پروژه درس آمار و احتمال مهندسی
جریان اطلاعات در شبکه های اجتماعی
قسمت اول: نفوذ *

سید ایمان حسینی زوارکی

28 اردیبهشت 1397

1 شرایط و ضوابط

با ارسال گزارش پروژه این درس، شرایط زیر را می پذیرید:
۱) شما از هرگونه حق اعتراض به سیستم نمره دهی و نمره نهایی تان چشم پوشی می‌کنید.
۲) سیستم نمره دهی نهایی نیست و می تواند بدون اطلاع قبلی تغییر کند.
۳) تحویل شما ممکن است در صورت اعتراض به سیستم نمره دهی، تصحیح نشود.
۴) ارسال پس از مهلت تعیین شده، تصحیح نخواهد شد.
۵) شما موظفید یک گزارش دقیق و خوانا به همراه موارد تکمیلی (به طور مثال کد برنامه‌تان) برای نشان دادن کار خود ارائه دهید.

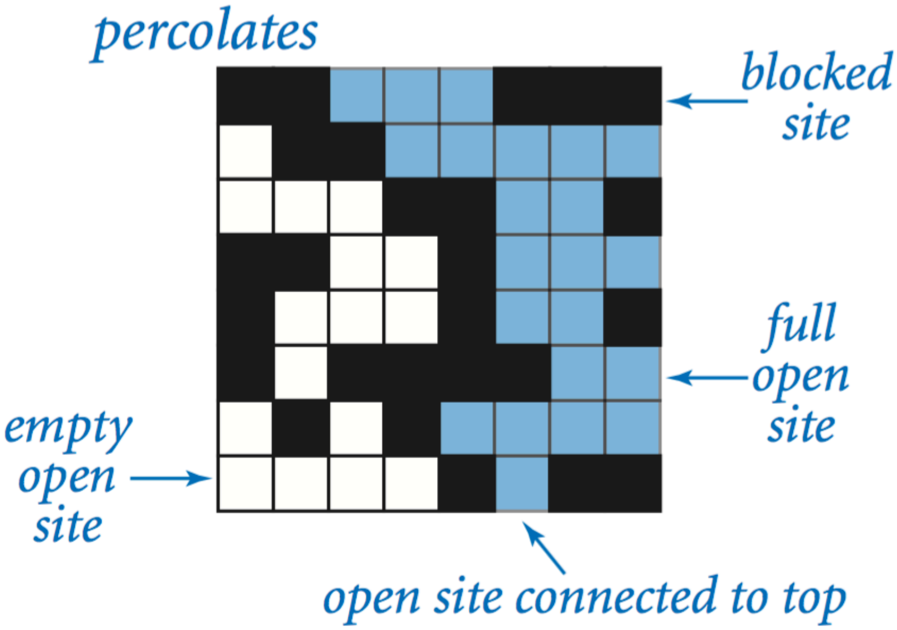
2 مقدمه

از اواخر دهه 1890 دو جامعه شناس مشهور داوید امیل دورکیم [1] و فردیناند تونیس [2] مفهوم شبکه های اجتماعی را با تحقیق در مورد گروه های اجتماعی در نظریه های خود بیان کردند. از آن زمان به بعد در بررسی تعامل بین نقش های اجتماعی در یک محیط، رویکرد ریاضیاتی رایج شد.

مدل های ریاضی (هر دو نوع تحلیلی و عددی) در طیف گسترده ای از مسائل مانند مدل رای دهندگان (برای پیش بینی نتایج انتخابات)، ارزیابی بازارهای مالی و حداکثر سازی نفوذ دراینستاگرام [3] و دیگر شبکه های اجتماعی مورد استفاده قرار می‌گیرد. ابزار مورد استفاده در این مطالعات همچنین می‌تواند در مسائلی مانند کنترل ازدحام [4] در طراحی شبکه های کامپیوتری و یا مسائل مربوط به قابلیت اطمینان برای پوشش شبکه‌های تنومند [5] مورد استفاده قرار گیرد.

---

*Percolation
[1] David Émile Durkheim
[2] Ferdinand Tönnies
[3] Instagram
[4] Congestion control
[5] Robust network coverage



شکل 1: نفوذ در فضای دو بعدی

در این پروژه ما قصد داریم مساله جریان پخش [6] اخبار و اطلاعات در یک شبکه را بررسی کنیم.
ما از مدل های مختلف برای شبکه خود و بازیگران آن (کاربران آن شبکه) استفاده خواهیم کرد و نتایج را برای آستانه ها و برای حالت اولیه سیستم می یابیم، مانند مدلی که در آن یک خبر خاص در کل شبکه پخش می‌شود.
در این پروژه می‌خواهیم نشان دهیم که چگونه آمار و احتمال، راهی برای فرمول بندی کردن و مواجه شدن با مسائل متنوع است و چگونه در زمینه های نظری و عملی ظاهر می‌شود.

3 قضیه قهوه، کولموگروف [7] و ویتنی [8]

قهوه بوسیله آب جوشی که بین دانه های قهوه جریان دارد ساخته می‌شود به این صورت که به علت واکنش شیمیایی قهوه با آب، آبی که از بین دانه های قهوه عبور می‌کند از طرف دیگر به عنوان «قهوه» خارج می شود. محیطی که در آن آب جوش موجود است، به عنوان مثال در یک قوری موکا، با قهوه پر شده است، اما بین دانه های قهوه فضای خالی برای جریان آب وجود دارد که به سمت پایین نفوذ کند.
یک شبکه سه بعدی از خانه‌هایی با مختصات صحیح در نظر بگیرید که در آن بعضی خانه‌ها با دانه‌های قهوه پر شده‌اند و بعضی از خانه‌های آن خالی هستند که آب می‌تواند در آن‌ها حرکت کند.
حال یک پیکربندی ساده‌تر دوبعدی را در نظر بگیریم. محیط ما یک شبکه‌ی دوبعدی با ابعاد NxN است که هر خانه‌ی آن می‌تواند اشغال یا آزاد باشد. برای تعیین حالت خانه‌ها(اشغال/آزاد) ما هر خانه را با احتمال p پر می‌کنیم. پس

[6]Flow
[7]**Kolmogorov**
[8]**Whitney**

شکل 2: ساختن قهوه با قوری موکا

مدل ما دو پارامتر دارد: N و p، که N نشان‌دهنده‌ی سایز محیط (محفظه‌ی بالایی قوری موکا) وp احتمال باز بودن یک خانه است که نشان‌دهنده‌ی میزان «فشردگی» قهوه است. به طور مثال در حالت حدی اعمال فشار بالا و فشرده سازی قهوه، چگالی دانه‌های قهوه بالا می‌رود و همه‌ی خانه‌ها با قهوه اشغال می‌شوند که در مدل ما این با $p = 0$ مشخص می‌شود.

مشخصا پر کردن این شبکه‌ی دو بعدی یک فرایند تصادفی است، پس با هر بار پر کردن شبکه یک پیکربندی جدید حاصل می‌شود. تصور کنید با یک مقدار مشخص N و p، ما هزار بار یک شبکه را می‌سازیم و هر بار بررسی می‌کنیم که آیا یک مسیر آزاد (بسته نشده) از ردیف بالایی به ردیف پایینی وجود دارد یا خیر.

حال اگر این کار را ۱۰۰۰۰۰ بار انجام شویم چه اتفاقی می‌افتد؟ به طور شهودی ما انتظار داریم که نسبت تعداد دفعاتی که یک مسیر یافت می‌شود به کل تعداد دفعات، به یک مقدار مشخص که فقط به p و N بستگی دارد میل کند.

تسک ۱.۱ (۵ امتیاز)

$\theta(p)$ را احتمال این که یک پیکربندی با پارامتر p (احتمال باز بودن یک سلول) نفوذپذیر باشد، یعنی یک مسیر باز از گوشه‌ی بالا به گوشه‌ی پایین وجود داشته باشد، تعریف می‌کنیم.

برای مقادیر مختلف N، $\theta(p)$ را رسم کنید. ($N \in 20, 50, 100, 400, 1000$)

چه چیزی مشاهده می‌کنید؟ اگر $N \to \infty$ چه اتفاقی می‌افتد؟

همانطور که در تسک ۱.۱ دیدید، به نظر می‌رسد $\theta(p)$ طبق انتظار ما یکنواخت است. همچنین به نظر می‌رسد که برای N های بزرگ یک p آستانه وجود دارد (به آن p_c می‌گوییم) که برای مقادیر $p > p_c$ داریم $\theta = 1$ و برای هر p

کوچکتر، $\theta = 0$. پس بیایید این ایده را دنبال کنیم و یک اثبات محکم با استفاده از مدلسازی که یاد گرفتیم بیابیم.

تسک ۱.۲ (۱۵امتیاز)
لم ۱.۳: لم زیر را اثبات کنید:
لم بورل‌ـ کانتلی
برای یک احتمال سه گانه‌ی Ω، فضای نمونه F، یک σـ‌جبر [9]از Ω و یک اندازه احتمال P،
اگر $A_1, A_2, .. \in F$ آنگاه داریم:

I) If $\sum_n P(A_n) < \infty$ then $P(A_n\ infinitely\ often) = 0$
II) If $\sum_n P(A_n) = \infty$, $\{A_n\}_{n=1}^{\infty}$ independent, then $P(A_n\ infinitely\ often) = 0$.

نکته‌ی مهم این لم چیست؟ بیایید یک کاربرد این لم را مشاهده کنیم:
فرایند پرتاب بی‌نهایت سکه‌ی شدیدا نامتقارن را در نظر بگیرید. رخدادهای مستقل این فرایند را $H_1, H_2, H_3, ...,$ می‌نامیم، که H_i رخداد شیر آمدن سکه‌ی iم است. همچنین فرض کنید سکه‌های ما شدیدا به سمت خط آمدن مایل هستند، با $P(H_n) = 1/n$ (به طور مثال سکه‌ی یک میلیونم فقط یک میلیونیم احتمال دارد که شیر بیاید). با استفاده از لم بالا ثابت کنید همچنان بی‌نهایت شیر می‌آید. (راهنمایی: از ناهمگرایی سری‌های هارمونیک استفاده کنید)

برای یک نتیجه‌ی بدیهی‌تر و غیر شهودی‌تر، مساله‌ی بالا را این بار با $P(H_n) = (\frac{99}{100})^n$ در نظر بگیرید. به عبارت دیگر، سکه‌ی اول به احتمال ٪۹۹ شیر می‌آید، سکه‌ی دوم به احتمال ٪ ۹۸.۰۱ شیر می‌آید و الی آخر .
حال ثابت کنید که در این حالت ممکن نیست بی‌نهایت بار شیر بیاید.

[10] برای بسط لم فوق، به یک تعریف دیگر نیز نیاز داریم.
برای یک دنباله‌ی داده شده از رویدادها،

$$A_1, A_2, ... \in F$$

«میدان دم» [11] آن‌ها به صورت زیر تعریف می‌شود:

$$\tau = \cap_{n=1}^{\infty}(A_n, A_{n+1}, ...)$$

میدان دم همچنین یک ساختار جبری است که اعضای آن را «رویدادهای دم» [12] می‌نامیم.
حال می‌توانیم قانون صفرـ‌یک کلموگروف [13] را بیان کنیم:

[10]Borel-Cantelli
[10]$\sigma - algebra$
[11]Tail field
[12]Tail events
[13]Kolmogorov's Zero-One Law

قضیه ۳.۲: قانون صفر_یک کلموگروف
با احتمال سه‌گانه (F,P,Ω) و یک دنباله از رخدادهای مستقل $A_1, A_2, .. \in F$ با میدان دم τ، اگر $T \in \tau$ باشد آنگاه $P(T) \in 0, 1$ است.

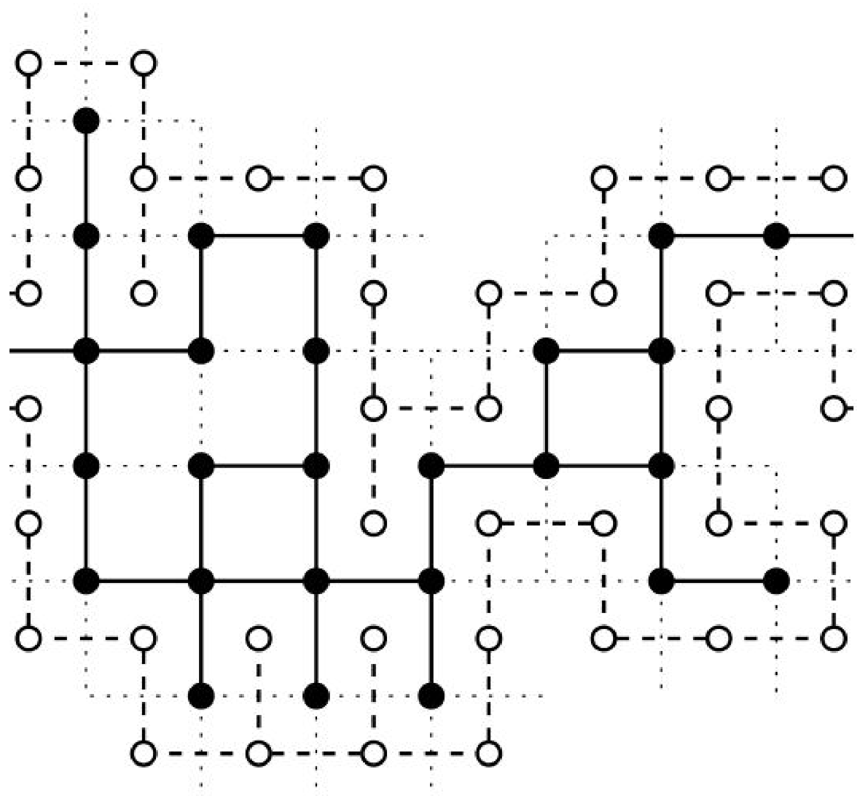
با استفاده از این قضیه می‌توان ثابت کرد در حقیقت، برای مسئله‌ی نفوذ که دربالا مطرح شد، و در واقع در حالت کلی‌تر(که لزوما به گراف شبکه‌ای که در بالا توصیف شد محدود نیست)، در حالت حدی که گراف بزرگ‌تر می‌شود، یک p_c آستانه وجود دارد که بالای این آستانه نفوذ اتفاق می‌افتاد. حال از آزمایش (تسک ۱.۱) شما احتمالا مقدار p_c را برای یک شبکه دو بعدی به دست آوردید. اما اثبات تحلیلی این مساله که p_c در واقع برابر (خطر لوث شدن!) ۱/۲ است اصلا بدیهی نیست و رسیدن به اثباتش بیش از ۲۰ سال به طول انجامیده است. با این حال ما می‌خواهیم کران‌هایی برای مقدار p_c ثابت کنیم.

تسک ۱.۳ (۱۰ امتیاز)
ثابت کنید که $p_c > 1/3$. یا به عبارت دیگر ثابت کنید که اگر $p < 1/3$ آنگاه $\theta(p) = 0$.
راهنمایی: شما می‌توانید (بدون نیاز به اثبات) از این استفاده کنید که این قضیه معادل مساله زیر است:
یک نقطه‌ی خاص را به عنوان مرکز در نظر بگیرید و تصور کنید که مشغول بازی اسنیک [14]هستید. با شروع از آن نقطه (مرکز) روی گره‌ها به صورتی حرکت کنید که مسیر خودتان را قطع نکنید. (به این مسیر یک مسیر «از خود بیزار» می‌گوییم) حال ما می‌خواهیم وجود یک مسیر از خود بیزار را جستجو کنیم که فقط شامل گره‌های باز باشد و اندازه‌اش بی‌نهایت باش. (در صورت وجود، نفوذ رخ می‌دهد). حال فرض کنید یک مسیر خاص به طول N داده شده است، چقدر احتمال دارد این مسیر یک مسیر معتبر باشد (همه‌ی گره‌های آن باز باشند)؟ حال با شروع از مرکز، چند مسیر مختلف از خود بیزار به طول N داریم؟
پاسخ دادن به این سوال دوم سخت است، اما می‌توانیم یک کران بالا روی تعداد این مسیرها بیابیم. سپس این احتمال تقریبی (که یک کران بالا برای احتمال دقیق است) را$P(n)$ می‌نامیم و تحقیق می‌کنیم که وقتی n به سمت بی‌نهایت میل می‌کند چه اتفاقی می‌افتد و چه شرایطی روی p باید داشته باشیم، که $P_\infty = 0$ باشد.

حال برای اثبات کران بالا روی p_c به یک قضیه از ویتنی نیاز داریم. این یک قضیه‌ی نظریه گراف محض است و ما اثبات آن را در اینجا ارائه نمی‌دهیم، اما شما می‌توانید با مشاهده‌ی شکل زیر ببینید که چگونه کار می‌کند.
نکته: هر مولفه‌ی همبندی ساخته شده از گره‌های باز، یک خوشه نامیده می‌شود.
قضیه ۳.۳: اگر یک دور از گره‌های بسته که مرکز را احاطه کرده است وجود داشته باشد، هیچ خوشه‌ی نامحدودی شامل مرکز وجود ندارد.
شهود این قضیه روشن است: اگر یک خوشه‌ی نامحدود داشته باشیم، نمی‌توانیم آن را در یک قفس از گره‌های بسته محصور کنیم.

تسک ۱.۴ (۲۵ امتیاز)
ثابت کنید که $p_c < 2/3$. این بسیار شبیه به نتیجه‌ی قبلی است. از قضیه ویتنی استفاده کنید و احتمال وجود یک دور« محصور کننده» از گره‌های بسته با اندازه‌ی N را بیابید و N را به سمت بی‌نهایت میل دهید.

[14] Snake



شکل 3: قضیه ویتنی