دانشگاه صنعتی شریف
دانشکده مهندسی کامپیوتر

عنوان

گزارش پروژه درسی

نام استاد درس

دکتر علی شریفی زارچی

اعضای گروه

حمیدرضا علیپور
رومینا نوبهاری
اشکان مجیدی
علیرضا کاظمینی

نام درس

آمار و احتمال مهندسی

نیم‌سال اول ۱۴۰۱-۱۴۰۲

۱ مقدمه

۲ ز هرچه رنگ تعلّق پذیرد آزاد است

پاسخ تئوری ۱. در این پرسش با این فرض پیش می رویم که هر فرد با تمام افراد هم خوشه با خودش هم سلیقه باشد(و به عبارتی دیگر در گراف حاصل به یکدیگر یال داشته باشند.) با توجه به این فرض که در هر خوشه سلیقه ها تماما یکسان است می توانیم نتیجه بگیریم که هر فرد در دقیقا یک خوشه عضو است. حال اگر هر رئوس هر یک از خوشه ها را با یک رنگ خاص رنگ کنیم، در گراف حاصل دو راس به یکدیگر وصلند اگر و فقط اگر همرنگ باشند.

در ماتریس مجاورت حاصل فرض می کنیم که اعداد روی قطر اصلی تماما ۱ باشند(یعنی فرض می کنیم که هر فرد با خودش هم سلیقه است) در آن صورت طبق توضیحات داده شده برای هر دو فرد هم سلیقه و یا به عبارت دیگر هم رنگ تمامی ستون های مربوط به هر یک از آن ها کاملا مثل هم کنید. برای مثال فرض کنید که افراد ۱ و ۳ و ۴ با هم و افراد ۲ و ۷ همچنین افراد ۵ و ۶ نیز با یکدیگر هم سلیقه باشند در آن صورت ماتریس مجاورت این افراد به شکل زیر میشود:

c۱	c۲	c۱	c۱	c۳	c۳	c۲
۱	۰	۱	۱	۰	۰	۰
۰	۱	۰	۰	۰	۰	۱
۱	۰	۱	۱	۰	۰	۰
۱	۰	۱	۱	۰	۰	۰
۰	۰	۰	۰	۱	۱	۰
۰	۰	۰	۰	۱	۱	۰
۰	۱	۰	۰	۰	۰	۱

پاسخ تئوری ۲. با توجه به صورت سوال این قسمت و قسمت بعدی و همچنین شکل ۳ با این فرض جلو می رویم که تمامی افراد عضو یک خوشه به یکدیگر یال دارند و تنها تفاوت این قسمت با قسمت قبلی این است که یک فرد می تواند همزمان در چند خوشه حضور داشته باشد.اگر این خوشه ها را به چشم نمودار ون نگاه کنیم می توانیم تمامی رئوس را دسته بندی کنیم و برای رئوس هر دسته یک رنگ در نظر بگیریم. برای مثال اگر دو خوشه A و B داشته باشیم می توانیم در نظر بگیریم که گرافمان به ۳ دسته A و B و $A\cap B$ افراز شده است و هر یک از این قسمت ها رنگ مخصوص به خود را دارند.

با این مدل سازی، در ماتریس مجاورت همانند بخش قبل هر دو ستون مربوط به هر دو راس همرنگ همچنان دقیقا مثل هم هستند، و همچنین اگر d راسی باشد که در خوشه های A_1 تا A_k حضور داشته باشد در آن صورت مجموعه ۱ های موجود در ستون مربوط به راس d برابر است با اجتماع ۱ های ستون های مربوط به رئوسی که تنها در یکی از خوشه های A_1 تا A_k قرار دارند.

برای مثال فرض کنید که رئوس ۱ و ۳ و ۴ تنها در خوشه A و رئوس ۲ و ۷ تنها در خوشه B و رئوس ۵ و ۶ هم در خوشه A و هم در خوشه B قرار داشته باشند. در آن صورت ماتریس مجاورت این رئوس به شکل زیر می شود:

c۱	c۲	c۱	c۱	c۳	c۳	c۲
۱	۰	۱	۱	۱	۱	۰
۰	۱	۰	۰	۱	۱	۱
۱	۰	۱	۱	۱	۱	۰
۱	۰	۱	۱	۱	۱	۰
۱	۱	۱	۱	۱	۱	۱
۱	۱	۱	۱	۱	۱	۱
۰	۱	۰	۰	۱	۱	۱

پاسخ تئوری ۳. فرض کنید این دو فرد در ژانر های ۱ تا k مشترک باشند در آن صورت احتمال اینکه با یکدیگر هم سلیقه باشند برابر است با ۱ منهای احتمال اینکه با یک دیگر هم سلیقه نباشند. احتمال اینکه این دو فرد در با هم هم سلیقه نشوند، با توجه به مستقل بودن این احتمالات، برابر است با احتمال اینکه در هیچ یک از خوشه ها با یکدیگر هم سلیقه نشوند که این احتمال هم برابر است با $1 - p_i$ پس نتیجه می گیریم که احتمال هم سلیقگی این دو فرد برابر است با:

$$1 - \pi_{i=1}^{k}(1 - p_i)$$

اگر دو نفر در جوامع زیادی عضو باشند و یا به عبارتی دیگر در خوشه های بیشتری قرار داشته باشند در آن صورت احتمال اینکه تعداد خوشه های مشترک شان بیشتر باشد بالاتر می رود(به عبارتی k بزرگتر می شود) پس با توجه به اینکه $1 - p_i$ عددی بین ۰ تا ۱ است پس حاصل $\pi_{i=1}^{k}(1 - p_i)$ کوچکتر و در نتیجه احتمال هم سلیقگی آن ها بزرگتر می شود.

نقص این مدل سازی این است که تفاوتی بین احتمال هم سلیقگی دو فرد که عضو یک خوشه هستند قائل نمی شود. به عبارتی دیگر p_i ها نباید از یکدیگر مستقل باشند.

برای مثال فرض کنید که دو فرد هر کدام در $\frac{M}{2}$ خوشه عضو باشند و تنها اشتراک این خوشه ها ژانر ۱ باشد. در نتیجه احتمال هم سلیقگی این دو فرد برابر با p_1 است.

حال دو نفر دیگر را در نظر بگیرید که تنها به ژانر ۱ علاقه داشته باشند. احتمال هم سلیقگی این دو فرد هم برابر با p_1 است، در حالی که از نظر منطقی سلایق دو نفر اول بسیار متفاوت تر از سلایق دو نفر دوم است و احتمال هم سلیقگی آن ها نباید با هم برابر باشد.

پاسخ تئوری ۴. مشابه بخش قبل اگر متمم احتمال هم سلیقگی آن ها را حساب کنیم نتیجه می گیریم که اگر این دو فرد در خوشه های ۱ تا c مشترک باشند در آن صورت احتمال هم سلیقگی آن ها برابر است با:

$$1 - \pi_{i=1}^{c}(e^{-F_{ui}F_{vi}})$$

با توجه به فرمول بالا چون F_{vi} ها اعدادی مثبت هستند پس هر یک از توان های e عددی بین ۰ تا ۱ است پس با زیاد تر شدن c، حاصل ضرب کاهش و در نتیجه احتمال هم سلیقگی افزایش پیدا می کند.

پاسخ تئوری ۵. فرض کنید که محل تقاطع سطر u و ستون v در ماتریس A برابر با d_{uv} باشد.

$$P[A|F] = \pi_{u,v}(P[d_{uv}|F])$$

داریم که d_{uv} به احتمال $1-\pi_{i=1}^{c}(e^{-F_{ui}F_{vi}})$ برابر با ۱ و به احتمال $\pi_{i=1}^{c}(e^{-F_{ui}F_{vi}})$ برابر با ۰ است پس با اندکی محاسبه نتیجه می گیریم که:

$$\pi_{u,v}(P[d_{uv}|F]) = \pi_{u,v}(d_{uv} + (-1)^{d_{uv}}\pi_{i=1}^{c}(e^{-F_{ui}F_{vi}}))$$

و لگاریتم این عبارت می شود:

$$\sum_{u,v} ln(d_{uv} + (-1)^{d_{uv}}\pi_{i=1}^{c}(e^{-F_{ui}F_{vi}}))$$

پاسخ تئوری ۶. اگر l(F) را به صورت تابعی بر حسب F_{ui} ها ببینیم و بقیه F_{vi} ها را ثابت فرض کنیم در آن صورت حرکت در راستای $\nabla_{F_{u,:}} l(F)$ مقدار این تابع را افزایش می دهد و با تکرار این روند به تعداد زیاد می توان به حالتی رسید که مقدار این گرادیان بسیار کم و نزدیک به ۰ باشد و بدین ترتیب مقدار ماکسیمم تابع بدست می آید.

پاسخ تئوری ۷. با استفاده از رابطه زیر هر یک از مولفه های گرادیان مورد نظر حساب می شود

$$\frac{\partial(l(F))}{\partial(F_{ut})} = \sum_{v=1,v\neq u}^{n} \frac{-F_{vt}(-1)^{d_{uv}} e^{\sum_{i=1}^{c}(-F_{ui}F_{vi})}}{d_{uv}+(-1)^{d_{uv}} e^{\sum_{i=1}^{c}(-F_{ui}F_{vi})}}$$

پاسخ شبیه‌سازی ۱. با توجه به پرسش های تئوری قبلی و فرمول هایی که برای likelihood و گرادیان بدست آوردیم کد مربوط به این بخش را کامل می کنیم و برای تست آن هم ماتریس A را مطابق روشی که در صفحه ۵ و ۶ داک پروژه گفته شده می سازیم و آن را به تابع train پاس میدهیم تا ماتریس F را خروجی دهد.
لگاریتم likelihood ماتریس F ابتدا از ۶۰۹- شروع شده و در نهایت به ۳۷۰- می رسد و همانطور که انتظار داشتیم ماتریس F هم به درستی ساخته می شود و رئوس ۱ تا ۱۵ تنها در خوشه ۲ و رئوس ۲۵ تا ۴۰ تنها در خوشه ۱ و رئوس ۱۵ تا ۲۵ در هر دو خوشه قرار می گیرند.

پاسخ تئوری ۸. احتمال اینکه درایه d_ij برابر k باشد را با استفاده از توزیع پوآسون با پارامتر $\mu = F_{i,:}.F_{j,:}^{T}$ می سازیم.
پس با داشتن ماتریس F احتمال اینکه ماتریس A ساخته شود برابر است با:

$$P[A|F] = \pi_{i,j}(P[d_{ij}|F]) = \pi_{i,j}(\frac{e^{-F_{i,:}.F_{j,:}^{T}}(F_{i,:}.F_{j,:}^{T})^{d_{ij}}}{d_{ij}!})$$

و لوگرایتم این عبارت برابر است با

$$log(P[A|F]) = \sum_{ij} log(\frac{e^{-\mu}\mu^{d_{ij}}}{d_{ij}}) = \sum_{ij}(-\mu + log(\frac{1}{d_{ij}}) + d_{ij}log(\mu))$$

با این تعاریف هر یک از مولفه های گرادیان $\nabla_{F_{u,:}} l(F)$ برابر می شود با

$$\frac{\partial(l(F))}{\partial(F_{ut})} = \sum_{v=1,v\neq u}^{n}(-F_{tv} + \frac{d_{uv}F_{tv}}{-F_{t,:}.F_{v,:}^{T}})$$

---

۳ گناه بخت پریشان و دست کوته ماست!

پاسخ تئوری ۹. ابتدا I_{ij} را مطابق زیر تغریف می‌کنیم.

$$I_{ij} = \begin{cases} 1, & \text{if } z_0[i] = z_0[j] \\ 0, & \text{otherwise} \end{cases}$$

حال اگر نحوه خوشه بندی یا همان z_0 در دسترس‌مان باشد آنگاه به وضوح داریم:

$$\begin{cases} A_{ij} \sim bernoulli(p), & \text{if } I_{ij} = 1 \\ A_{ij} \sim bernoulli(q), & \text{otherwise} \end{cases}$$

اما اگر z_0 را نداشته باشیم آنگاه

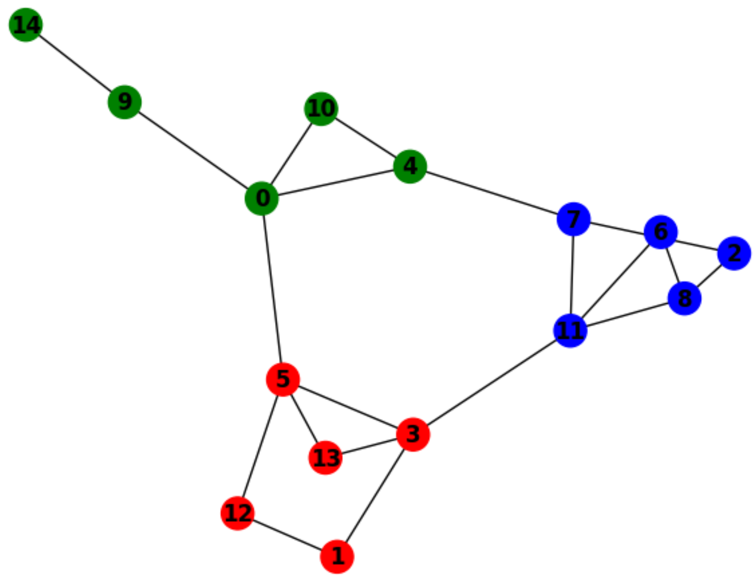
$$\begin{aligned} p(A_{ij} = 1) &= P(A_{ij} = 1 | I_{ij} = 1) p(I_{ij} = 1) + p(A_{ij} = 1 | I_{ij} = 0) p(I_{ij} = 0) \\ &= p.\frac{1}{3} + q.\frac{2}{3} \end{aligned}$$

$$\rightarrow A_{ij} \sim bernoulli(\frac{p+2q}{3})$$

پاسخ تئوری ۱۰. میدانیم اگر i و j هم‌سلیقه باشند آنگاه $A_{ij} = A_{ji} = 1$ پس این دو درایه به وضوح مستقل نیستند همچنین درایه های روی قطر اصلی اهمیت چندانی ندارند و می‌توانیم آن ها را ۱ فرض کنیم.

پاسخ شبیه‌سازی ۲. همانطور که در پرسش تئوری نه توضیح داده شد با استفاده از توزیع برنولی این ۱۰ نمونه را ساختیم. در واقع از تابع np.random.binomial استفاده کردیم. این تابع توزیع binomial را در اختیار ما قرار می‌دهد. با توجه به این که می‌دانیم اگر در توزیع binomial اندازه را برار با یک قرار دهیم توزیعbernoulli خواهیم داشت پس این تابع به ما در ساخت ماتریس کمک می‌کند.

پاسخ شبیه‌سازی ۳. با استفاده از کتابخانه networkx این گراف را کشیدیم و خوشه های آن را رنگ کردیم.



شکل ۱: گراف هم‌سلیقگی افراد

پاسخ شبیه‌سازی ٤. همانطور که خواسته شده این تابع دو بردار ورودی میگیرد و با یک پیمایش ساده روی آن ها فاصله همینگ را برمی‌گرداند.

پاسخ شبیه‌سازی ٥. حال از بین تمامی نام‌گذاری های ممکن برای خوشه‌های یکی از بردار های ورودی، آن را که فاصله همینگ کمتری دارد را انتخاب می‌کنیم. تمامی جایگشت های ممکن را با استفاده از کتابخانه itertools و تابع permutaions می‌یابیم.

پاسخ تئوری ۱۱. همانطور که گفته شد درایه های زیر قطر اصلی از هم مستقل هستند. پس $\frac{n(n-1)}{2}$ درایه مستقل از هم داریم.

پاسخ تئوری ۱۲. متغیر های زیر را تعریف می‌کنیم:

	A_{ij}	I_{ij}
a	1	1
b	0	1
c	1	0
d	0	0

برای مثال a تعداد درایه هایی در زیر قطر است که $A_{ij} = 1$ و هم سلیقه هستند.

$$L(z) = p(A|Z) = \Pi p(A_{ij}|z) = p^a(1-p)^b q^c (1-q)^d$$

پاسخ تئوری ۱۳.

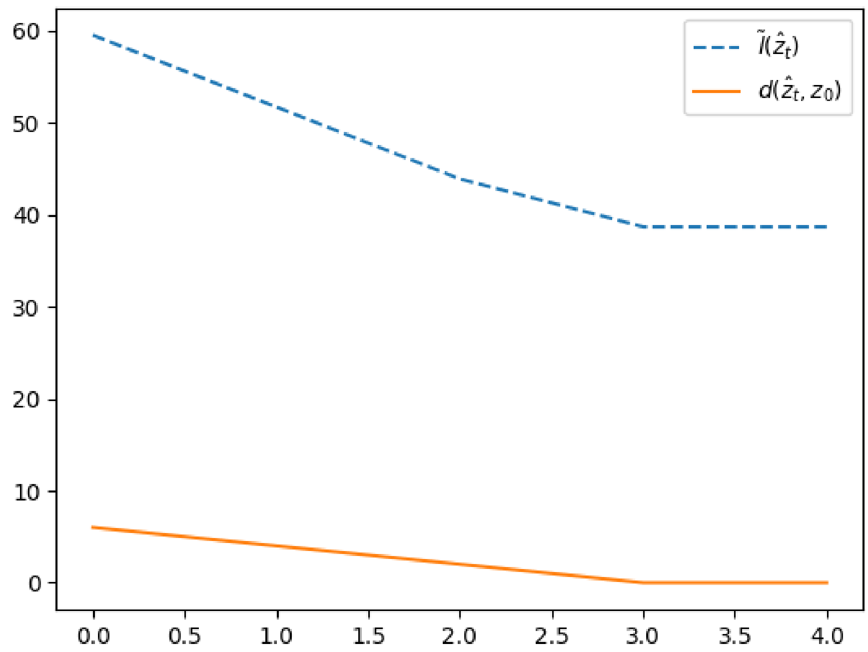
$$log(L(z)) = alog(p) + blog(1-p) + clog(q) + dlog(1-q)$$

پاسخ شبیه‌سازی ۶. مطابق با روش گفته شده ابتدا a،b،cوd را حساب می‌کنیم. سپس تخمین بیشینه درست‌نمایی آن را در تابع mle محاسبه میکنیم و سپس در ادمه log_mle_hat محاسبه می‌شود.

پاسخ شبیه‌سازی ۷. مطابق با روش خواسته شده الگوریتم پیاده سازی شده و خط نارنجی فاصله همینگ کمینه است و همانطور که مشاهده می‌کنید در حالت $T = 5$ این فاصله برابر با صفر شده به معنای این که به خوشه‌بندی درست رسیده‌ایم. دقت کنید که طبیعتا هرچه T عددی بزرگتر باشد تخمین های دقیق تری بدست خواهیم آورد اما در شبیه‌سازی های ما T برابر با ۵ فرض شده.

پاسخ شبیه‌سازی ۸. نتایج به این صورت هستند که ۷ تا از بردار های تخمین زده شده درست هستند(فاصله همینگ کمینه آن‌ها صفر هست). اما ۳ تای دیگر به این صورت نیستند. یکی از این ۳ تا با افزایش T تخمین درستی میدهد اما دو خروجی دیگر که فاصله همینگ آن‌ها ۴ هست با افزایش T نیز به بردار خواسته شده نمیرسند. خروجی ها به صورت زیر هستند:

Estimation: [1, 3, 2, 3, 1, 3, 2, 2, 2, 1, 1, 2, 3, 3, 1] ,min_hamming: 0
Estimation: [2, 1, 3, 1, 2, 1, 3, 3, 3, 2, 2, 3, 1, 1, 2] ,min_hamming: 0
Estimation: [3, 1, 2, 1, 3, 1, 2, 2, 2, 3, 3, 2, 1, 1, 3] ,min_hamming: 0
Estimation: [3, 2, 1, 2, 3, 2, 3, 1, 1, 1, 3, 3, 2, 2, 1] ,min_hamming: 4
Estimation: [2, 1, 3, 1, 2, 1, 3, 3, 3, 2, 2, 3, 1, 1, 2] ,min_hamming: 0
Estimation: [2, 1, 3, 1, 2, 1, 2, 3, 3, 3, 2, 2, 1, 1, 3] ,min_hamming: 4
Estimation: [2, 3, 1, 3, 2, 3, 1, 1, 1, 2, 2, 1, 3, 3, 2] ,min_hamming: 0
Estimation: [3, 2, 1, 2, 3, 2, 1, 1, 1, 3, 3, 1, 2, 2, 3] ,min_hamming: 0
Estimation: [2, 3, 1, 3, 2, 3, 1, 1, 1, 2, 2, 1, 3, 3, 2] ,min_hamming: 0
Estimation: [2, 2, 3, 1, 2, 1, 3, 3, 3, 2, 2, 3, 1, 1, 1] ,min_hamming: 2



شکل ۲: نمودار تخمین بیشینه درست‌نمایی و همینگ کمینه بر حسب t

پاسخ شبیه‌سازی ۹. همانطور که در خروجی کد شبیه‌سازی مشخص شده، حالاتی وجود دارند که $\tilde{l}(\hat{z}_T) = \tilde{l}(z_0)$ و در تعدادی از آن ها فاصله همینگ کمینه با z_0 برابر با صفر است و در تعدادی دیگر نه.

پاسخ شبیه‌سازی ۱۰. همانطور که مشاهده می‌کنید در بسیاری از آن ها فاصله همینگ کمینه با z_0 برابر با صفر است به این معنی که درست تخمین زده‌ایم.

پاسخ شبیه‌سازی ۱۱. در ابتدا ده نمونه ماتریس ساخته بودیم و برای این سوال از دو نمونه آن‌ها بهره می‌گیریم. در بهترین نتیجه به تخمین درستی از z_0 رسیده ایم اما اگر به تخمین هایی که از روی ماتریس دوم ($A[2]$) انجام شده نگاه کنیم متوجه اختلاف کم اما مشهودی می‌شویم که این موضوع با بزرگ کردن T رفع خواهد شد.

---

۴ او نه خیال است و نه طیف

پاسخ تئوری ۱۴. متغیر A_{ij} از توزیع برنولی است که در صورتی که i و j در یک دسته قرار داشته باشند احتمال آن برابر p و در غیر این صورت احتمال آن برابر q است.
تابع توزیع احتمال متغیر برنولی با احتمال p می شود:

$$pmf = \begin{cases} p & A_{ij} = 1 \\ 1-p & A_{ij} = 0 \end{cases}$$

پاسخ تئوری ۱۵. اگر i و j در یک خوشه باشند به احتمال p به یکدیگر وصلند پس $E[A_{ij}] = p$ و در صورتی این دو راس در یک خوشه نباشد مقدار این امید ریاضی برابر q است. پس درایه های ماتریس W با اعداد p و q پر می شود.
برای مثال فرض کنید که رئوس ۱ و ۲ در یک خوشه و رئوس ۳ و ۴ در یک خوشه دیگر قرار داشته باشند در آن صورت W به شکل زیر می شود

p	p	q	q
p	p	q	q
q	q	p	p
q	q	p	p

پاسخ تئوری ۱۶. ماتریس W در این حالت همان حالت خاصی است که در پرسش تئوری قبلی کشیدیم. فرض کنید که مقدار ویژه ماتریس W برابر k باشد در آن صورت باید بدست بیاوریم که به ازای چه مقادیری از k دترمینان ماتریس زیر ۰ می شود:

p-k	p	q	q
p	p-k	q	q
q	q	p-k	p
q	q	p	p-k

دترمینان این ماتریس برابر است با

$$k^4 - 4k^3p + 4k^2p^2 - 4k^2q^2 = k^2(k-2p-2q)(k-2p+2q)$$

پس مقدار ویژه های W برابر هستند با ۰ و ۲p+۲q و ۲p-۲q.
با جایگذاری هر یک از این مقادیر و حل معادلات چند مجهولی بردار های ویژه برای هر یک از مقادیم ویژه بدست می آیند برای حالت k=۰

$$v = \begin{bmatrix} ۱ \\ -۱ \\ ۱ \\ -۱ \end{bmatrix}$$

برای حالت k=۲p+۲q

$$v = \begin{bmatrix} ۱ \\ ۱ \\ ۱ \\ ۱ \end{bmatrix}$$

برای حالت k=۲p-۲q

$$v = \begin{array}{|c|}\hline ۱ \\ \hline ۱ \\ \hline ۱- \\ \hline ۱- \\ \hline \end{array}$$

پاسخ تئوری ۱۷. اگر راس i از گراف عضو خوشه اول باشد سطر i ام را از نوع یک و اگر عضو خوشه ۲ باشد این سطر را از نوع ۲ در نظر می گیریم. با توجه به تعریف ماتریس W و روابطی که برای وجود یا عدم وجود یک یال بین دو راس یک خوشه یا دو راس از دو خوشه مختلف وجود دارد نتیجه می گیریم که هر نوع از سطر ها در ماتریس W دقیقا مثل هم پر می شوند.

پس نتیجه می گیریم که مرتبه این ماتریس ۲ است یعنی حداکثر دو سطر مستقل خطی دارد. از جبر خطی می دانیم که ماتریس با مرتبه n حداکثر n مقدار ویژه ناصفر دارد پس ثابت کردیم که W حداکثر دو مقدار ویژه ناصفر دارد.

حال با توجه به این فرض که هر یک از خوشه ها دقیقا $\frac{n}{2}$ تا عضو دارد ثابت می کنیم که مقادیر $\frac{n(p+q)}{2}$ و $\frac{n(p-q)}{2}$ مقادیر ویژه ماتریس هستند و بردار های ویژه متناظر با هر یک از آن ها را هم حساب می کنیم. برای این کار باید معادله زیر را حل کنیم.

$$WV = \lambda V$$

در حالتی که مقدار ویژه ما برابر با $\frac{n(p+q)}{2}$ باشد، اگر تمامی مولفه های V را برابر ۱ قرار دهیم تساوی بالا برقرار می شود(در این حالت دو طرف تساوی بالا برابر با برداری با مولفه های $\frac{n(p+q)}{2}$ است)

و در حالتی که مقدار ویژه برابر با $\frac{n(p-q)}{2}$ باشد می توان بردار V را بدین صورت ساخت:

اگر سطر i ام در ماتریس W از نوع یک باشد مولفه ی i ام بردار را برابر ۱ و در غیر این صورت برابر با ۱- قرار می دهیم. با توجه به این روش ساختن V، تمامی درایه های حاصل ضرب WV برابر با $\frac{n(p-q)}{2}$ می شود پس این بردار در تساوی بالا صدق می کند و نتیجه می گیریم که این همان بردار متناظر با مقدار ویژه $\frac{n(p-q)}{2}$ است.

پاسخ تئوری ۱۸. با توجه به این فرض که تعداد اعضای دو خوشه با یکدیگر برابرند نتیجه می گیریم که در هر سطر از W دقیقا $\frac{n}{2}$ تا p و $\frac{n}{2}$ تا q داریم پس به ازای هر i داریم که:

$$[D_W]_{i,i} = \sum_{j=1}^{n} W_{i,j} = \frac{n(p+q)}{2}$$

پس ماتریس D_W ماتریسی است که تمام درایه های قطر اصلی آن برابر $\frac{n(p+q)}{2}$ و بقیه درایه های آن برابر با ۰ است.

پاسخ تئوری ۱۹. معادله زیر را باید حل کنیم

$$\lambda V = L_W V = (D_W - W)V = D_W V - WV$$

با توجه به اینکه ماتریس D_W قطری است و مقدار تمام درایه های روی قطر اصلی آن برابر مقدار ثابت $\frac{n(p+q)}{2}$ است پس $D_W V = \frac{n(p+q)}{2} V$ پس تساوی بالا را می توان به شکل زیر نوشت

$$WV = (\frac{n(p+q)}{2} - \lambda)V$$

پس دوباره برگشتیم به مسئله پیدا کردن مقادیر ویژه W. ثابت کردیم که این ماتریس تنها ۳ مقدار ویژه ۰ و $\frac{n(p+q)}{2}$ و $\frac{n(p-q)}{2}$ را دارد پس λ تنها ۳ مقدار ۰ و $\frac{n(p+q)}{2}$ و nq را می تواند داشته باشد و از طرفی به ازای هر یک از این مقادیر ویژه، بردار ویژه متناظر با آن نیز از بردار های ویژه ماتریس W که در پرسش تئوری ۱۸ آن ها را حساب کردیم، محاسبه می شود.

پاسخ تئوری ۲۰. دو مقدار ویژه کوچکتر ماتریس L_W برابر با ۰ و nq (با توجه به اینکه q از p کوچکتر است) هستند.
پس اگر این مقادیر را به جای λ در تساوی $WV = (\frac{n(p+q)}{2} - \lambda)V$ بگذاریم به دو بردار ویژه u_1 و u_2 می رسیم که تمام مولفه های u_1 برابر با ۱ است و در u_2 هم داریم که اگر سطر i ام در ماتریس W از نوع ۱ باشد، درایه i ام از بردار برابر ۱ و در غیر این صورت برابر -۱ است(اثبات در پرسش تئوری ۱۸)
حال اگر تابع گفته شده با استفاده از این دو بردار را بر روی رئوس گراف سوال ۱۶ اعمال کنیم رئوس ۱ و ۲ به نقطه (۱،۱) و رئوس ۳ و ۴ به نقطه (۱،-۱) map می شوند.
با توجه به اینکه در این نمایش رئوس در صفحه R^2 نقاط متناظر با رئوس ۱ و ۲ بسیار نزدیک به هم قرار گرفتند نتیجه می گیریم که این دو راس باید در یک خوشه باشند. استدلال مشابه برای رئوس ۳ و ۴ نیز برقرار است.

پاسخ شبیه‌سازی ۱۲. با استفاده از کتابخانه‌های numpy و `sklearn.cluster` توابع `A` و `W` را ساخته و میزان خطای الگوریتم برای ۲ ماتریس بصورت $A = 0.7$ و $B = 1$ میباشد.

پاسخ تئوری ۲۱.

پاسخ شبیه‌سازی ۱۳. با استفاده از n های 100 و 150 برابر $W = 1$ و برای $n = 100 -> A = 0.54$ و $n = 150 -> A = 0.51$ میباشد

پاسخ تئوری ۲۲. در اینجا هم مشابه الگوریتم با k=۲ عمل می کنیم. ابتدا ماتریس مجاورت رئوس را میسازیم. سپس ماتریس لاپلاسین آن را بدست می آوریم(روش ساختن ماتریس لاپلاسین بدین صورت است که مشابه روش گفته شده در پرسش تئوری ۱۸ ماتریس D_W را میسازیم و $L_W = D_W - W$ می شود ماتریس لاپلاسین).
حال k مقدار ویژه کوچکتر ماتریس لاپلاسین را حساب می کنیم و به ازای آن ها بردار ویژه متناظر را بدست می آوریم. این k بردار را به ترتیب در کنار هم قرار می دهیم تا ماتریسی n در k تشکیل شود. راس i ام از گرافمان به سطر i ام از این ماتریس و در واقع به یک نقطه از فضای R^k متناظر می شود. با اعمال الگوریتم k-means روی این نقاط آن ها را به k خوشه افراز می کنیم.

پاسخ شبیه‌سازی ۱۴.
پاسخ شبیه‌سازی ۱۵.
پاسخ شبیه‌سازی ۱۶.

۵ من از دیار حبیبم نه از بلاد غریب

پاسخ تئوری ۲۳. جواب این مسئله برابر است با:

$$p^m(1-p)^{\binom{n}{2}-m}$$

دلیلش این است احتمال وجود m یالی که باید رسم شوند برابر p است و برای بقیه یال ها به احتمال ۱-p یال کشیده نمی شود و با توجه به استقلال این احتمالات، احتمال کل برابر است با حاصل ضرب این احتمالات که می شود همان عبارت بالا.

پاسخ تئوری ۲۴. در این پرسش می دانیم که دقیقا m یال داریم. پس در مجموع انتخاب m از تعداد کل یال ها حالت برای گرافمان وجود دارد. با توجه به اینکه احتمال وجود یا عدم وجود هر دو یالی برابر است نتیجه می گیریم که احتمال اینکه دقیقا m یال مطلوب را انتخاب کنیم برابر است با:

$$\frac{1}{\binom{\binom{n}{2}}{m}}$$

پاسخ تئوری ۲۵. اگر منظور این سوال این است که ۲۰ درصد یال ها در این گراف کشیده شده باشند با فرض اینکه m یال داشته باشیم جواب برابر است با:

$$\binom{m}{\frac{m}{5}}p^{\frac{m}{5}}(1-p)^{\binom{n}{2}-\frac{m}{5}}$$

اما اگر منظور سوال این است که ۲۰ درصد از کل روابط ممکن بین رئوس گراف انتخاب می شود(چه وصل بودن و چه وصل نبودن) در آن صورت باید یک پنجم کل $\binom{n}{2}$ حالت رابطه بین رئوس راانتخاب کنیم و اگر یک یال کشیده شده انتخاب شده بود باید احتمال p برای آن برقرار و اگر یک یال کشیده نشده انتخاب شده بود باید احتمال ۱-p برای آن برقرار شود.
و به همین ترتیب برای یک یال کشیده شده انتخاب نشده باید احتمال ۱-p و برای یک یال کشیده نشده انتخاب نشده باید احتمال p برقرار شود.
فرض کنید a یال از یال ها کشیده شده و b یال از یال های کشیده نشده انتخاب شده باشند در آن صورت احتمال مورد نظر برای این حالت می شود:

$$p^{\binom{n}{2}-m-b+a}(1-p)^{m-a+b}$$

و باید این احتمالات را روی تمام انتخاب های روابط هم سلیقگی جمع بزنیم.

پاسخ شبیه‌سازی ۱۷. ابتدا یک گراف با ویژگی های خواسته شده شبیه‌سازی می‌کنیم. هر یال از گراف به احتمال p رسم می‌شود(از توزیع برنولی تبعیت می‌کند). با استفاده از تابع bernoulli.rvs(p,size) درون scipy.stats این کار را انجام می‌دهیم. میانگین مجموع تعداد یال ها را محاسبه می‌کنیم که همانطور که مشاهده می‌کنید برابر است با ۱۶۹۰/۲ و خطای m ، ۷۷/۴۹۴ درصد است که عدد بالایی می‌باشد(البته توجه کنید که توابع بهتری برای این شبیه سازی در numpy و random موجود اند).

پاسخ تئوری ۲۶. هر یال به احتمال p کشیده می شود پس در کل امید ریاضی تعداد یال های گراف برابر است با $\binom{n}{2}p$ که برای اعداد این سوال برابر است با 1698.3

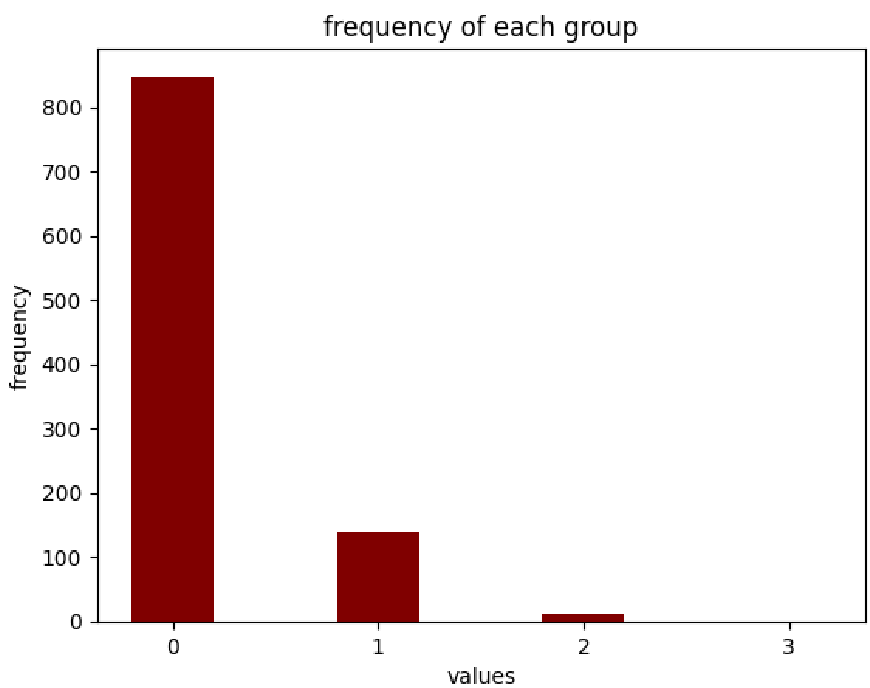
در کل برای تحلیل تعداد یال های رسم شده در گراف توجه کنید که کشیده شدن هر یال از توزیع برنولی با احتمال p پیروی می کند و تعداد کل یال ها برابر با جمع این توزیع های برنولی است پس تعداد کل یال ها از توزیع دو جمله ای است. در توزیع دو جمله ای می دانیم که امید ریاضی و واریانس به ترتیب برابر است با np و npq پس در اینجا امید ریاضی و واریانس می شود $\binom{n}{2}p$ و $\binom{n}{2}p(1-p)$

اگر خطای m کمتر از ۵ درصد باشد به این معناست که:

$$\frac{|\binom{n}{2}p - m|}{\binom{n}{2}p} \leq 5\% \rightarrow \frac{19}{20}\binom{n}{2}p \leq m \leq \frac{21}{20}\binom{n}{2}p \rightarrow 1613.385 \leq m \leq 1783.215$$

$$m \in [1613.385, 1783.215]$$

پاسخ شبیه‌سازی ۱۸. با توجه به قطعه کد نوشته شده میانگین تعداد افراد همرنگ برابر است با ۱۵۲ همچنین نمودار خواسته شدا را در زیر مشاهده می‌کنید.



پاسخ تئوری ۲۷. میانگین تعداد هم سلیقه های هر فرد می شود جمع تعداد هم سلیقه های همه ی افراد تقسیم بر n. از طرفی جمع تعداد هم سلیقه های همه افراد در واقع همان جمع درجات تمام رئوس گراف یا همان دو برابر تعداد یال های گراف است. از طرفی امید ریاضی تعداد رئوس گراف را در پرسش تئوری قبلی بدست آوردیم پس :

$$L = \frac{2m}{n} = \frac{2\binom{n}{2}p}{n} = \frac{n(n-1)p}{n} = p(n-1)$$

پاسخ تئوری ۲۸. در سوال قبل فرمول L را حساب کردیم و در اینجا مقدارش می شود 0.15984. پس اگر درجه یک راس بیشتر از ۰ بود آن راس همرنگ و در غیر این صورت رسوا است.

احتمال اینکه یک فرد همرنگ باشد برابر است با ۱ منهای احتمال اینکه رسوا باشد یعنی به هیچ راس دیگری یال نداشته باشد که این احتمال هم برابر است با $(1-p)^{999}$. پس احتمال این که یک شخص همرنگ باشد برابر است با $1-(1-p)^{999}$

فرض کنید که X_i برابر با متغیر تصادفی همرنگ بودن فرد iام و X برابر با تعداد افراد همرنگ باشد در آن صورت:

$$E[X] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = n(1-(1-p)^{999}) = 1000 * 0.148 = 148$$

پاسخ شبیه‌سازی ۱۹. با استفاده از تابع random.choices گرافمان را به صورت تصادفی می سازیم و با پیمایش روی رئوس گراف تعداد روابط هم سلیقگی با خاصیت تراگذری و زنجیره ای را می شماریم و از اعداد بدست آمده میانگین می گیریم. به طور میانگین 4487.8 خاصیت تراگذری و 1335162.9 رابطه هم سلیقگی با خاصیت زنجیره ای وجود دارد.

پاسخ تئوری ۲۹. فرض کنید که X_i متغیر تصادفی برنولی باشد که ۱ است اگر رئوس ۳ تایی i ام خاصیت تراگذری داشته باشند و ۰ است اگر نداشته باشند(در مجموع $\binom{n}{3}$ تا ۳ تایی از رئوس داریم)

احتمال ۱ بودن هر یک از این متغیر های تصادفی برابر با p^3 است و تعداد کل ۳ تایی های رئوس با خاصیت تراگذری می شود جمع این متغیر های تصادفی پس با توجه به خاصیت خطی بودن امید ریاضی

$$E[X] = E[\sum_{i=1}^{\binom{n}{3}} X_i] = \sum_{i=1}^{\binom{n}{3}} E[X_i] = \binom{n}{3} p^3$$

برای حساب کردن تعداد روابط هم سلیقگی با خاصیت زنجیره ای هم مانند قسمت قبل متغیر های تصادفی تعیین می کنیم با این تفاوت که اینبار احتمال یک بودن هر یک از آن ها برابر است با $3p^2(1-p)$

و حاصل جمع آن ها می شود:

$$E[X] = E[\sum_{i=1}^{\binom{n}{3}} X_i] = \sum_{i=1}^{\binom{n}{3}} E[X_i] = 3\binom{n}{3} p^2(1-p)$$

پاسخ تئوری ۳۰. اگر در هر گروه ۳ تایی از رئوس هر راس حداقل به یک راس دیگر وصل باشد در آن صورت تنها دو حالت برای این ۳ راس وجود دارد. حالت اول اینکه خاصیت تراگذری و حالت دوم اینکه خاصیت زنجیره ای داشته باشند.

در پرسش تئوری قبل امید ریاضی تعداد ۳ تایی ها با خاصیت تراگذری و زنجیره ای را حساب کردیم پس در اینجا چون کل فضای احتمالمان این است که هر ۳ تایی یا خاصیت زنجیره ای دارد یا تراگذری جواب برابر است با:

$$\frac{\binom{n}{3}p^3}{\binom{n}{3}p^3 + 3\binom{n}{3}p^2(1-p)} = \frac{p}{p+3(1-p)} = \frac{p}{3-2p}$$

پاسخ شبیه‌سازی ۲۰. در هر مرحله برای هر راس از گراف، رئوس مجاور با آن را در یه لیست قرار می دهیم تا تعداد یال های بین اعضای این لیست را بشماریم و این اعداد را با هم جمع می زنیم و در نهایت تقسیم بر تعداد رئوس می کنیم.

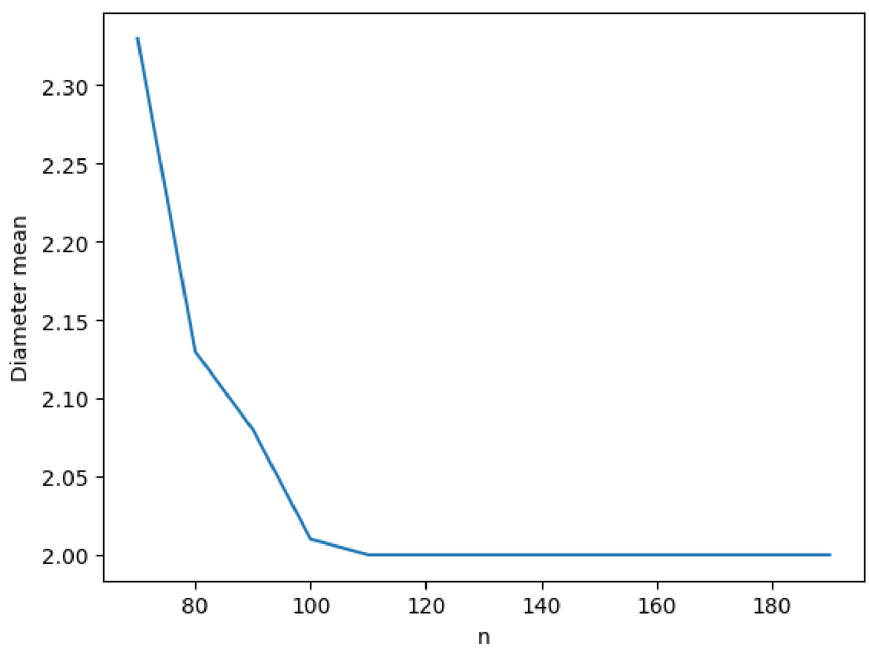
۱۰۰ بار گراف را می سازیم و از اعدادی که در هر بار بدست آمده میانگین می گیریم تا متوسط تعداد روابط هم سلیقگی میان همسایه های یک راس بدست بیاید. جواب تقریبا برابر با 0.013 است.

پاسخ تئوری ۳۱. یک راس دلخواه را در نظر بگیرید. ما متوسط تعداد مثلث های شامل این راس را می خواهیم. به $\binom{n-1}{2}$ طریق میتوان ۲ راس دیگر انتخاب کرد و احتمال اینکه این ۳ راس تشکیل یک مثلث دهند برابر با p^3 است. پس جواب می شود $\binom{n-1}{2}p^3$

پاسخ شبیه‌سازی ۲۱. با توجه به این که گراف حاصل به احتمال بسیار بالایی همبند نیست پس عملا محاسبه چنین میانگینی معنا دار نیست. از آنجاییکه گراف امکان دارد ناهمبند باشد، آنرا برای حالت همبند باشد حساب میکنیم.

پاسخ شبیه‌سازی ۲۲. با استفاده از تابع diameter در networkx قطر گراف را در ۱۰۰ نمونه محاسبه میکنیم. همانطور که مشاهده میکنید میانگین برابر با ۲/۸۱ شده است.

پاسخ شبیه‌سازی ۲۳. با انجام شبیه‌سازی خواسته شده نمودار زیر را مشاهده خواهیم کرد. همانطور در نمودار



شکل ۳: نمودار میانگین قطر گراف بر حسب n

مشخص است با زیاد شدن n با توجه به این که تعداد یال ها زیاد میشود پس قطر گراف بسیار کوچک خواهد شد و به ۲ میل می‌کند.

پاسخ تئوری ۳۲. هر یک از ۲-n راس دیگر بجز u و v به احتمال $1-p^2$ به حداقل یکی از این دو راس وصل نیست و اگر $I_{uv}=1$ این احتمال باید برای تمام آن ها برقرار شود و پس با توجه به مستقل بودن این احتمالات جواب می شود:

$$(1-p^2)^{n-2}$$

درواقع از خطی بودن امید ریاضی استفاده کرده این و Indicator Variable هایمان را وجود مثلث هم‌سلیقگی بین ۳ راس درنظر گرفته ایم.

پاسخ تئوری ۳۳. می دانیم که X_n در واقع جمع تمام I_{uv} هاست و با توجه به خاصیت خطی بودن امید ریاضی

$$E[X_n]=E[\sum_{uv}I_{uv}]=\sum_{uv}E[I_{uv}]=\binom{n}{2}(1-p^2)^{n-2}$$

پاسخ تئوری ۳۴. طبق نامساوی مارکف داریم

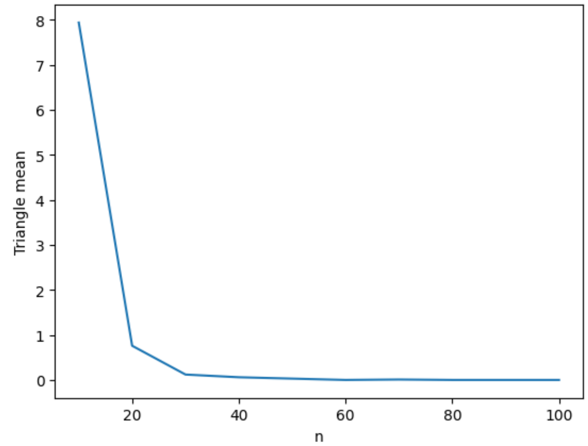
$$lim_{n\to\infty}p(1\leq X_n)\leq lim_{n\to\infty}E[X_n]=lim_{n\to\infty}\frac{n(n-1)}{2}(1-p^2)^{n-2}$$

$1-p^2$ کمتر از ۱ است و با بزرگ شدن n دارد به صورت نمایی به ۰ میل می کند در حالی که $\frac{n(n-1)}{2}$ با اردر چند جمله ای دارد رشد می کند پس چون رشد تابع نمایی سریع تر است، با میل کردن n به بی نهایت حاصل ضرب آن ها به ۰ میل می کند پس $p(1\leq X_n)$ کمتر مساوی ۰ می شود و چون احتمال همیشه عددی نامنفی است برابر با ۰ می شود.

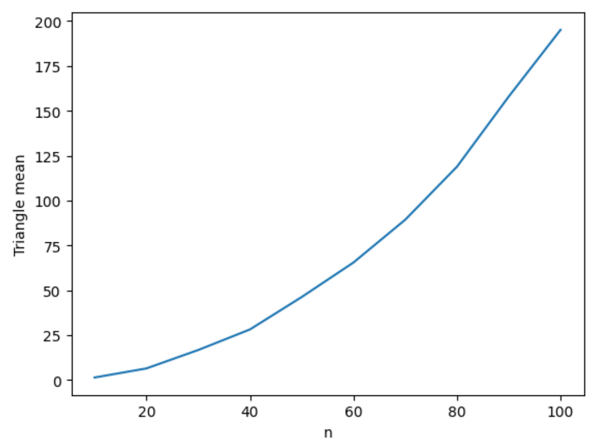
پاسخ تئوری ۳۵. ثابت کردیم که با زیاد شدن n احتمال اینکه X_n بیشتر مساوی ۱ باشد برابر با ۰ می شود پس $X_n=0$ یعنی هر دو راس باید حداقل یک همسایه مشترک داشته باشند پس قطر گراف حداکثر برابر با ۲ می شود.

پاسخ شبیه‌سازی ۲۴. همانطور که در کد مشاهده می‌کنید این میانگین برابر است با ۱۸۸/۷۵

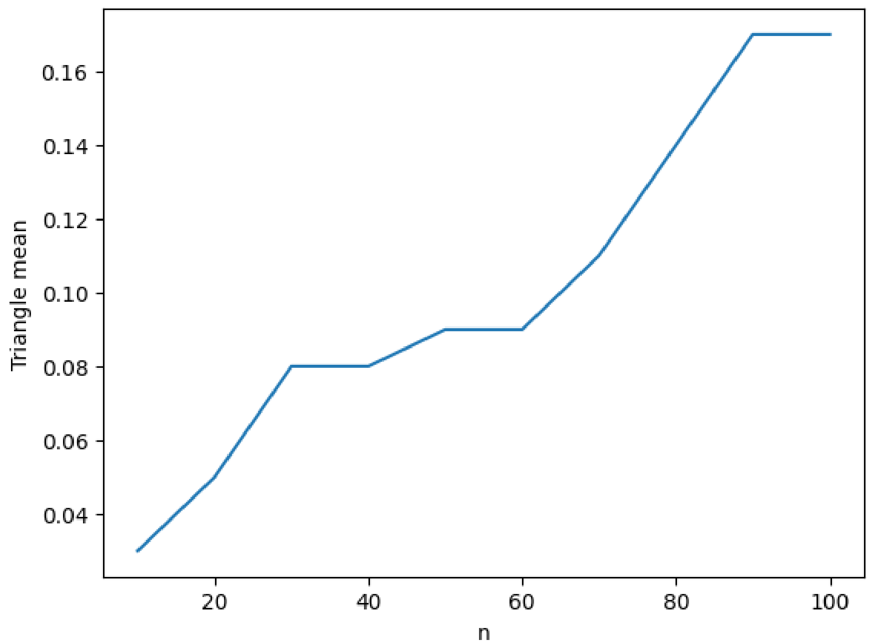
پاسخ شبیه‌سازی ۲۵. با توجه به این که احتمال بسیار کم است همانطور که می‌بینید به ۰ میل میکند.



پاسخ شبیه‌سازی ۲۶. به بینهایت میل میکند.



پاسخ شبیه‌سازی ۲۷. به بنظر باید با توجه به این cmf عددی ثابت باشد.



۶ سل المصانع رکبا تهیم فی الفلوات!

پاسخ تئوری ۳۶. برای حرکت از راس i به راس j اگر بین این دو راس، راس دیگری وجود داشته باشد، احتمال انتقال بین آنها با توجه به اینکه حرکت روی راس ها یکنواخت است، $\frac{1}{\mathrm{d}(i)}$ میشود. درصورتی هم که بین رئوس i و j راس دیگری نباشد، امکان انتقال از راس i به راس j وجود ندارد. پس برای احتمال انتقال از راس i به راس j داریم:

$$P_{i,j} = \begin{cases} \frac{1}{\mathrm{d}(i)} & i \texttt{ and } j \texttt{ are connected} \\ 0 & \texttt{otherwise} \end{cases}$$

پاسخ تئوری ۳۷. از آنجاییکه ماتریس $\mathrm{P} \in \mathbb{R}^{n\times n}$ راس‌های قرار گرفته بین هر ۲ راس $\mathrm{A} \in \mathbb{R}^{n\times n}$ را نشان میدهد، میتوان از هر $A_{i,j}$ برای چک کردن وصل بودن ۲ راس استفاده کرد. میدانیم هم که در ماتریس $\mathrm{D} \in \mathbb{R}^{n\times n}$ هم برای $D_{i,i} = \mathrm{d}(i)$ درجه راس مورد نظر را داریم، پس میتوان برای احتمال انتقال از راس i به j که ماتریس $\mathrm{P} \in \mathbb{R}^{n\times n}$ را تشکیل میدهد را بصورت زیر نوشت:

$$P_{i,j} = \frac{A_{i,j}}{D_{i,i}}$$

پاسخ تئوری ۳۸. در ضرب $\mathrm{P} \in \mathbb{R}^{n\times n} \times \mathrm{P} \in \mathbb{R}^{n\times n}$ وقتی سطر i را در ستون j ضرب میکنیم، برای راس‌هایی که هم به i و هم به j متصلند، از آنجاییکه درایه‌ی آن 0 نیست (چون به i و j متصلند)، $P^2_{i,j} = \frac{1}{\mathrm{d}(i)\mathrm{d}(j)}$. در صورتی هم که به یکی از رئوس i یا j وصل نباشد یا به هیچ کدام از رئوس i و j وصل نباشد، $P^2_{i,j} = 0$. پس برای P^2 داریم:

$$P^2_{i,j} = \begin{cases} \frac{1}{\mathrm{d}(i)\mathrm{d}(j)} & i \texttt{ and } j \texttt{ are connected} \\ 0 & \texttt{otherwise} \end{cases}$$

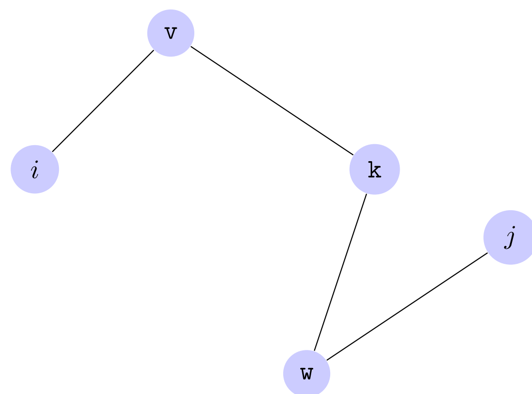
که ${P^2}_{i,j}$ احتمال انتقال از i به j با دو حرکت را نشان میدهد.

پاسخ تئوری ۳۹. مانند پرسش تئوری ۳۸، برای رسیدن از راس i به راس j در t مرحله، ۲ حالت زیر را داریم که در صورتیکه از i به j مسیری موجود بود، حاصل درایه $P_{i,j}$ پس از $\prod^t P^{n\times n}$ است. پس داریم:

$$P^{(t)}_{i,j} = \begin{cases} \frac{1}{\mathrm{d}(i)\mathrm{d}(j)}, & i \text{ and } j \text{ are connected after } \Pi^t P^{n\times n} \\ 0, & \text{otherwise } z \end{cases}$$

پاسخ تئوری ۴۰. با فرض اینکه یکی از مسیر هایی که باید $\mathtt{t}$ مرحله در آن طی شود که از i به j برسیم مانند مسیر زیر باشد و $\mathtt{v}$، $\mathtt{w}$ و $\mathtt{k}$ نیز بخش هایی از این مسیر باشد، احتمال طی کردن این مسیر برای رسیدن از i به j، برابر احتمال طی کردن مسیرهای از i برروی راس‌های بین i و $\mathtt{v}$ تا رسیدن به راس $\mathtt{v}$، سپس از $\mathtt{v}$ به $\mathtt{k}$، پس از آن از $\mathtt{k}$ به $\mathtt{w}$ و در نهایت طی کردن رئوس بین $\mathtt{w}$ و j است.
در ظی مسیر از i بسمت j، برای مثال در مرحله $\mathtt{v}$، انتخاب اینکه به $\mathtt{k}$ برود، $\frac{1}{\mathrm{d}(v)}$ است. در مرحله $\mathtt{k}$ نیز انتخاب اینکه به $\mathtt{w}$ برود، $\frac{1}{\mathrm{d}(k)}$ و از $\mathtt{w}$ به راس بعدی بسمت j، $\frac{1}{\mathrm{d}(w)}$ میباشد.
در طی مسیر از j به i نیز، برای رفتن از $\mathtt{w}$ به $\mathtt{k}$، $\frac{1}{\mathrm{d}(w)}$، برای رفتن از $\mathtt{k}$ به $\mathtt{v}$، $\frac{1}{\mathrm{d}(k)}$ و برای رفتن از $\mathtt{v}$ به راس بعدی بسمت

i، $\frac{1}{\mathrm{d}(v)}$ میباشد.
پس برای طی مسیرها از i به j و از j به i در هر دو حالت، احتمال حرکت روی راس های وسط برابر است. تنها تفاوت در این دو، انتخاب مسیر از راس i یا انتخاب مسیر از راس j است که بسته به اینکه چند راس به آن‌ها متصل باشد، به احتمال $\frac{1}{\mathrm{d}(i)}$ از راس i و $\frac{1}{\mathrm{d}(j)}$ از راس j مسیر مورد نظر طی خواهد شد.
پس تفاوت احتمال طی مسیر i–j از i به j یا از j به i، تنها به درجه‌ی رئوس i و j بستگی دارد.



پاسخ تئوری ۴۱. در صورتیکه کاربر i و j همسلیقه باشند، پس در خوشه‌ای که i و j حاضر هستند، تعداد یال‌های بیشتری قرار دارد. به همین علت، تعداد مسیرهای t‌تایی از i به j، بیشتر از تعداد مسیرهای t‌تایی از i به کاربر دیگری است که در خوشه‌ی ij قرار ندارد.
پس در صورت هم سلیقه بودن i و j، احتمال اینکه از i به j با طی t مرحله برسیم، بیشتر از آن است که از i به j در صورت هم سلیقه نبودن با طی t مرحله برویم چون در حالت هم سلیقه بودن مسیرهای بیشتری از i به j موجود میباشد.
با توجه به اینکه i و j همسلیقه هستند، اگر k یک فرد درون دسته i و j باشد، احتمال i به k و j به k بیشتر از زمانیست که k درون دسته‌ی i و j نباشد.

پاسخ تئوری ۴۲. از آنجاییکه برای تعریف اختلاف سلیقه دو کاربر داریم:

$$r_{i,j} = \sqrt{\sum_{k=1}^{n} \frac{(P_{ik}^{(t)} - P_{jk}^{(t)})^2}{\mathrm{d}(k)}}$$

اگر هر راس c_i را از خوشه c داشته باشیم، پس احتمال خواسته شده که از یکی از رئوس c_i به k برسیم، برابر است با میانگین احتمال‌هایی که از k به رئوس c_i رسیده‌ایم. پس داریم:

$$P_{ck}^{(t)} = \frac{\sum_{i=1}^{|c|} P_{kc_i}^{(t)}}{|c|}$$

با توجه به تعریف اختلاف سلیقه برای هر کدام از c_i و c_2 داریم:

$$r_{c_1,c_2} = \sqrt{\sum_{k=1}^{n} \frac{(P_{c_1 k}^{(t)} - P_{c_2 k}^{(t)})^2}{\mathrm{d}(k)}}$$

حال با توجه به احتمال هر ck داریم:

$$r_{c_1,c_2} = \sqrt{\sum_{k=1}^{n} \frac{(\frac{\sum_{i=1}^{|c_1|} P_{kc_i}}{|c_1|} - \frac{\sum_{i=1}^{|c_2|} P_{kc_i}}{|c_2|})^2}{\mathrm{d}(k)}}$$

پاسخ تئوری ٤٣. در یک دسته بندی خوب، بیشتر یال‌ها داخل دسته هستند و بین دسته‌ها یال‌های کمتری وجود دارد. پس برای معیار مورد نظر میتوان در هر دسته اختلاف یال‌های داخل دسته و یال‌های خارج آنرا در نظر گرفت. با توجه به این موضوع برای هر خوشه داریم:

$$\sum_{c\in P} in(c) - out(c)$$

برای بهترین شکل دسته بندی، حاصل عبارت بالا باید به مقدار بیشینه خود برسد زیرا هر چه یال‌های درون دسته‌ها بیشتر باشد، ارتباط بین آن‌ها بیشتر بوده و دسته بندی بهینه تر صورت گرفته است.

پاسخ شبیه‌سازی ۲۸. پس از `import` کردن گراف `karate_club` از کتابخانه `networkx`، ماتریس مجاورت آنرا حساب کرده و سپس با استفاده از پاسخ تئوری ۱۵، ماتریس $P \in \mathbb{R}^{n\times n}$ را حساب کرده و با استفاده از توابع پرسش‌های تئوری ۳۷ تا ۴۲، الگوریتم خواسته شده را اجرا میکنیم. با استفاده از گراف `karate_club`، برای $t = 2$ پس از قدم زدن تصادفی گروه — میماند.

پاسخ شبیه‌سازی ۲۹. مانند پرسش شبیه سازی ۲۸، پس از قدم زدن تصادفی با $t = 5$، گروه — باقی میماند.
