F}$، $\mathcal{S}_k$- مسأله را حل می‌کند، برای هر $u \in V'$، و هر $C \subseteq V' \setminus \{u\}$ از اندازه‌ی $\frac{k}{۲}$، مجموعه‌ی $F \in \mathcal{G}'$ وجود دارد به‌طوری‌که $u \in F$ و $F \cap C = \emptyset$. زیرا در غیر این صورت، $\mathcal{F}$ نمی‌تواند بین دو ستاره که مرکز آن‌ها u است، در $k - ۱$ رأس از $B \cup C$ اشتراک دارند و رأس مکمل یکی از آن‌ها یک عضو از A و رأس مکمل دیگری عضو دیگر A است، تمایز قائل شود. فرض کنید

$$m = |\mathcal{G}'|,$$

$$n' = |V'| = n - \frac{k}{۲} - ۱$$

و M یک ماتریس $m \times n'$ باشد که سطرهای آن بردارهای وقوع مجموعه‌ها در $\mathcal{G}'$ است. اکنون می‌توان ستون‌های M را به‌عنوان بردار وقوع زیرمجموعه‌هایی از یک مجموعه‌ی دیگر از اندازه‌ی m درنظر گرفت.

به ازای هر ستون i و هر مجموعه‌ی J شامل $\frac{k}{2}$ ستون به جز i، سطری وجود دارد که i-امین مؤلفه‌ی آن ۱ است و به ازای هر $j \in J$، j-امین مؤلفه ۰ است. بنابراین، هیچ زیرمجموعه‌ی متناظر با یک ستون در اجتماع $\frac{k}{2}$ زیرمجموعه متناظر با هر $\frac{k}{2}$ ستون دیگر قرار نمی‌گیرد و طبق لم ۷.۳.۴

$$|\mathcal{G}'| = m > \frac{(\frac{k}{2})^2 \log(n' - \frac{k}{4})}{10 \log \frac{k}{2}} = \Omega\Big(\frac{k^2 \log n}{\log k}\Big),$$

و بنابراین

$$|\mathcal{F}| \geq \Omega(k|\mathcal{G}|) \geq \Omega(k|\mathcal{G}'|) \geq \Omega\Big(\frac{k^3 \log n}{\log k}\Big).$$

□

فصل ۵
جستجوی خوشه

۱.۵ مقدمه

گراف مخفی که در این فصل جستجو می‌کنیم، گرافی است روی $V = \{۱, \ldots, n\}$ که یک مؤلفه‌ی همبندی آن خوشه و مؤلفه‌های دیگر آن رأس‌های منفرد هستند. همانند جستجوی ستاره‌ها، این مسأله را نیز، هم برای حالتی که اندازه‌ی خوشه نامعلوم است و هم برای حالتی که اندازه‌ی آن معلوم است بررسی می‌کنیم. برای حل این مسأله، آلن و اُزدی در [۱] تنها از الگوریتم‌های قطعی و غیرانطباقی استفاده کردند و نشان دادند که حتی برای خانواده‌ی همه‌ی خوشه‌ها، مسأله با $O(n \log^۲ n)$ پرسش حل می‌شود.

خانواده‌ی گراف‌های روی $V = \{۱, ۲, \ldots, n\}$ که شامل یک کپی از K_k و $n-k$ رأس منفرد است را با $\mathcal{C}_k$ نشان داده و قرار می‌دهیم

$$\mathcal{C} = \cup_{k=۲}^{n} \mathcal{C}_k.$$

ابتدا کران بالا و کران پایینی برای اندازه‌ی کوچک‌ترین خانواده‌ای که $\mathcal{C}$-مسأله را حل می‌کند ارائه می‌دهیم و در ادامه با فرض اطلاع از مقدار k، این کران‌ها را برای $\mathcal{C}_k$-مسأله و به ازای مقادیر مختلف k، بررسی می‌کنیم.

۲.۵ جستجوی خوشه‌ای با اندازه‌ی نامعلوم

در قضیه‌ی زیر، یک کران بالا و یک کران پایین برای اندازه‌ی کوچک‌ترین خانواده‌ای که $\mathcal{C}$-مسأله را حل می‌کند، بیان می‌شود.

قضیه ۱.۲.۵. [۱] *هر خانواده‌ای که $\mathcal{C}$- مسأله را حل می‌کند دارای اندازه‌ی حداقل $\Omega(n \log n)$ است و خانواده‌ای از اندازه‌ی $O(n \log^2 n)$ وجود دارد که $\mathcal{C}$- مسأله را حل می‌کند.*

ابتدا اثبات کران پایین ارائه شده در قضیه‌ی فوق را بیان می‌کنیم.

گزاره ۲.۲.۵. [۱] *اندازه‌ی کوچک‌ترین خانواده‌ی $\mathcal{F}$ که $\mathcal{C}$- مسأله را حل می‌کند حداقل برابر است با $\Omega(n \log n)$.*

برهان. فرض کنید $\mathcal{F}$، خانواده‌ای است که $\mathcal{C}$- مسأله را حل می‌کند. رأسی دلخواه مانند $u \in V$ و مجموعه‌ی $V_1 = V \setminus \{u\}$ را در نظر می‌گیریم. به منظور تمیز دادن بین گراف کامل روی V_1 و گراف کامل روی $V_1 \cup \{u\}$، $\mathcal{F}$ باید شامل گروه $F_1(u) = \{u, v_1\}$ باشد که $v_1 \in V_1$. حال قرار می‌دهیم

$$V_2 = V_1 \setminus \{v_1\}.$$

به منظور تمیز دادن بین گراف کامل روی V_2 و گراف کامل روی $V_2 \cup \{u\}$، $\mathcal{F}$ باید شامل گروه $F_2(u)$ باشد، به‌طوری‌که

$$u \in F_2(u),\ |F_2(u) \cap V_2| = 1.$$

رأس موجود در $F_2(u) \cap V_2$ را با v_2 نشان می‌دهیم. به همین ترتیب ادامه می‌دهیم و برای هر $1 \leq i \leq n-2$ تعریف می‌کنیم:

$$V_i = V_{i-1} \setminus \{v_{i-1}\}$$

و مجموعه‌ی $F_i(u) \in \mathcal{F}$ را پیدا می‌کنیم که بین گراف کامل روی V_i و گراف کامل روی $V_i \cup \{u\}$ تمایز قائل می‌شود. در این صورت $u \in F_i(u)$ و $|F_i(u) \cap V_i| = 1$. رأس موجود در $F_i(u) \cap V_i$ را v_i می‌نامیم. برای هر $1 \leq i \leq n-2$،

$$|V_i| = n - i,$$

$$|F_i(u) \cap V_i| = 1$$

و

$$|F_i(u)| \leq i + ۱.$$

همچنین، همه‌ی مجموعه‌های $F_i(u)$ متمایزند. زیرا رئوس متمایز v_i وجود دارند که $v_i \in F_i(u)$، اما برای هر $j < i$، $v_i \notin F_j(u)$. برای هر $u \in V$، $\mathcal{F}$ شامل همه‌ی مجموعه‌های $F_i(u)$ است. برای هر رأس $u \in V$، و هر $۱ \leq i \leq n - ۲$، به زوج (u, i) وزن $w(u, i) = \frac{۱}{|F_i(u)|}$ را نسبت می‌دهیم. برای مجموعه‌ی $F \in \mathcal{F}$، حداکثر به تعداد $|F|$، رأس u وجود دارد (رأس‌های موجود در F) که به ازای هر i، $F = F_i(u)$. بنابراین، وزن کلی متناظر با مجموعه‌ی $F \in \mathcal{F}$، حداکثر برابر است با ۱، یعنی

$$\sum_{(u,i):F_i(u)=F} w(u, i) \leq |F| \cdot \frac{۱}{|F|} = ۱.$$

بنابراین

$$\begin{aligned}
|F| &\geq \sum_{F\in\mathcal{F}} \sum_{(u,i):F_i(u)=F} w(u, i) \\
&= \sum_{u\in V} \sum_{i=۱}^{n-۲} w(u, i) \\
&= \sum_{u\in V} \sum_{i=۱}^{n-۲} \frac{۱}{|F_i(u)|} \\
&\geq \sum_{u\in V} \sum_{i=۱}^{n-۲} \frac{۱}{i + ۱} \\
&= \Omega(n \log n).
\end{aligned}$$

□

گزاره ۵.۲.۳. [۱] *خانواده‌ی $\mathcal{F}$ از اندازه‌ی $O(n \log^۲ n)$ وجود دارد به‌طوری‌که C- مسأله را حل می‌کند.*

برهان. خانواده‌ی $\mathcal{F}$ را به‌طور بازگشتی به‌صورت زیر می‌سازیم.

نخست، مجموعه‌ی V در $\mathcal{F}$ قرار دارد. V را به دو نیمه‌ی $V_۱$ و $V_۲$ افراز می‌کنیم و بخشی از خوشه را که در هر نیمه قرار دارد جستجو می‌کنیم. خوشه‌ی مورد نظر اجتماع خوشه‌های پیدا شده در $V_۱$ و $V_۲$ است. این روش تا زمانی که قسمتی از خوشه در هر V_i از اندازه‌ی ۰ یا از اندازه‌ی حداقل ۲ است جواب می‌دهد. اما

اگر بخشی از خوشه در V_i از اندازه‌ی ۱ باشد، آنگاه پاسخ Q_{V_i} "منفی" است و در نتیجه به‌منظور شناسایی این رأس به تعدادی پرسش تکمیلی نیاز داریم. بدون کاستن از کلیت، فرض کنید خوشه دارای یک رأس در $V_۱$ باشد. نشان می‌دهیم این رأس را می‌توان با پاسخ پرسش‌های زیر جستجو کرد. $(۰,۱)$-بردارهای متمایزی، به طول $\lceil \log |V_۱| \rceil$، به رئوس $V_۱$ نسبت می‌دهیم. به ازای هر $۱ \leq i \leq \lceil \log |V_۱| \rceil$ و $j \in \{۰,۱\}$ و $u \in V_۲$، گروه

$$F(i,j,u) = \{v \in V_۱ \mid \text{امین مؤلفه‌ی } v\text{، } j \text{ است-} i\} \cup \{u\}$$

را در $\mathcal{F}$ قرار می‌دهیم. اگر پاسخ Q_V "مثبت" و پاسخ $Q_{V_۱}$ "منفی" باشد، در این صورت، حداقل یک رأس u از خوشه در $V_۲$ قرار دارد. اگر هیچ رأسی از خوشه در $V_۱$ نباشد آنگاه پاسخ هر پرسش $Q_{F(i,j,u)}$ "منفی" است. در غیر این صورت، یک رأس v از خوشه در $V_۱$ قرار دارد. پاسخ $Q_{F(i,j,u)}$ "مثبت" است اگر و تنها اگر u در خوشه باشد و i-امین مؤلفه‌ی v، j باشد. از آنجا که حداقل یک رأس از خوشه در $V_۲$ قرار دارد، می‌توان v را از پاسخ این پرسش‌ها به‌دست آورد. همچنین برای حالتی که $V_۲$ شامل تنها یک رأس از خوشه است باید پرسش‌های مشابهی در نظر بگیریم. فرض کنید $f(n)$ تعداد پرسش‌های مورد نیاز برای n رأس باشد. پس طبق بحث بالا داریم:

$$f(n) \leq ۴ \cdot \frac{n}{۲} \cdot \log \frac{n}{۲} + ۲f(\frac{n}{۲}) + ۱ = O(n \log^۲ n).$$

□

۳.۵ جستجوی خوشه‌ای با اندازه‌ی معلوم

اکنون کران‌های بالا و کران‌های پایینی را برای حالتی که اندازه‌ی خوشه داده شده است، ارائه می‌دهیم.

قضیه ۱.۳.۵. [۱] *به ازای هر k، خانواده‌ی $\mathcal{F}$ از اندازه‌ی $O(k^۲ \log n)$ وجود دارد که $\mathcal{C}_k$-مسأله را حل می‌کند و هر خانواده‌ای که $\mathcal{C}_k$-مسأله را حل می‌کند یا شامل $\Omega(n)$ جفت است یا اندازه‌ی آن حداقل برابر $\Omega(\frac{k^۲}{\log n})$ است. علاوه بر این، برای هر $k \leq n^{\frac{۱}{۳}}$، اندازه‌ی هر خانواده که $\mathcal{C}_k$-مسأله را حل می‌کند حداقل برابر است با $\Omega(\frac{k^۲ \log n}{\log k})$. همچنین به ازای $k \leq \sqrt{n}$، این کران حداقل $\Omega(k^۲)$ است. برای هر s، خانواده‌ای*

از اندازه‌ی $\lceil \frac{n}{2} \rceil(1+s)$ موجود است که $\mathcal{C}_{n-s}$- مسأله را حل می‌کند.

بهترین کران‌های به‌دست آمده، برای مقادیر مختلف k، در جدول ۱.۵ خلاصه شده است. در ادامه‌ی این بخش، این نتایج اثبات می‌شوند.

گزاره ۲.۳.۵. [۱] *به ازای هر k، خانواده‌ی $\mathcal{F}$ از اندازه‌ی $O(k^2 \log n)$ وجود دارد که $\mathcal{C}_k$- مسأله را حل می‌کند.*

برهان. به ازای مقدار ثابت ومثبت c، قرار دهید

$$m = ck^2 \log n.$$

فرض کنید $F_1, F_2, \dots, F_m$ ،m زیرمجموعه‌ی تصادفی از V باشند که به‌طور مستقل و به‌صورت زیر انتخاب شده‌اند. در انتخاب F_i، هر $v \in V$ به‌طور مستقل و با احتمال $\frac{1}{k}$ در F_i قرار می‌گیرد. برای هر دو خوشه‌ی $C, C' \in \mathcal{C}_k$، پیشامد $A_{C,C'}$ عبارت است از این که هیچ یک از F_iها بین C و C' تمایز قائل نشود. C_1 و C_2 را دو گراف کامل از اندازه‌ی k در نظر بگیرید به‌طوری‌که

$$|V(C_1) \setminus V(C_2)| = |V(C_2) \setminus V(C_1)| = 1.$$

فرض کنید $v_1, \dots, v_{k-1}$ رئوس مشترک C_1 و C_2 باشند و برای $i = 1, 2$، u_i رأس مکمل C_i باشد. F_i بین C_1 و C_2 تمایز قائل می‌شود اگر و تنها اگر دقیقاً یکی از دو رأس u_1 و u_2 و دقیقاً یکی از رأس‌های $v_1, \dots, v_{k-1}$ در F_i قرار گیرد. بنابراین، احتمال این‌که F_i بین C_1 و C_2 تمایز قائل شود برابر است با

$$\frac{2}{k} \cdot \frac{k-1}{k}\Big(1 - \frac{1}{k}\Big)^{k-1} = \Omega\Big(\frac{1}{k}\Big).$$

پس، به ازای مقدار مناسب c، احتمال این‌که هیچ یک از F_iها بین C_1 و C_2 تمایز قائل نشود برابر است با

$$\Big(1 - \Omega\Big(\frac{1}{k}\Big)\Big)^m \leq n^{-2k}.$$

برای دو خوشه که در تعداد بیشتری رأس اختلاف دارند، این احتمال کم‌تر است. از طرفی،

$$|\mathcal{C}_k| = \binom{n}{k}$$

جدول ۱.۵ : کران‌هایی برای اندازه‌ی خانواده‌ای که $\mathcal{C}_k$- مسأله را حل می‌کند.

k	کران پایین	کران بالا
$k \leq n^{\frac{1}{3}}$	$\Omega(\frac{k^2 \log n}{\log k})$	$O(k^2 \log n)$
$n^{\frac{1}{3}} < k \leq \sqrt{n}$	$\Omega(k^2)$	$O(k^2 \log n)$
$\sqrt{n} < k < \sqrt{n \log n}$	$\Omega(\frac{k^2}{\log n})$	$O(k^2 \log n)$
$\sqrt{n \log n} \leq k \leq n - \log^2 n$	$\Omega(n)$	$O(n \log^2 n)$
$k = n - s,\ s < \log^2 n$	$\Omega(n)$	$(s+1)\lceil \frac{n}{2} \rceil$

و در نتیجه، تعداد جفت خوشه‌ها برابر است با:

$$\binom{|\mathcal{C}_k|}{2} < \binom{n}{k}^2 < n^{2k}.$$

بنابراین

$$Pr(\bigcup_{C,C' \in \mathcal{C}_k} A_{C,C'}) \leq \sum_{C,C' \in \mathcal{C}_k} Pr(A_{C,C'}) < 1$$

و در نتیجه

$$Pr(\bigcap_{C,C' \in \mathcal{C}_k} \bar{A}_{C,C'}) > 0.$$

پس خانواده‌ی $\mathcal{F} = \{F_1, F_2, \ldots, F_m\}$ وجود دارد که $\mathcal{C}_k$- مسأله را حل می‌کند. □

گزاره ۳.۳.۵. [۱] *به ازای هر k، اگر خانواده‌ی $\mathcal{F}$، $\mathcal{C}_k$- مسأله را حل کند آنگاه $\mathcal{F}$ یا شامل $\Omega(n)$ جفت است یا این خانواده از اندازه‌ی حداقل $\Omega(\frac{k^2}{\log n})$ است.*

برهان. به‌وضوح، می‌توان فرض کرد $k^2 > \log n$، چون در غیر این صورت چیزی برای اثبات نداریم. فرض کنید $\mathcal{F}$ خانواده‌ای باشد که $\mathcal{C}_k$- مسأله را حل می‌کند. در این صورت برای هر $A, B \subseteq V$ که $|A| = 2$ و $|B| = k - 1$ و $A \cap B = \emptyset$، مجموعه‌ی $F \in \mathcal{F}$ وجود دارد به‌طوری‌که $|F \cap A| = 1$ و $|F \cap B| = 1$. زیرا در غیر این صورت، $\mathcal{F}$ نمی‌تواند بین گراف کامل روی B و یک رأس از A، و گراف کامل روی B و رأس دیگر A تمایز قائل شود. قرار می‌دهیم $\mathcal{F}_0$ خانواده‌ی همه‌ی مجموعه‌های $F \in \mathcal{F}$ از اندازه‌ی ۲ باشد.

فرض کنید:

$$m = c \cdot \frac{n \log n}{k},$$

تعریف می‌کنیم:

$$\mathcal{F}_1 = \{F \in \mathcal{F} \mid 2 < |F| \leq m\}$$

و

$$\mathcal{F}_2 = \mathcal{F} \setminus (\mathcal{F}_0 \cup \mathcal{F}_1).$$

نشان می‌دهیم اگر

$$|\mathcal{F}_0| \leq \frac{1}{10} n$$

آنگاه به ازای مقدار ثابت c_1 که فقط به c وابسته است داریم:

$$|\mathcal{F}_1 \cup \mathcal{F}_2| > c_1 \cdot \frac{k^2}{\log n}.$$

فرض کنید:

$$|\mathcal{F}_0| \leq \frac{1}{10} n$$

و

$$|\mathcal{F}_1 \cup \mathcal{F}_2| \leq c_1 \cdot \frac{k^2}{\log n}.$$

زیرمجموعه‌ی $A = \{u, v\}$ از V را به‌طور یکنواخت انتخاب می‌کنیم و تعریف می‌کنیم:

$$\mathcal{F}_1' = \{F \in \mathcal{F}_1 \mid |F \cap A| = 1\}.$$

برای هر $F \in \mathcal{F}_1$ داریم:

$$\begin{aligned} Pr(F \in \mathcal{F}_1') &= 2 \cdot \frac{|F|}{n} \cdot \frac{n - |F|}{n - 1} \\ &\leq 2 \cdot \frac{|F|}{n}. \end{aligned}$$

بنابراین

$$
\begin{aligned}
E(|\mathcal{F}'_۱|) &\leq ۲ \sum_{F\in\mathcal{F}_۱} \frac{|F|}{n} \\
&\leq ۲|\mathcal{F}_۱| \cdot \frac{m}{n}.
\end{aligned}
$$

طبق نامساوی مارکوف (۱.۱.۳)، احتمال این‌که $|F'_۱| > ۴|\mathcal{F}_۱| \cdot \frac{m}{n}$ حداکثر $\frac{۱}{۲}$ است.

برای هر $F \in \mathcal{F}_۰$ داریم:

$$
\begin{aligned}
Pr(F\cap A \neq \emptyset) &= ۲\frac{|F|}{n} \cdot \frac{n-|F|}{n-۱} + \frac{|F|(|F|-۱)}{n(n-۱)} \\
&= \frac{۴n-۶}{n(n-۱)}
\end{aligned}
$$

چون $|\mathcal{F}_۰| \leq \frac{۱}{۱۰}n$، احتمال این‌که مجموعه‌ی $F \in \mathcal{F}_۰$ وجود داشته باشد به‌طوری‌که $F\cap A \neq \emptyset$، کم‌تر است از

$$
\sum_{F\in\mathcal{F}_۰} \frac{۴n-۶}{n(n-۱)} < \frac{۲}{۵}.
$$

بنابراین، مجموعه‌ی A وجود دارد به‌طوری‌که برای هر $F \in \mathcal{F}_۰$،

$$
F\cap A = \emptyset
$$

و

$$
|F'_۱| \leq ۴|\mathcal{F}_۱| \cdot \frac{m}{n}.
$$

به ازای چنین انتخاب A و مقادیر مناسب c و $c_۱$ داریم:

$$
\begin{aligned}
|\mathcal{F}'_۱| &\leq ۴|\mathcal{F}_۱| \cdot \frac{m}{n} \\
&\leq ۴c_۱c \cdot \frac{k^۲}{\log n} \cdot \frac{n\log n}{kn} \\
&\leq \frac{k-۱}{۴}.
\end{aligned}
$$

بنابراین، مجموعه‌ی $B_۱ \subseteq V\setminus A$ از اندازه‌ی $\frac{k-۱}{۲}$ وجود دارد به‌طوری‌که برای هر $F \in \mathcal{F}'_۱$، $|F\cap B_۱| \geq ۲$.

زیرمجموعه‌ی تصادفی $B_۲ \subseteq V$ را از اندازه‌ی $\frac{k-۱}{۲}$ انتخاب می‌کنیم. برای هر $F \in \mathcal{F}_۲$ و به ازای مقدار

ثابت $c_۲ = \Theta(c)$ داریم:

$$\begin{aligned} Pr(|F \cap B_۲| \leq ۱) &= \frac{\binom{n-|F|}{\frac{k-۱}{۲}}}{\binom{n}{\frac{k-۱}{۲}}} + \frac{k-۱}{۲} \cdot \frac{|F|}{n} \cdot \frac{\binom{n-|F|}{\frac{k-۱}{۲}-۱}}{\binom{n-۱}{\frac{k-۱}{۲}-۱}} \\ &\leq \left(۱ - \frac{|F|}{n}\right)^{\frac{k-۱}{۲}} + \frac{k-۱}{۲} \cdot \frac{|F|}{n} \cdot \left(۱ - \frac{|F|-۱}{n-۱}\right)^{\frac{k-۱}{۲}-۱} \\ &\leq e^{-\frac{k-۱}{۲} \cdot \frac{m}{n}} + \frac{k-۱}{۲} \cdot \frac{n}{n} \cdot e^{-(\frac{k-۱}{۲}-۱)\frac{m-۱}{n-۱}} \\ &\leq n^{-c_۲}. \end{aligned}$$

بنابراین، اگر c به اندازه‌ی کافی بزرگ باشد آنگاه با احتمال بالا $A \cap B_۲ = \emptyset$ و برای هر $F \in \mathcal{F}_۲$، خواهیم داشت $|F \cap B_۲| \geq ۲$. قرار می‌دهیم

$$B' = B_۱ \cup B_۲.$$

به‌وضوح، $B' \subseteq V \setminus A$ و $|B'| \leq k - ۱$. فرض کنید B یک توسیع دلخواه از B' به زیرمجموعه‌ای از $V \setminus A$ از اندازه‌ی $k - ۱$ باشد. همچنین فرض کنید $C_۱$ یک گراف کامل روی $B \cup \{u\}$ و $C_۲$ یک گراف کامل روی $B \cup \{v\}$ باشد. از آنجا که برای هر $F \in \mathcal{F}_۰$، $u, v \notin F$، هیچ مجموعه‌ای در $\mathcal{F}_۰$ نمی‌تواند بین $C_۱$ و $C_۲$ تمایز قائل شود. از مجموعه‌های موجود در $\mathcal{F}_۱$، نه آن مجموعه‌هایی که شامل هر دو رأس u و v هستند و نه آن مجموعه‌هایی که شامل هیچ‌کدام از آن‌ها نیستند نیز نمی‌توانند بین این دو خوشه تمایز قائل شوند. همه‌ی مجموعه‌های دیگر در $\mathcal{F}_۱$، یعنی مجموعه‌هایی که دقیقاً شامل یک رأس از میان u و v هستند و همه‌ی مجموعه‌های موجود در $\mathcal{F}_۲$، شامل حداقل دو رأس از B هستند. بنابراین، این مجموعه‌ها نیز نمی‌توانند بین این دو خوشه تمایز قائل شوند. بنابراین، $\mathcal{F}$ نمی‌تواند دو گراف $C_۱$ و $C_۲$ را تشخیص دهد. پس با این فرض که این خانواده $\mathcal{C}_k$-مسأله را حل می‌کند به تناقض رسیده‌ایم. □

اکنون یک کران پایین بهتر برای حالت $k \leq n^{\frac{۱}{۴}}$ ارائه می‌دهیم. ابتدا لم زیر را بیان می‌کنیم.

لم ۴.۳.۵. [۱] *فرض کنید S یک مجموعه از اندازه‌ی m و $\mathcal{A}$ خانواده‌ای شامل n زیرمجموعه از S باشد. همچنین فرض کنید مجموعه‌های متمایز $A, B_۱, \ldots, B_r, C_۱, \ldots, C_r \in \mathcal{A}$ وجود ندارند به‌طوری که*

$$A \subseteq \bigcup_{i=۱}^{r} B_i$$

و

$$A \subseteq \bigcup_{i=1}^{r} C_i$$

که $r \leq n^{\frac{1}{3}}$. *در این صورت*

$$m = \Omega(\frac{r^2 \log n}{\log r}).$$

برهان. قرار دهید $\mathcal{B} = \emptyset$. تا زمانی که مجموعه‌های $A, B_1, \dots, B_r \in \mathcal{A}$ وجود دارند به‌طوری‌که

$$A \subseteq \bigcup_{i=1}^{r} B_i$$

$A, B_1, \dots, B_r$ را از $\mathcal{A}$ خارج کرده و A را به $\mathcal{B}$ اضافه می‌کنیم. فرض کنید $\mathcal{A}'$ خانواده‌ی به‌دست آمده از $\mathcal{A}$ در پایان این فرایند باشد و اندازه‌ی $\mathcal{B}$ را با l نشان می‌دهیم. بنابراین

$$|\mathcal{A}'| = n - l(r+1).$$

دو خانواده‌ی $\mathcal{A}'$ و $\mathcal{B}$، r-پوشش-آزاد هستند. به‌وضوح $\mathcal{A}'$، r-پوشش-آزاد است. زیرا در غیر این صورت، فرایند بالا متوقف نخواهد شد. $\mathcal{B}$ نیز r-پوشش-آزاد است، چون اگر مجموعه‌های $A, C_1, \dots, C_r \in \mathcal{B}$ موجود باشند به‌طوری‌که

$$A \subseteq \bigcup_{i=1}^{r} C_i,$$

آنگاه مجموعه‌های $B_1, \dots, B_r \in \mathcal{A}$ نیز وجود دارند که با A از $\mathcal{A}$ خارج شده‌اند و داریم:

$$A \subseteq \bigcup_{i=1}^{r} B_i$$

که با فرض تناقض دارد. اگر $l \geq \frac{n^{\frac{2}{3}}}{4}$، از آنجا که $r \leq n^{\frac{1}{3}}$، طبق لم ۷.۳.۴ داریم:

$$m > \frac{r^2 \log(l - \frac{r}{2})}{10 \log r} = \Omega\Big(\frac{r^2 \log n}{\log r}\Big).$$

و اگر $l < \frac{n^{\frac{2}{3}}}{4}$ ، چون $r \leq n^{\frac{1}{3}}$، داریم:

$$|\mathcal{A}'| = n - l(r+1) > \frac{n}{2}.$$

بنابراین، طبق لم ۷.۳.۴

$$m > \frac{r^2 \log(\frac{n}{2} - \frac{r}{2})}{10 \log r} = \Omega\Big(\frac{r^2 \log n}{\log r}\Big).$$

□

گزاره ۵.۳.۵. [۱] *به ازای هر* $k \leq n^{1/3}$، *اگر* $\mathcal{F}$ *خانواده‌ای باشد که* $\mathcal{C}_k$*- مسأله را حل کند آنگاه*

$$|\mathcal{F}| = \Omega\Big(\frac{k^2 \log n}{\log k}\Big).$$

برهان. فرض کنید $\mathcal{F}$ خانواده‌ای است که $\mathcal{C}_k$- مسأله را حل می‌کند. تعریف می‌کنیم $m = |\mathcal{F}|$ و M یک ماتریس $m \times n$ است که سطرهای آن بردار وقوع مجموعه‌های موجود در $\mathcal{F}$ است. ستون‌های M را به‌عنوان بردار وقوع زیرمجموعه‌های یک مجموعه‌ی دیگر از اندازه‌ی m در نظر می‌گیریم. برای هر $1 \leq i \leq n$، G_i را زیرمجموعه‌ی متناظر با i- امین ستون M تعریف می‌کنیم. خانواده‌ی $\mathcal{G}$ را به‌صورت زیر تعریف می‌کنیم.

$$\mathcal{G} = \Big\{G_{2i-1} \cup G_{2i} \mid 1 \leq i \leq \frac{n}{2}\Big\}.$$

ادعا می‌کنیم مجموعه‌های متمایز $A, B_1, \ldots, B_{\frac{k-1}{4}}, C_1, \ldots, C_{\frac{k-1}{4}} \in \mathcal{G}$ وجود ندارند به‌طوری‌که

$$A \subseteq \bigcup_{i=1}^{\frac{k-1}{4}} B_i \tag{۱.۵}$$

و

$$A \subseteq \bigcup_{i=1}^{\frac{k-1}{4}} C_i. \tag{۲.۵}$$

فرض کنید چنین مجموعه‌هایی موجود باشند. A اجتماع دو زیرمجموعه متناظر با دو ستون متمایز از M است. فرض کنید u و v رأس‌های متناظر با این دو ستون باشند. به‌طور مشابه، فرض کنید $w_1, \ldots, w_{k-1}$ رئوس متناظر با $B_1, \ldots, B_{\frac{k-1}{4}}, C_1, \ldots, C_{\frac{k-1}{4}}$ باشند. اعضای A گروه‌هایی شامل u یا v هستند. طبق (۱.۵) و (۲.۵) هر چنین گروهی شامل حداقل دو رأس از $w_1, \ldots, w_{k-1}$ است. بنابراین، هیچ گروهی بین گراف کامل روی $u, w_1, \ldots, w_{k-1}$ و گراف کامل روی $v, w_1, \ldots, w_{k-1}$ تمایز قائل نمی‌شود. در نتیجه، چنین مجموعه‌هایی در $\mathcal{G}$ وجود ندارد و طبق لم ۴.۳.۵ با $r = \frac{k-1}{4}$ و $\mathcal{A} = \mathcal{G}$، داریم:

$$|\mathcal{F}| = m = \Omega\Big(\frac{k^2 \log n}{\log k}\Big).$$

□

در ادامه ثابت می‌شود برای هر $n^{\Omega(1)} \leq k \leq \sqrt{n}$، هر خانواده که $\mathcal{C}_k$- مسأله را حل می‌کند از اندازه‌ی حداقل $\Omega(k^2)$ است.

تعریف ۶.۳.۵. [۱] فرض کنید A زیرمجموعه‌ای از مجموعه‌ی S و $\mathcal{A}$ خانواده‌ای از زیرمجموعه‌های S باشد. گفته می‌شود A دوبار با $\mathcal{A}$ پوشانده می‌شود اگر برای هر $a \in A$، حداقل دو مجموعه در $\mathcal{A}$ و شامل a وجود داشته باشد.

لم ۷.۳.۵. [۱] *S مجموعه‌ای از اندازه‌ی m و $\mathcal{A}$ خانواده‌ای شامل n زیرمجموعه از S است. فرض کنید هیچ مجموعه‌ای در $\mathcal{A}$ با هر r مجموعه‌ی دیگر در $\mathcal{A}$ دوبار پوشانده نمی‌شود که $n^{\Omega(1)} \leq r \leq \sqrt{n}$. در این صورت $m = \Omega(r^2)$.*

برهان. فرض کنید به ازای ثابت کوچک $\epsilon > 0$،

$$m \leq \epsilon r^2.$$

نشان می‌دهیم اگر ϵ به اندازه‌ی کافی کوچک باشد آنگاه مجموعه‌ی $A \in \mathcal{A}$ وجود دارد که با r مجموعه‌ی دیگر در $\mathcal{A}$، دوبار پوشانده می‌شود. تا زمانی که عضو $a \in S$ وجود دارد به‌طوری‌که متعلق به یک یا دو مجموعه در $\mathcal{A}$ است، این مجموعه‌ها را از $\mathcal{A}$ خارج می‌کنیم. بعد از خارج کردن این مجموعه‌ها a به هیچ مجموعه‌ای در $\mathcal{A}$ تعلق ندارد. بنابراین، این روند بعد از حداکثر m مرحله متوقف می‌شود و از آن به بعد هر $a \in S$ یا در هیچ مجموعه‌ای قرار ندارد یا متعلق به حداقل سه مجموعه است. فرض کنید $\mathcal{A}'$ خانواده‌ی مجموعه‌های باقی‌مانده است و اندازه‌ی آن را با n' مشخص می‌کنیم. بنابراین،

$$n' \geq n - 2m \geq n - 2\epsilon r^2 \geq (1 - 2\epsilon)n.$$

اگر مجموعه‌ی $A \in \mathcal{A}'$ موجود باشد به‌طوری‌که $|A| \leq \frac{r}{2}$، آنگاه این مجموعه با خانواده‌ای شامل حداکثر r مجموعه از $\mathcal{A}' \setminus \{A\}$، با انتخاب دو مجموعه‌ی دلخواه که هر عضو A را دربردارد، دوبار پوشانده می‌شود. حال فرض می‌کنیم برای هر $\mathcal{A}'$

$$|A| > \frac{r}{2}.$$

$\frac{r}{2}$ مجموعه‌ی $B_1, \ldots, B_{\frac{r}{2}} \in \mathcal{A}'$ را به‌طور تصادفی انتخاب می‌کنیم. مجموعه‌ی C را مجموعه‌ی شامل هر $a \in S$ در نظر می‌گیریم که حداکثر به یک مجموعه از $B_1, \ldots, B_{\frac{r}{2}}$ تعلق دارد. اکنون مجموعه‌ی دیگر $A \in \mathcal{A}'$ را به‌صورت تصادفی انتخاب می‌کنیم. اگر $|A \cap C| \leq \frac{r}{4}$ آنگاه برای هر $a \in A \cap C$، دو مجموعه در $\mathcal{A}' \setminus \{A\}$ شامل a انتخاب می‌کنیم. این مجموعه‌ها با $B_1, \ldots, B_{\frac{r}{2}}$ ، خانواده‌ای از حداکثر r مجموعه تشکیل می‌دهند که A را دوبار می‌پوشاند. نشان می‌دهیم

$$E(|A \cap C|) \leq \frac{r}{5}$$

و بنابراین، انتخاب $A, B_1, \ldots, B_{\frac{r}{2}}$ وجود دارد به‌طوری‌که برای هر $i = 1, \ldots, \frac{r}{2}$، $A \neq B_i$ و $|A \cap C| \leq \frac{r}{4}$. بنابراین، A با r مجموعه‌ی دیگر دوبار پوشانده می‌شود که با فرض تناقض دارد. عضو $a \in S$ را در نظر گرفته و تعداد مجموعه‌ها در $\mathcal{A}'$ را که شامل a هستند با k نشان می‌دهیم. احتمال این‌که $a \in A \cap C$ حداکثر برابر است با

$$\begin{aligned}
\frac{k}{n'}\left(\frac{\binom{n'-k}{\frac{r}{2}}}{\binom{n'}{\frac{r}{2}}} + \frac{k\binom{n'-k}{\frac{r}{2}-1}}{\binom{n'}{\frac{r}{2}}}\right) &= \frac{k}{n'}\left(\frac{\binom{n'-k}{\frac{r}{2}}}{\binom{n'}{\frac{r}{2}}} + \frac{kr}{2n'}\frac{\binom{n'-k}{\frac{r}{2}-1}}{\binom{n'-1}{\frac{r}{2}-1}}\right) \\
&\leq \frac{k}{n'}\Big(1 - \frac{k}{n'}\Big)^{\frac{r}{2}} + \frac{k^2 r}{2n'^2}\Big(1 - \frac{k-1}{n'-1}\Big)^{\frac{r}{2}-1} \\
&\leq \frac{k}{n'}e^{-\frac{kr}{2n'}} + \frac{k^2 r}{2n'^2}e^{-\frac{kr}{4n'}}.
\end{aligned}$$

اکنون نشان می‌دهیم که به ازای مقدار ثابت c، این احتمال حداکثر $\frac{c}{r}$ است. ابتدا عبارت $\frac{k}{n'}e^{-\frac{kr}{2n'}}$ را در نظر می‌گیریم. اگر $k \leq \frac{2n'}{r}$، آنگاه این عبارت حداکثر برابر $\frac{2}{r}$ است. اگر $k > \frac{2n'}{r}$، قرار می‌دهیم

$$x = \frac{kr}{2n'}.$$

چون $x > 1$ داریم:

$$\frac{k}{n'}e^{-\frac{kr}{2n'}} = \frac{2}{r}xe^{-x} < \frac{2}{er}.$$

اکنون عبارت $\frac{k^2 r}{2n'^2}e^{-\frac{kr}{4n'}}$ را در نظر می‌گیریم. اگر $k \leq \frac{8n'}{r}$، آنگاه این عبارت حداکثر برابر $\frac{32}{r}$ است. اگر $k > \frac{8n'}{r}$، آنگاه قرار می‌دهیم

$$x = \frac{kr}{4n'}.$$

$x > ۲$ و

$$\frac{k^۲ r}{۲n'^۲} e^{-\frac{kr}{۴n'}} = \frac{۸}{r} x^۲ e^{-x}.$$

به‌سادگی می‌توان بررسی کرد که برای هر $x > ۲$، $x^۲ e^{-x}$ نزولی است و بنابراین

$$\frac{۸}{r} x^۲ e^{-x} < \frac{۳۲}{e^۲ r}.$$

در نتیجه، به ازای مقدار ثابت c، احتمال این‌که $a \in A \cap C$ حداکثر $\frac{c}{r}$ است. بنابراین داریم:

$$E(|A \cap C|) \leq \frac{cm}{r} \leq \frac{c\epsilon r^۲}{r} \leq \frac{r}{۵},$$

به شرط آن که ϵ به اندازه‌ی کافی کوچک باشد و اثبات لم کامل می‌شود. □

گزاره ۸.۳.۵. [۱] *برای هر $n^{\Omega(۱)} \leq k \leq \sqrt{n}$، اگر $\mathcal{F}$ خانواده‌ای باشد که $\mathcal{C}_k$- مسأله را حل کند آنگاه*

$$|\mathcal{F}| = \Omega(k^۲).$$

برهان. برای $n^{\Omega(۱)} \leq k \leq n^{\frac{۱}{۴}}$ این نتیجه از گزاره‌ی ۵.۳.۵ استنباط می‌شود. بنابراین، فرض می‌کنیم $k > n^{\frac{۱}{۴}}$ و $\mathcal{F}$ خانواده‌ای است که $\mathcal{C}_k$-مسأله را حل می‌کند. تعریف می‌کنیم $m = |\mathcal{F}|$ و M ماتریس $m \times n$ است که سطرهای آن بردار وقوع مجموعه‌ها در $\mathcal{F}$ است. ستون‌های M را به‌عنوان بردار وقوع زیرمجموعه‌های یک مجموعه‌ی دیگر از اندازه‌ی m در نظر می‌گیریم. برای هر $۱ \leq i \leq n$، G_i را زیرمجموعه‌ی متناظر با i- امین ستون M تعریف می‌کنیم و قرار می‌دهیم:

$$\mathcal{G} = \Big\{ G_{۲i-۱} \cup G_{۲i} \mid ۱ \leq i \leq \frac{n}{۲} \Big\}.$$

ادعا می‌کنیم مجموعه‌های متمایز $A, B_۱, \ldots, B_{\frac{k-۱}{۲}} \in \mathcal{G}$ وجود ندارند به‌طوری‌که A با $B_۱, \ldots, B_{\frac{k-۱}{۲}}$ دوبار پوشانده شود. فرض کنید این مجموعه‌ها موجود باشند. A اجتماع دو زیرمجموعه متناظر با دو ستون متمایز از M است. رأس‌های متناظر با این دو ستون را u و v می‌نامیم. به‌طور مشابه، فرض می‌کنیم $w_۱, \ldots, w_{k-۱}$ رأس‌های متناظر با $B_۱, \ldots, B_{\frac{k-۱}{۲}}$ باشند. اعضای A گروه‌هایی شامل u یا v هستند. چون A دوبار با $B_۱, \ldots, B_{\frac{k-۱}{۲}}$ پوشانده می‌شود، هریک از این گروه‌ها شامل حداقل دو رأس از رأس‌های $w_۱, \ldots, w_{k-۱}$ هستند. بنابراین، هیچ گروهی بین گراف کامل روی $u, w_۱, \ldots, w_{k-۱}$ و گراف کامل روی

$v, w_1, \ldots, w_{k-1}$ تمایز قائل نمی‌شود. درنتیجه این مجموعه‌ها در $\mathcal{G}$ وجود ندارند. بنابراین طبق لم ۷.۳.۵ داریم:

$$|\mathcal{F}| = m = \Omega(k^2).$$

□

این بخش را با یک کران بالای ساده پایان می‌دهیم. این کران، تقریب‌های قبل را برای خوشه‌هایی که تقریباً همه‌ی رئوس را دربردارند بهبود می‌دهد.

گزاره ۹.۳.۵. [۱] *به‌ازای هر s، خانواده‌ای با اندازه‌ی حداکثر $\lceil \frac{n}{2} \rceil (s+1)$ هست که $\mathcal{C}_{n-s}$- مسأله را حل می‌کند.*

برهان. برای هر $u \in V$، $s+1$ جفت از رئوس را که شامل u است، مورد پرسش قرار می‌دهیم. u در خوشه است اگر و تنها اگر پاسخ حداقل یکی از این پرسش‌ها ”مثبت“ باشد. □

مراجع

[1] N. Alon and V. Asodi. Learning a hidden subgraph. *31st International Colloquium on Automata, Languages and programming*, pages 110–121, 2004.

[2] N. Alon, R. Beigel, S. Kasif, S. Rudich, and B. Sudakov. Learning a hidden matching. *Proceedings of the 43rd IEEE FOCS*, pages 197–206, 2002.

[3] N. Alon and J. H. Spencer. *The Probabilistic Method.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., New York, second edition, 2000.

[4] R. Beigel, N. Alon, M. S. Apaydin, L. Fortnow, and S. Kasif. An optimal procedure for gap closing in whole genome shotgun sequencing. *ACM Press*, pages 22–30, 2001.

[5] R. Dorfman. The detection of defective members of large populations. *Ann. Math. Statist*, pages 436–440, 1943.

[6] D. Z. Due and F. K. Hwang. *Combinatorial group testing and it' s applications.* Series on Applied Mathematics. Co. Pte. Ltd. second edition, 2000.

[7] D. Z. Due and F. K. Hwang. *pooling Designs and nonadaptive group testing important tools for DNA sequencing.* World scientific, 2006.

[8] W. Feller. *An introduction to Probability Theory and it' Applications.* Series on Applied Mathematics. 1950.

[9] V. Grebinski and G. Kucherov. Optimal query bounds for reconstructing a hamiltonian cycle in complete graphs. *Proc. 5th Israeli Symposium on Theoretical Computer Science*, pages 166–173, 1997.

[10] V. Grebinski and G. Kucherov. Reconstructing a hamiltonian cycle by querying the graph: Application to dna physical mapping. *Discrete Applied Math.*, pages 147–165, 1998.

[11] V. Grebinski and G. Kucherov. Optimal reconstruction of graphs under the additive model. *Algorithmica*, pages 104–124, 2000.

[12] S. Jukna. *Extremal Combinatorics With Applications in Computer Science.* Frankfurt, 2001.

[13] C. H. Lie. A sequential method for screening exprimental variables. *J. Amer. Statist. Assoc*, pages 455–477, 1962.

[14] M. Sobel and P. A. Groll. Group testing to eliminate afficiently all defectives in a binomial sample. *Bell System Tech. J*, pages 1179–1252, 1959.

[15] R. M. Wilson. Decomposition of complete graphs into subgraphs isomorphic to a given graph. *Congressus Numerantium* ***XV***, pages 647–659, 1975.

واژه‌نامه فارسی به انگلیسی

(۰, ۱) - بردار	(0,1)-vector
p - تنک	p-sparse
r - پوشش-آزاد	r-cover-free
t - وزن	t-weight
آزمون گروهی	group testing
الگوریتم احتمالاتی	probabilistic algorithm
الگوریتم انطباقی	adaptive algorithm
الگوریتم ترتیبی	sequential algorithm
الگوریتم چند مرحله‌ای	multi-stage algorithm
الگوریتم غیرانطباقی	nonadaptive algorithm
الگوریتم قطعی	deterministic algorithm
بردار وقوع	incidence vector
پرسش	query
تطابق	matching
تنک	sparse
چهاروجهی	tetrahedra
خوشه	clique
ستاره	star
شش ضلعی	hexagon
صفحه تصویری	projective plane

گراف پنهان .... hidden graph

گراف تصادفی .... random graph

ماتریس مجاورت .... adjacency matrix

مثبت .... positive

مجموعه‌ی زمینه .... ground set

مرز .... boundary

مقدار ویژه .... eigenvalue

منفرد .... isolated

منفی .... negative

نمایه

Surname: Sharifi Name: Elahe

Title: Learning a Hidden Graph

Supervisor: Dr. Meysam Alishahi
Advisor: Reza Musawi

Degree: Master of Science Subject: Applied Mathematics
Field: Graph theory and Combinatorics

Shahrood University of Technology Faculty of Mathematical Sciences
Date: July 2013 Number of pages: 83

Keywords: group testing, nonadaptive algorithm, genome sequencing

Abstract

In this thesis, we consider the problem of learning a hidden graph from a family of graphs on n vertices in a model where the only allowed operation is to query whether a set of vertices induces an edge. Our objective is to estimate the minimum number of queries needed to identify a hidden graph. Toward this end, we apply nonadaptive algorithms generally and study family of matchings, family of stars and family of cliques. We further describe some bounds that apply to general graphs.

Shahrood University of Technology
Faculty of Mathematical Sciences

Applied Mathematics

Learning a Hidden Graph

by

Elahe Sharifi

Supervisor

Dr. Meysam Alishahi

Advisor

Reza Musawi

Dissertation Submitted in Partial
Fulfillment of The Requirements For The
Degree of Master of Science in

July 2013