



دانشگاه صنعتی امیر کبیر
(پلی تکنیک تهران)

دانشکده مهندسی کامپیوتر

تحلیل شبکه‌های پیچیده

تمرین دوم

استاد درس: دکتر مصطفی حقیر چهرقانی

آذر ۱۴۰۱

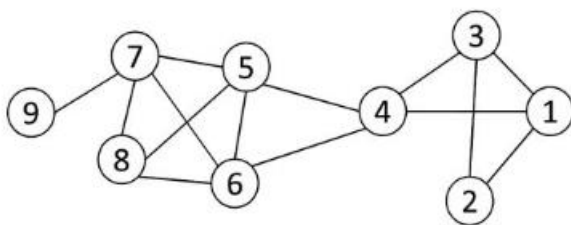
مهلت تحویل تمرین :

۱۱:۵۹ ب.ظ

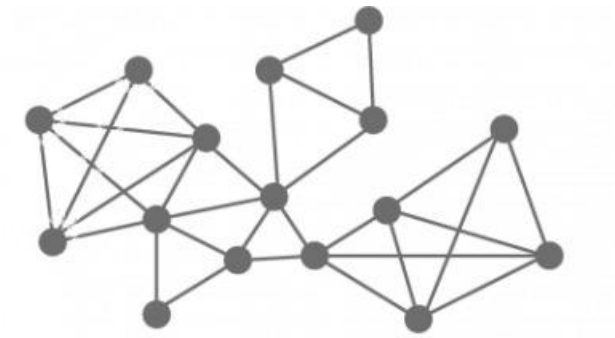
۱۴۰۱/۱۵/۱۵

سوال اول

الف) معیار $\text{betweenness centrality}$ را برای هر یال از گراف زیر محاسبه کنید. پس از تعیین یک مقدار آستانه برای میزان مرکزیت هر یال، Community های موجود در این گراف را مشخص کنید.



ب) با استفاده از الگوریتم $\text{Clique Percolation}$ ، Community های موجود در گراف زیر را مشخص کنید.



سوال دوم

مجموعه دادگان زیر با توضیحات آورده شده برای هر کدام را در نظر بگیرید.

Graph Name	Edge list	Node names	Description
AirportDataset	airports. edges	airports. codes	Two airports are connected by an edge if there is a flight between those two airports.
BibleDataset	bible. edges	bible. names	Network of nouns. Two nouns are connected by an edge if they appear in the same verse.

الف) در این بخش لازم است توابعی برای محاسبه معیارهای مرکزیت زیر بنویسید.

Closeness •

$$C_i = \frac{n}{\sum_j d_{ij}}$$

efficiency •

$$C_i = \frac{1}{n-1} \sum_{j(\neq i)} \frac{1}{d_{ij}}$$

Degree •

Katz •

برای این سوال باید فایل edges. مربوط به هر گراف که در ضمیمه‌ی تمرین موجود می‌باشد را به عنوان ورودی در نظر بگیرید و در نهایت خروجی برنامه فایلی است که در آن برای هر گره یک خط با فرمت nodeID: Centrality Value مرتب شده بر اساس مقادیر مرکزیت گره‌ها وجود دارد. به ازای هر معیار و هر گراف، در این بخش یک فایل خروجی لازم است.

ب) مقادیر مرکزیت محاسبه شده در هر فایل را به صورت نموداری که بردار افقی آن شماره گره‌ها، مرتب شده بر اساس مقادیر مرکزیت‌شان و بردار عمودی مقدار مرکزیت آن‌ها می‌باشد، نشان دهید. در این بخش نیز باید به ازای هر معیار در هر گراف یک نمودار رسم شود.

ج) در این بخش لازم است نمودارهای خروجی در قسمت ب را تحلیل کرده و نکات جالب توجه خود را در چند سطر شرح دهید.

د) برای هر گراف یک جدول با سه ستون تشکیل دهید. در این جداول هر ستون متناسب با یک معیار مرکزیت در نظر گرفته شود و در آن نام ۱۰ گره برتر با مقدار مرکزیت متناظر را ذکر کنید. لازم به ذکر است برای تشکیل این جداول باید حتما نام واقعی گره‌ها و نه شناسه آن‌ها آورده شود.

ه) بر اساس جداول قسمت قبل، در رابطه با اینکه برای هر گراف کدام معیار در یافتن گره‌های مهم بهتر عمل کرده است چند خط توضیح دهید. در این بخش لازم است با توجه به نظر شخصی خود و بررسی نتایج، توضیحات خود را ارائه دهید.

نکاتی که در حل این سوال باید در نظر بگیرید:

۱. برای هر گراف دو فایل `names` و `edges`. در ضمیمه‌ی تمرین قرار داده شده‌اند. فایل `names`. مربوط به هر گراف، شامل نگاشت اسامی واقعی گره‌ها به آیدی متناظرشان است. توجه کنید که ورودی برنامه‌های شما فایل `edges`. خواهد بود و از اسامی گره‌ها در توضیح بخش‌های معنایی گراف‌ها استفاده می‌شود.

۲. در این سوال تمامی گراف‌ها به صورت بدون وزن و بدون جهت در نظر گرفته می‌شوند.

۳. کدهای مربوط به مجموعه دادگان فرودگاه‌ها `airports.codes` با استفاده از فایل `name_of_airports.csv` به اسامی واقعی فرودگاه‌ها نگاشت می‌شوند.

۴. برای محاسبه‌ی معیار Closeness تنها بزرگترین جزء متصل گراف را در نظر بگیرید.

۵. در این سوال استفاده از کتابخانه به غیر از استفاده مستقیم در محاسبه معیارها مجاز است.

سوال سوم^۱

مجموعه دادگان در نظر گرفته شده برای این سوال نیز همانند سوال دوم می‌باشد.

در این سوال لازم است یکی از گراف‌های داده شده را به دلخواه انتخاب کرده و آن را با محاسبه Eigenvectorها خوشه بندی کنید.

الف) در ابتدا باید تکه کدی بنویسید که با در نظر گرفتن فایل edges. به عنوان ورودی ماتریس مجاورت متناظر با گراف و ماتریس درجه را بسازد و با استفاده از آنها، ماتریس Laplacian را محاسبه کند.

ب) پس از ساخت این ماتریس Eigenvectorهای آن را مشخص کرده و بعد از مرتب سازی مقادیر ویژه، بردار متناظر با دومین کوچکترین مقدار را نیز مشخص کنید.

ج) با کمک گرفتن از بردار مشخص شده در قسمت قبل با یکی از روش های DBSCAN و KMeans گراف را خوشه بندی کنید.

د) در انتها نیز دو معیار Modularity و Min-Cut را به منظور بررسی خوشه بندی نهایی محاسبه و گزارش کنید.

لازم به ذکر است در این بخش امکان استفاده از کتابخانه وجود ندارد و برای محاسبه این دو معیار می‌توانید تعریف ارائه شده در کلاس را پیاده سازی کنید.

نکاتی که در این سوال باید در نظر بگیرید:

۱. در این سوال گراف را باید به صورت بدون جهت و بدون وزن در نظر بگیرید.
۲. برای هر گراف دو فایل names و edges. در ضمیمه‌ی تمرین قرار داده شده‌اند. فایل names. مربوط به هر گراف، شامل نگاشت اسامی واقعی گره‌ها به آیدی متناظرشان است. توجه کنید که ورودی برنامه‌های شما فایل edges. خواهد بود.
۳. در این سوال استفاده از کتابخانه برای محاسبه بردارهای ویژه و مقادیر ویژه مجاز است.

^۱ در این سوال استفاده از کتابخانه‌هایی نظیر networkx مجاز است.

سوال چهارم

مجموعه دادگان مورد نظر در این سوال دادگان [Wiki-Vote](#) می‌باشد. در این سوال با استفاده از تکنیک Modularity optimization به صورت سلسله مراتبی هربار با aggregate کردن اجتماع‌های حاصل شده در مرحله قبل طبق بردارهای ویژه اجتماع جدیدی تشکیل دهید و زمانی که معیار Modularity gain بدست آمده برای اجتماع حاصل شده مقداری کمتر از مرحله قبل پیدا کرد این روند را متوقف کرده و اجتماع های نهایی را به همراه اعضاء هر اجتماع گزارش کنید.

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 HW02_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