



دانشگاه صنعتی امیر کبیر
(پلی تکنیک تهران)

دانشکده مهندسی کامپیوتر

تحلیل شبکه‌های پیچیده

تمرین اول

استاد درس: دکتر مصطفی حقیر چهرقانی

آبان ۱۴۰۱

مهلت تحویل تمرین :

۱۱:۵۹ ب.ظ

۱۴۰۱/۰۸/۲۲

سوال اول

الف) یک مدل از گراف تصادفی Erdos-Renyi با تعداد گره ۷۶۲۴ و تعداد ۲۷۸۰۶ یال انتخابی تصادفی تولید کنید.

ب) یک نمونه از مدل تصادفی Small-World تولید کنید. برای این گراف نیز تعداد گره ۷۶۲۴ و تعداد ۲۷۸۰۶ یال در نظر بگیرید.

ج) مجموعه دادگان LastFM Asia Social Network را از این [لینک](#) دانلود کنید.

د) برای هر سه مدل ذکر شده در بالا نمودار توزیع درجه را در مقیاس log-log رسم کرده و با یکدیگر مقایسه کنید و تفاوت‌های بین گراف‌های تصادفی و گراف واقعی را با توجه به توزیع درجه متناظر با آن‌ها بیان کنید.

ه) در این قسمت ضریب خوشه بندی میانگین را برای هر سه گراف تولید شده در مراحل قبلی گزارش کنید. برای محاسبه این ضریب پیاده سازی خود را از فرمول‌های زیر انجام دهید.

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

$$C = \frac{1}{N} \sum_{i=1}^N C_i$$

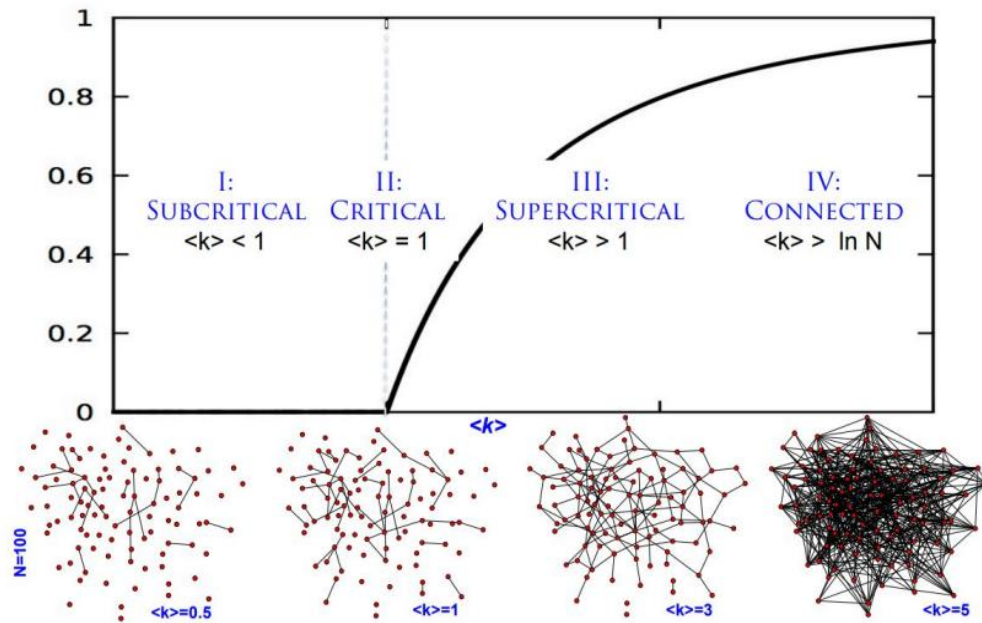
سوال دوم^۱

در این سوال باید با مصورسازی گراف ثابت کنید اندازه نسبی بزرگترین مولفه یک گراف تصادفی به شکل (۱) تغییر می‌کند.

شبیه سازی در این سوال به این صورت شکل می‌گیرد که گراف‌های تصادفی با تعداد گره یکسان N و احتمال وجود یال p تولید شده و اندازه بزرگترین مولفه گراف در آن‌ها محاسبه می‌شود. اندازه نسبی این مولفه نیز از تقسیم اندازه مولفه به تعداد گره‌ها حاصل می‌شود.

در این قسمت نشان دهید با افزایش میانگین درجه گره‌های گراف اندازه این مولفه‌ها به صورت شکل (۱) تغییر خواهد کرد. ویژگی ۴ رژیم نشان داده شده در تصویر نیز برای اطلاعات بیشتر در شکل (۲) آورده شده است. همچنین با توجه به این که گزاره‌های ادعا شده در این تصویر در حالتی برقرار است که تعداد گره‌ها به سمت بی‌نهایت میل کند، باید نشان دهید هرچه N بزرگتر شود دقت گزاره‌ها بیشتر می‌شود. در انتها نیز مقادیر پارامترهای N و p برای استنتاج هر یک از گزاره‌ها را گزارش دهید.

^۱ در حل این سوال می‌توانید از کتابخانه‌های معروف کار با گراف نظیر `Networkx` استفاده کنید.



Structural (percolation) phase transition at $\langle k \rangle = 1$ (or equivalently when $p = 1/N$)

شکل ۱

Subcritical ($\langle k \rangle < 1, p < p_c = 1/N$)

No giant component;
 $N-L$ isolated clusters, cluster size distribution is exponential;
 The largest cluster is a tree, its size $\sim \ln N$.

Supercritical ($\langle k \rangle > 1, p > p_c = 1/N$)

Unique giant component: $NG \sim (p - p_c)N$;
 GC has loops;
 Cluster size distribution: exponential.

Critical ($\langle k \rangle = 1, p = p_c = 1/N$)

Unique giant component: $NG \sim N^{2/3}$
 Contains a vanishing fraction of all nodes, $NG/N \sim N^{-1/3}$
 Small components are trees, GC has loops.

Connected ($\langle k \rangle > \ln N, p > (\ln N)/N$)

Only one cluster: $NG = N$;
 GC is dense;
 Cluster size distribution: None.

شکل ۲

سوال سوم

یک گراف دوبخشی با تعداد گره‌های N_1 و N_2 در دو مجموعه را در نظر بگیرید و باتوجه به ویژگی‌های این گراف به سوالات زیر پاسخ دهید.

الف) بیشترین تعداد لینکی که این شبکه می‌تواند دارا باشد چقدر است؟

ب) محاسبه کنید چه تعداد لینک نمی‌تواند در مقایسه با حالت دوبخشی نبودن این گراف با اندازه $N = N_1 + N_2$ اتفاق بیفتد.

ج) اگر $N_1 < N_2$ برقرار باشد، در رابطه با چگالی شبکه که نسبت تعداد لینک‌ها به تعداد بیشترین تعداد لینک می‌باشد، چه می‌توان گفت؟

د) عبارتی برای ارتباط بین N_1 و N_2 و میانگین درجه هر دو بخش یعنی k_1 و k_2 گراف بیابید.

سوال چهارم

برای این سوال ابتدا مجموعه دادگان Astro Physics collaboration network را از این [لینک](#) دانلود کنید.

همانطور که در کلاس گفته شده است در حل مسئله پیشینه‌سازی تاثیر از دو الگوریتم مرسوم حریصانه و CELF استفاده می‌شود. در این قسمت نیز لازم است با کمک گرفتن از این دو الگوریتم مجموعه گره‌های تاثیرگذار در این شبکه را با اندازه ۱۰ که توسط آن گره‌ها میزان تاثیر پیشینه خواهد شد را بیابید.

تفاوت عملکرد این دو الگوریتم را از نظر تئوری و عملی با مقایسه زمان اجرا و هم‌چنین خروجی نهایی هر دو الگوریتم شرح دهید.

سوال پنجم

در این سوال با در نظر گرفتن مجموعه دادگان داده شده در سوال چهارم، این بار با هدف یافتن مجموعه بهینه‌ای از گره‌های گراف به منظور کشف شیوع از الگوریتم‌های CELF و حریصانه استفاده کنید. در پاسخ به این سوال الگوریتم مورد استفاده را شرح دهید و همچنین روشی برای افزایش سرعت آن ارائه کنید.

میزان پاداش و هزینه در این مسئله را نیز به شکل زیر در نظر بگیرید.

پاداش = $0.25 * \text{حداکثر تعداد گره‌های تاثیرگذاشته شده}$

هزینه = $0.8 * \text{جمع قدر مطلق وزن یال‌های متصل به گره‌های انتخاب شده در مجموعه}$

نکات مربوط به تحویل تمرین

- برای قسمت‌های مرتبط با پیاده‌سازی زبان مورد قبول، زبان پایتون است.
- نظم و خوانایی در نوشتن گزارش و کدها از اهمیت بالایی برخوردار است. کدهای خود را تا حد امکان کامنت‌گذاری کنید. خروجی کدها و نتایج سوالات را درون گزارش بنویسید و از توضیح اضافی کد و موارد دیگر خودداری فرمایید (کد بدون گزارش ارزشی ندارد).
- در پیاده‌سازی بخش‌های مختلف امکان استفاده از کتابخانه‌های آماده مربوط به کار با گراف را ندارید؛ مگر در مواردی که در صورت سوال ذکر شده باشد.
- برای هر سوال یک پوشه جداگانه در نظر گرفته، کد و مواردی از قبیل خروجی برنامه و نمودارها را در آن ذخیره نمایید. این پوشه‌ها را به همراه یک فایل report.pdf برای گزارش و توضیح سوالات، درون یک فایل فشرده شده با فرمت HW01_StdNumber.zip که StdNumber شماره دانشجویی شماست، در سامانه بارگذاری کنید.
- مطابق قوانین دانشگاه هر گونه کپی‌برداری ممنوع است و در صورت مشاهده، نمره هر دو طرف صفر داده می‌شود. اما استفاده از توضیحات موجود در اینترنت به منظور یادگیری الزاما با ذکر منبع بلامانع است.
- در صورت وجود هر گونه سوال یا ابهامی با ایمیل درس در تماس باشید:

cna.assistant@gmail.com