



تحلیل شبکه‌های پیچیده

تمرین اول

{GNP, Influence Maximization, Outbreak Detection}



برای ارسال تمرین به نکات زیر توجه کنید.

۱. ملاک اصلی انجام تمرین گزارش آن است و ارسال کد بدون گزارش فاقد ارزش است. برای این تمرین یک فایل گزارش در قالب pdf تهیه کنید و در آن برای هر سوال، تصاویر ورودی، تصاویر خروجی و توضیحات مربوط به آن را ذکر کنید. سعی کنید توضیحات کامل و جامعی تهیه کنید.
۲. زبان برنامه‌نویسی برای انجام تمرین‌ها، پایتون(Python) در نظر گرفته شده است.
۳. در نظر داشته باشید کدهای شما باید قابلیت اجرا در هنگام ارائه را داشته باشند. همچنین بر روی کدهای خود مسلط باشید
۴. کدهای ارسالی خود را برای افزایش خوانایی و درک بهتر به صورت مناسب کامنت‌گذاری کنید.
۵. پاسخ سوالات تشریحی، فایل‌های کد و گزارش خود را در یک فایل فشرده قرار داده و با نام با فرمت HW1_StudentID.zip ارسال نمایید.
۶. در صورت وجود ابهام یا اشکال می‌توانید از طریق [کانال](#) با ما در ارتباط باشید (داخل کانال لینک گروه نیز موجود می‌باشد)
۷. برای اطلاع از ددلاین تکلیف، به تاریخ اعلام شده در سایت توجه کنید.
۸. توجه داشته باشید به ازای هر روز تاخیر در تحویل تکلیف، ۵ درصد از نمره کسب شده در آن تکلیف کسر خواهد شد و پس از هفت روز نمره تکلیف صفر خواهد شد. (سامانه تحویل تمرین در ابتدای روز هشتم بسته خواهد شد)
۹. مطابق با قوانین دانشگاه هر نوع کپی برداری و تقلب در انجام تکالیف و پروژه‌های این درس مجاز نبود و در صورت مشاهده نمره طرفین صفر منظور می‌گردد.
۱۰. استفاده از کتابخانه‌های موجود گرافی نظیر networkx تنها برای سوال سوم و چهارم مجاز است.

سوال اول: بررسی ویژگی‌های گرافی

- الف) برای این سوال سه گراف در پیوست شما قرار داده شده است. گراف‌ها به صورت پیش فرض بدون جهت می‌باشند (مگر اینکه در حالتی خواسته شود گراف را جهت‌دار در نظر بگیرید). از بین گراف‌های داده شده یک گراف را به دلخواه انتخاب کنید و تعداد گره‌ها و یال‌های این گراف دلخواه را به دست آورید.
- ب) یک گراف بر اساس مدل تصادفی Erdos- Reyni با تعداد گره و یال گراف انتخابی قسمت الف تولید کنید. مقدار احتمال در تولید گراف Erdos-Reyni را به سلیقه خود انتخاب کنید.
- ج) یک گراف بر اساس مدل تصادفی Small-World با تعداد گره و یال گراف انتخابی قسمت الف تولید کنید. برای گام rewire در گراف Small-World مقدار احتمال را 0.2 در نظر بگیرید.
- د) برای تمامی گراف‌ها (سه گراف پیوست و دو گراف قسمت ب و ج) نمودار توزیع درجه را در مقیاس $\log\text{-}\log$ رسم کنید. تفاوت بین گراف‌های دنیای واقعی و گراف‌های تصادفی را با توجه به توزیع درجه آن‌ها بیان کنید. آیا بین گراف‌های دنیای واقعی نیز تفاوتی وجود دارد؟ تحلیل خود را بیان کنید.
- ه) برای سه گراف موجود در پیوست طول مسیر متوسط (average path length) و قطر گراف را به دست آورید. برای پاسخ به این بخش یک بار گراف‌ها را به صورت جهت‌دار و یک بار بدون جهت در نظر بگیرید. برای هر گراف تحلیل کنید که طول مسیر متوسط و قطر گراف نسخه جهت‌دار و بدون جهت آن چه تفاوتی با یکدیگر دارد. در نسخه جهت‌دار گراف‌ها دلیل تفاوت دو مقدار مذکور چیست و وابسته به چه عواملی می‌تواند باشد؟ به طور مشابه تفاوت دو مقدار مذکور برای نسخه بدون جهت را بررسی و تحلیل کنید.
- و) ضریب خوشه‌بندی میانگین برای تمامی گراف‌ها را محاسبه کنید و تحلیل خود را از اختلاف مقادیر بین گراف‌های دنیای واقعی و گراف‌های تصادفی ارائه دهید. در صورتی که گرافی دارای ضریب خوشه‌بندی میانگین 0 باشد، علت آن چه چیست؟ این گراف‌ها چه نوع گرافی می‌توانند باشد؟

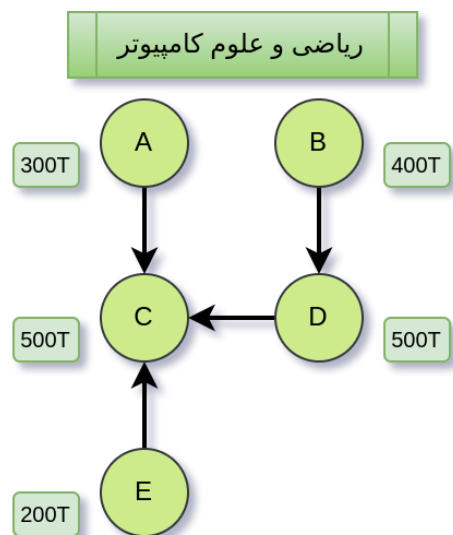
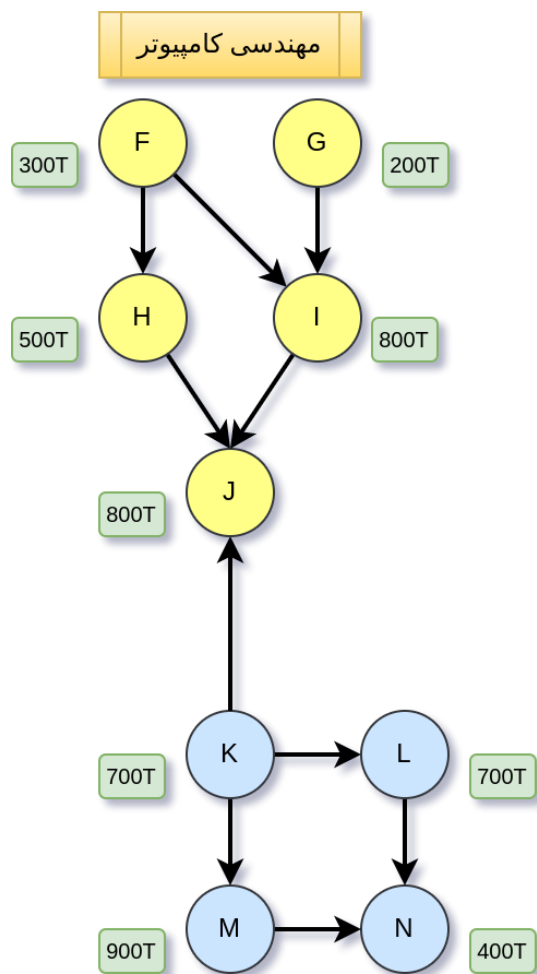
سوال دوم: مفاهیم گراف تصادفی

- الف) قطر گراف Watts-Strogatz را محاسبه کنید. (مراحل محاسبه را به طور کامل ذکر کنید).
- ب) اثبات کنید در یک گراف k -regular که دارای n راس و میزان expansion آن معادل α باشد، بین هر دو راس این گراف یک مسیر به طول $O(\log(n)/\alpha)$ وجود دارد.
- ج) cluster coefficient را در یک گراف G_{np} محاسبه کنید.

- الف) تعریف تابع monotonic و sub modularity را بیان کنید. آیا تابعی که تعداد زیرمجموعه‌های یک مجموعه را حساب می‌کند، دو ویژگی مذکور را دارد؟ اگر دارد، آن را اثبات کنید و در غیر این صورت، مثال نقضی بیاورید.
- ب) دو نوآوری الگوریتم CELF را به صورت کامل توضیح دهید.
- ج) دانشگاه امیرکبیر قصد دارد جلوی شیوع موش‌ها را بگیرد! برای این کار تنها یک میلیون تومان بودجه در نظر گرفته شده است. برای گیرانداختن آن‌ها می‌توان از تله‌گذاری استفاده کرد. استفاده از تله‌ها در سالن‌های مختلف هزینه‌های متفاوتی دارد. گراف شکل ۱ نقشه نمادین دانشکده‌های مهندسی کامپیوتر، ریاضی و علوم کامپیوتر و ساختمان بهداشت که درگیر مشکل مذکور شده‌اند را نشان می‌دهد. هر یک از گره‌ها مربوط به یک سالن است و در کنار آن هزینه تله‌گذاری به هزار تومان نوشته شده است (مثلاً 300T به معنای ۳۰۰ هزار تومان است). همچنین پال‌ها مسیر حرکت موش‌ها را نشان می‌دهد. فرض کنید که احتمال پدید آمدن موش در هر یک از سالن‌ها یکسان باشد. **قصد داریم تله‌گذاری را جوری انجام دهیم که بیشترین تعداد موش را گرفته و شیوع را کاهش دهیم.** الگوریتم حریصانه را به صورت گام به گام برای مسئله طرح شده انجام دهید. مطابق این الگوریتم با شرایط ذکرشده موش‌های چند سالن را می‌توان گرفت؟
- د) الگوریتم CELF را به صورت گام به گام برای مسئله طرح شده انجام دهید. مطابق این الگوریتم با شرایط ذکرشده موش‌های چند سالن را می‌توان گرفت؟
- ه) کدام الگوریتم برای شرایط این مسئله عملکرد بهتری داشته است؟ چرا؟
- و) ارتباط مسئله بیشینه‌سازی تاثیر (Influence Maximization) با مسئله تشخیص شیوع (Outbreak Detection) در چیست؟ آیا این دو مسئله قابل تبدیل به یکدیگر هستند؟ آیا قابل استفاده در یک گراف یکسان هستند؟ جهت پال‌ها چه اهمیتی در هر یک از دو مسئله دارد؟

ملاحظات:

- ۱- هدف از این مسئله گرفتن بیشترین موش از سالن‌های مختلف است. بنابراین تاخیر در گرفتن موش و یا آلوده شدن سالن‌ها اهمیتی ندارد. به عنوان مثال اگر در سالن J تله‌گذاری شود، تمام موش‌های دانشکده مهندسی کامپیوتر را می‌گیرد بنابراین $f(\{J\}) = 5$. طبیعتاً در این حالت اگر شیوع موش از سالن F و G صورت گرفته باشد، هم با تاخیر شیوع شناسایی می‌شود و هم سالن‌های زیادی درگیر می‌شوند ولی در مسئله ما اهمیت ندارد!
- ۲- با گذاشتن یک تله در یک سالن بی‌نهایت موش را می‌توان گرفت. بنابراین نیازی به گذاشتن چند تله در یک سالن نیست.
- ۳- اگر یک موش از یک منبع شیوع را بگیریم، فرض می‌شود که تمام موش‌های آن منبع گرفته شده است. به عنوان مثال اگر در H تله‌گذاری شود، موش‌های سالن F به طور کامل گرفته می‌شوند؛ یعنی خواهیم داشت: $f(\{H\}) = 2$
- ۴- گراف داده شده قطعی است و نیازی به تولید realization از گراف نیست.



بهداشت

شکل ۱ - گراف سوال ۳

یک انتشاراتی به تازگی کتابی در حوزه تحلیل شبکه‌های پیچیده به چاپ رسانده است. جهت شناساندن و تبلیغ بیشتر کتاب تصمیم گرفته شده است که پیش از روز اصلی رونمایی، هشت نسخه آن را برای افرادی خاص با بیشترین تاثیر ارسال بکند. ارتباط دوستی افراد در شبکه اجتماعی فیس‌بوک موجود است و بر مبنای آن قصد انتخاب افراد را داریم. فرض کنید که ارتباط دوستی، دوطرفه است. مطمئن هستیم که هشت نفر اول حتماً از کتاب خوششان خواهد آمد. همچنین فرض کنید که اگر فردی تبلیغ را دریافت کرده باشد و از آن خوشش آمده باشد، تبلیغ را در صفحه فیس‌بوکش به اشتراک می‌گذارد. افرادی که با فرد تبلیغ‌کننده در تعامل هستند، با احتمال ۱۰٪ از کتاب خوششان خواهد آمد و آن‌ها نیز به طور مشابه کتاب را تبلیغ می‌کنند. در غیر این صورت کاری انجام نمی‌دهند. هدف بیشینه کردن افرادی است که از کتاب خوششان می‌آید.

الف) فایل مربوط به گراف شبکه اجتماعی فیس‌بوک را از این [لینک](#) بارگیری نمایید. این گراف ذاتاً بدون جهت و بدون وزن است. هر سطر از این فایل بیانگر یک یال از گراف است که شامل مبدا و مقصد هر یال است. بر اساس این فایل گراف مذکور را ایجاد کنید. تعداد گره‌ها یعنی افراد و تعداد یال‌ها یعنی دوستی‌ها در گراف چقدر است؟

ب) همانطور که بیان شد احتمال تبلیغ در طول یک یال ۰.۱ است؛ به بیان دقیق‌تر اگر یک طرف یال فعال (activate) شده باشد (یعنی از کتاب خوشش آمده باشد)، طرف دیگر با احتمال ۱۰٪ فعال خواهد شد (یعنی از کتاب خوشش خواهد آمد و آن را تبلیغ خواهد کرد). در نظر داشته باشید که هر یال تنها یک بار می‌تواند فعال شود. مفهوم realization را توضیح دهید و بیان کنید ضرورت استفاده از آن در چیست؟ از گراف موجود ۱۰ realization ایجاد کنید و تا انتهای سوال با realization های ایجاد شده کار کنید. پیشنهاد می‌شود که realization های موجود را ذخیره کنید یا با استفاده از random seed کاری کنید که با هر بار اجرای برنامه آن‌ها به شکل ثابت ایجاد شوند.

ج) در ابتدا برای انتخاب مجموعه هشت گره برتر از دو الگوریتم ابتکاری و ساده استفاده کنید؛ نخست، ۸ گره با بیشترین درجه را انتخاب کنید. دوم به صورت تصادفی مجموعه‌ای ۸ عضوی از گره‌ها را انتخاب کنید. پس از اجرای دو الگوریتم شماره گره‌های انتخاب‌شده و مقدار $f(S)$ را برای هر یک از دو الگوریتم گزارش کنید.

د) حال برای انتخاب مجموعه مناسب دو الگوریتم حریصانه و CELF را مطابق اسلاید درس پیاده‌سازی کنید. پس از اجرای دو الگوریتم، شماره گره‌های انتخاب‌شده و مقدار $f(S)$ را برای هر یک از دو الگوریتم گزارش کنید.

ه) مقدار $f(S)$ مجموعه‌داده‌های ۸ عضوی هر یک از چهار الگوریتم بیان‌شده را با یکدیگر مقایسه کنید. تحلیل کنید که کدام روش مناسب‌تر است. آیا الگوریتم CELF از الگوریتم حریصانه خروجی دقیق‌تری دارد؟ توضیح دهید.

و) زمان اجرای هر چهار الگوریتم را با یکدیگر مقایسه کنید. آیا الگوریتم CELF از الگوریتم حریصانه سریع‌تر است؟ چرا؟