



تحلیل شبکه‌های پیچیده

تمرین دوم

{Community Detection}

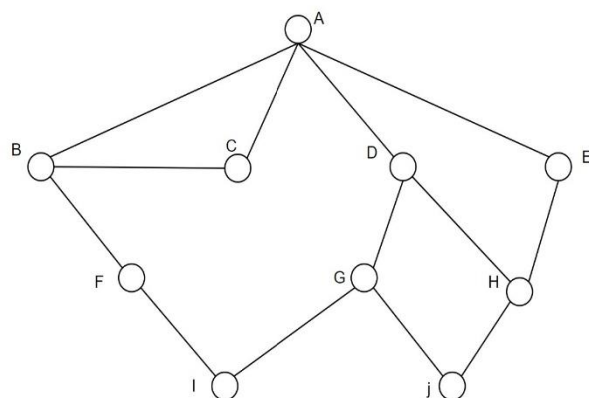


برای ارسال تمرین به نکات زیر توجه کنید.

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 HW2_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 گرافیکی نظیر networkx غیرمجاز است. استفاده از کتابخانه numpy در این تمرین مجاز است.
۱۱. سوالات ۱، ۴، ۵ و ۷ تماماً توضیحی و محاسباتی هستند و نباید پیاده‌سازی شوند. سوالات ۲، ۳ و ۶ علاوه بر بخش‌های توضیحی و محاسباتی دارای بخش پیاده‌سازی هم هستند.
۱۲. عدد جلوی هر بخش بارم آن بخش را نشان می‌دهد.

سوال اول

معیار betweenness centrality را برای یال‌های گراف شکل ۱ محاسبه کنید. با تعیین یک حد آستانه مناسب جامعه^۱های این گراف را مشخص کنید. با حذف هر یال نیازی به محاسبه مجدد معیار وجود ندارد، بلکه محاسبه یکباره آن کفایت می‌کند. (۹ نمره)



شکل ۱. گراف سوال ۱

سوال دوم

الف) مجموعه دادگان v2v.txt را در نظر بگیرید. در این فایل هر سطر، یک یال از گراف جهت‌دار را نشان می‌دهد که گره اول مبدا و گره دوم مقصد است. الگوریتم pagerank را با توجه به اسلایدهای درس پیاده‌سازی کرده و بر روی مجموعه‌دادگان v2v اعمال کنید. از بین همسایگان گره‌های ۶۸۸، ۳۸۷، ۲۷۷، ۸۷۶، ۹۹۹، ۱۷۷۷ و ۶۳۱۹؛ ۱۰ گره با بیش‌ترین شباهت را گزارش کنید. (۱۶ نمره)

ب) با توجه به این که الگوریتم pagerank نوعی شباهت ساختاری را می‌سنجد، به نظر شما با کمک این الگوریتم می‌توان یک الگوریتم خوشه‌بندی روی یک گراف ارائه داد؟ اگر جواب مثبت است، راهکار خود را با ذکر دلیل بیان کنید؛ در غیر این صورت استدلال کنید چرا این کار شدنی نیست. این قسمت اختیاری است و نمره امتیازی دارد. (۲ نمره)

الف) شباهت دو گره در یک گراف را می‌توان به صورت تکرارشونده^۲ با کمک روابط زیر بدست آورد. در این روابط R شباهت دو گره a و b را نشان می‌دهد.

$$R_0(a, b) = \begin{cases} 0 & (\text{if } a \neq b) \\ 1 & (\text{if } a = b) \end{cases} \quad \text{رابطه ۱}$$

$$R_{k+1}(a, b) = \frac{C}{|I(a)||I(b)|} \sum_{i=1}^{|I(a)|} \sum_{j=1}^{|I(b)|} R_k(I_i(a), I_j(b)) \quad \text{رابطه ۲}$$

که C یک عدد ثابت بین صفر و یک و I(a) مجموعه گره‌های همسایه-ورودی گره a هستند. منظور از همسایه‌های-ورودی a گره‌هایی است که یال از آن‌ها خارج شده و به گره a وارد شده است. اگر I(a) یا I(b) برابر صفر باشد؛ مقدار R(a,b)=0 است. رابطه ۲ را با شرایط بالا روی گراف v2v اعمال کنید. معیار شباهت R را برای گره‌ها و موارد خواسته شده در سوال ۳ گزارش کنید. تعداد مراحل (مقدار k) را برابر جزء صحیح تقسیم قطر گراف بر ۴ در نظر بگیرید. (۱۵ نمره)

ب) نشان دهید رابطه ۲ خاصیت یکنوایی دارد. این سوال اختیاری است و نمره امتیازی دارد. (۳ نمره)

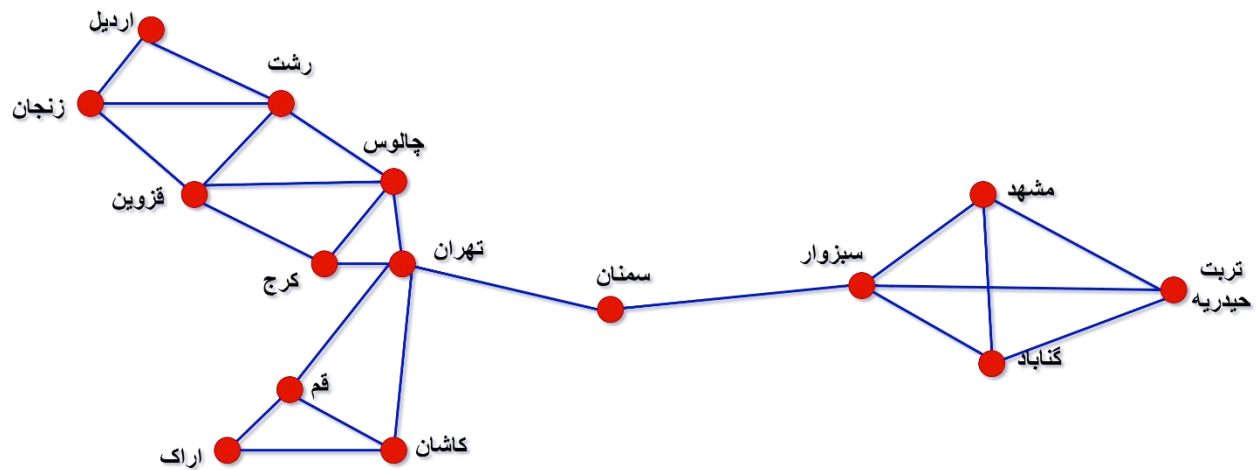
سوال چهارم

- الف) تابع $Q(G, S)$ Modularity در مبحث تقسیم‌بندی گراف را بنویسید. (۲ نمره)
- ب) متغیرهای G, S, m, A و k_i را توضیح دهید. (۳ نمره)
- ج) اثبات کنید که می‌توان تابع Modularity را به فرم زیر نوشت؛ به گونه‌ای که s یک بردار از عناصر $+1$ و -1 باشد که اندیس i -ام بیانگر عضو بودن گره i -ام در جامعه اول هست یا خیر. (۸ نمره)

$$Q(G, s) = \frac{1}{4m} s^t B s$$

سوال پنجم

به تازگی یک شرکت در حوزه زیرساخت اینترنت تصمیم گرفته است که تعدادی از شهرهای ایران را با استفاده از فیبر نوری مطابق گراف شکل ۲ به هم متصل کند. این شرکت قصد دارد تا تعدادی نمایندگی برای هر مجموعه شهرهایی که با هم ارتباط زیادی دارند تاسیس کند. برای این منظور باید جوامع گراف مشخص شود. با استفاده از CPM^3 با $k=3$ شهرهایی که در یک حوزه نمایندگی قرار می‌گیرند را مشخص کنید. مراحل را به صورت کامل ذکر کنید. (۹ نمره)



شکل ۲. گراف سوال ۵

در این سوال می‌خواهیم شناسایی جوامع را با کمک روش Spectral Clustering انجام دهیم. مجموعه‌داده این سوال را از طریق [لینک](#) دریافت کنید. این مجموعه‌داده یک گراف است که می‌بایست یال‌های آن را بدون جهت و بدون وزن فرض کنید.

الف) توضیح دهید که روش Spectral Clustering چگونه کار می‌کند؟ (۵ نمره)

ب) ماتریس Laplacian گراف فوق را بدست آورید. سپس مقادیر ویژه و بردارهای متناظر با آن را بیابید. بدست آوردن بردارهای ویژه با استفاده از کتابخانه‌های موجود نظیر numpy مجاز است. برای هر گره یک نمایش جدید یک بعدی مبتنی بر بردارهای ویژه ایجاد کنید. (۶ نمره)

ج) حال با استفاده از روش k-means گراف را خوشه‌بندی کنید. در این بخش می‌بایست مقدار k ای که بهترین نتیجه در معیار Min-Cut و Modularity را برای ما ایجاد می‌کنند به صورت جداگانه گزارش کنید. (۱۴ نمره)

د) آیا می‌توانستیم از همان ابتدا روش k-means را بر روی گراف اجرا کنیم و با استفاده از آن گراف را بخش‌بندی کنیم؟ اگر بله، ماتریس Laplacian چه مزیتی داشته است؟ اگر خیر، چرا؟ (۵ نمره)

الف) از میان روش‌های Girvan-Newman, Fast Modularity, Spectral Clustering, CPM و AGM کدام یک از آن‌ها همپوشان بودن جوامع را پشتیبانی می‌کند و کدام یک جوامع را به صورت غیر همپوشان مدل‌سازی می‌کنند؟ توضیح دهید. (۵ نمره)

ب) از بین روش‌هایی که همپوشانی را پشتیبانی می‌کند، کدام یک به این فرض که «به طور کلی، در محل همپوشانی جوامع مختلف، تراکم یال‌ها بیشتر از تراکم یال‌ها در هر یک از جوامع است» معتقد است؟ توضیح دهید. (۳ نمره)