به نام خدا



دانشگاه صنعتی امیرکبیر

(پلی‌تکنیک تهران)

درس تحلیل شبکه‌های پیچیده

استاد حقیرچهرقانی

تمرین دوم

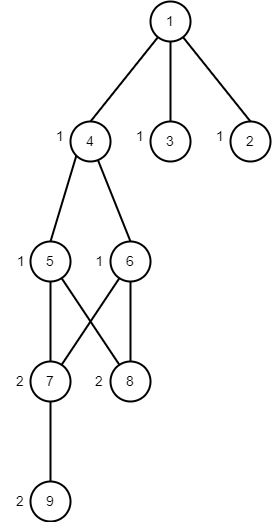
علیرضا مازوچی

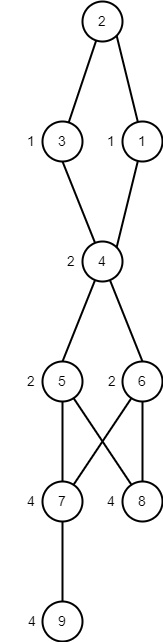
۴۰۰۱۳۱۰۷۵

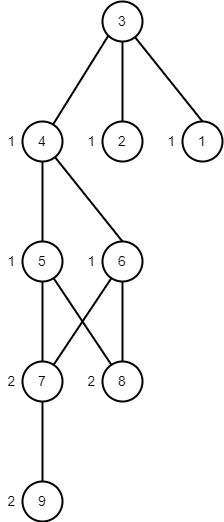
سوال ۱

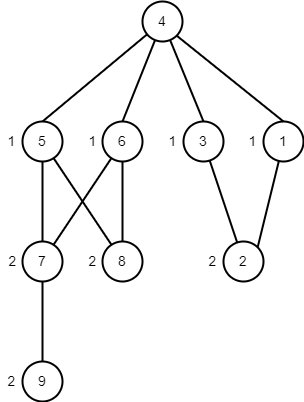
**الف)** ابتدا باید برای هر گره یک BFS تولید کرد و سپس امتیاز وابستگی (dependency score) گره به تمام یال‌های باقی‌مانده در BFS را محاسبه کرد. طبیعتا گره ریشه در BFS دارای وابستگی صفر به یال‌های غایب در درخت خواهد بود.

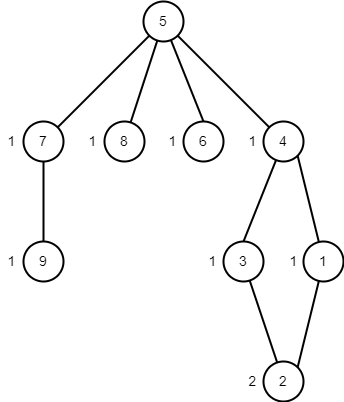
در قسمت‌های بعد محاسبات مربوط به هر BFS به ترتیب برای تمام گره‌ها آورده شده است:

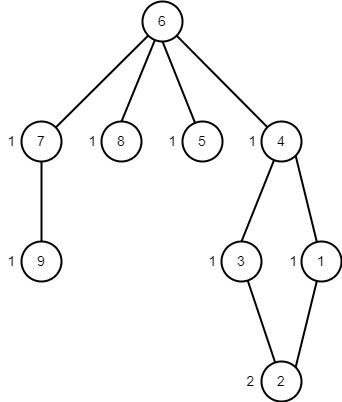


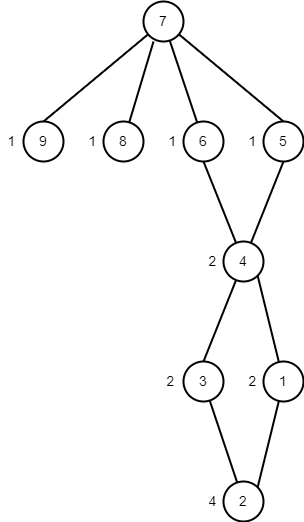


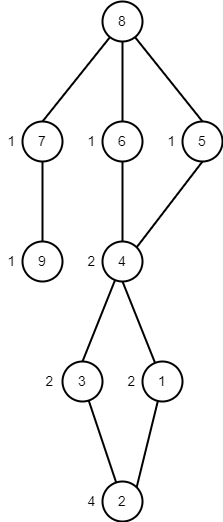


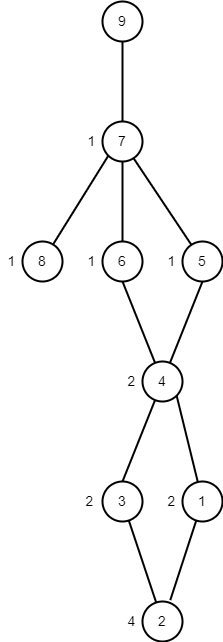




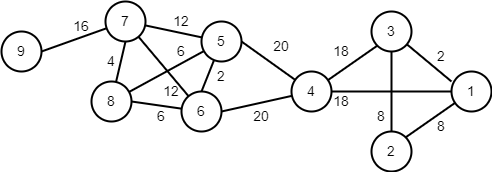






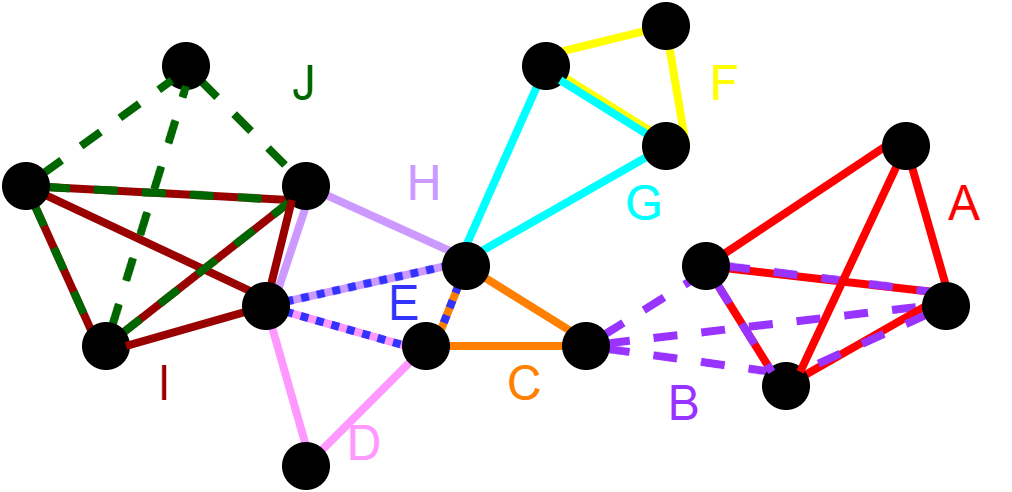


با تجمیع نتایج خواهیم داشت:



حال اگر به عنوان مثال حد آستانه را ۲۰ در نظر بگیریم، یک اجتماع با گره‌های ۱ تا ۴ و یک اجتمال با گره‌های ۵ تا ۹ خواهیم داشت. قطعا پس از این تقسیم مقدار میانگی (betweenness) تمام گره‌ها بیشتر نمی‌شود. پس فرآیند سلسله مراتبی همینجا می‌تواند پایان بیابد.

**ب)** ابتدا باید clique‌های ماکسیمال را پیدا کرد که در تصویر زیر نشان داده شده است:



سپس تعداد گره مشترک بین هر دو clique را در جدول زیر وارد می‌کنیم:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E | F | G | H | I | J |
| A | ۴ | ۳ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ |
| B | ۳ | ۴ | ۱ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ |
| C | ۰ | ۱ | ۳ | ۱ | ۲ | ۰ | ۱ | ۱ | ۰ | ۰ |
| D | ۰ | ۰ | ۱ | ۳ | ۲ | ۰ | ۰ | ۱ | ۱ | ۰ |
| E | ۰ | ۰ | ۲ | ۲ | ۳ | ۰ | ۱ | ۲ | ۱ | ۰ |
| F | ۰ | ۰ | ۰ | ۰ | ۰ | ۳ | ۲ | ۰ | ۰ | ۰ |
| G | ۰ | ۰ | ۱ | ۰ | ۱ | ۲ | ۳ | ۱ | ۰ | ۰ |
| H | ۰ | ۰ | ۱ | ۱ | ۲ | ۰ | ۱ | ۳ | ۲ | ۱ |
| I | ۰ | ۰ | ۰ | ۱ | ۱ | ۰ | ۰ | ۲ | ۴ | ۳ |
| J | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۱ | ۳ | ۴ |

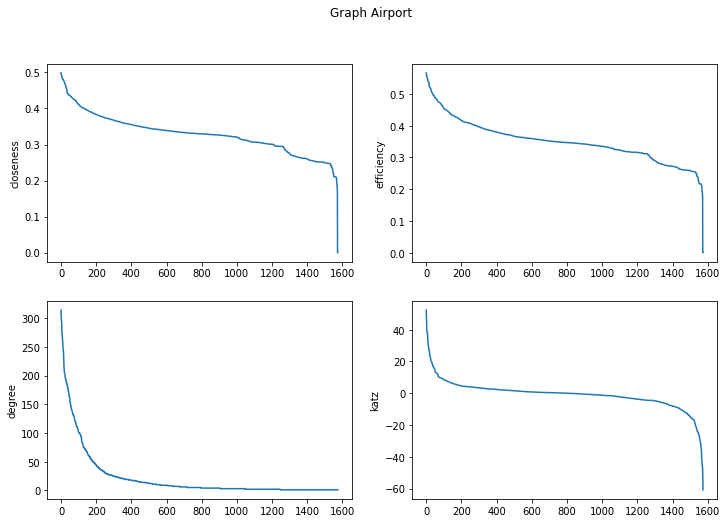
در گام بعد باید مقدار k را مشخص کنیم و متناسب با آن درایه‌های بیشتر مساوی k-1 را برابر با ۱ و مابقی را برابر با صفر قرار دهیم. چون مقدار k در صورت سوال مشخص نشده است به صورت شهودی آن را برابر با ۳ در نظر می‌گیریم و داریم:

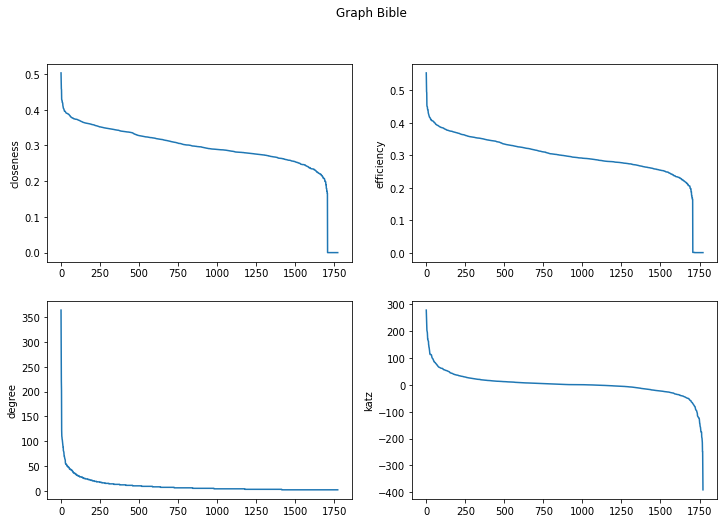
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E | F | G | H | I | J |
| A | ۱ | ۱ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ |
| B | ۱ | ۱ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ |
| C | ۰ | ۰ | ۱ | ۰ | ۱ | ۰ | ۰ | ۰ | ۰ | ۰ |
| D | ۰ | ۰ | ۰ | ۱ | ۱ | ۰ | ۰ | ۰ | ۰ | ۰ |
| E | ۰ | ۰ | ۱ | ۱ | ۱ | ۰ | ۰ | ۱ | ۰ | ۰ |
| F | ۰ | ۰ | ۰ | ۰ | ۰ | ۱ | ۱ | ۰ | ۰ | ۰ |
| G | ۰ | ۰ | ۰ | ۰ | ۰ | ۱ | ۱ | ۰ | ۰ | ۰ |
| H | ۰ | ۰ | ۰ | ۰ | ۱ | ۰ | ۰ | ۱ | ۱ | ۰ |
| I | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۱ | ۱ | ۱ |
| J | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۰ | ۱ | ۱ |

متناسب با این ماتریس A و B در یک اجتماع (Community)، F و G در یک اجتماع و مابقی در یک اجتماع دیگر قرار دارند.

سوال ۲

**ب)** به ترتیب برای دو گراف Airport و Bible داریم:





**ج)** می‌توان نکات زیر را از نمودارها متوجه شد:

* شاخص مرکزیت Degree با سه شاخص دیگر تفاوت بسیار زیادی دارد ولی سه تای دیگر رفتار کلی مشابهی داشته‌اند.
* در سه شاخص Closeness، Efficiency و Katz می‌بینیم که تعداد بسیار کمی از گره‌ها مرکزیت بسیار بالایی دارند و به وضوح از مابقی جدا شده‌اند و در حین حال تعداد هم بسیار مرکزیت پایینی دارند ولی اکثریت قریب به اتفاق مقداری نزدیک به هم داشته‌اند.
* در شاخص مرکزیت Degree گره‌ها با درجه مرکزیت پایین که متفاوت از عموم گره‌ها باشد دیده نمی‌شود.
* در معیار Katz‌ مقدار صفر برای اکثر گره‌ها بدست آمده است که می‌تواند یک حدآستانه مشخص باشد و اعداد منفی هم دیده می‌شود ولی چنین چیزی در سایر شاخص‌ها دیده نمی‌شود.

**د)** نتایج گراف Bible در جدول زیر آورده شده است.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ردیف | Closeness | | Efficiency | | Degree | | Katz | |
| نام | امتیاز | نام | امتیاز | نام | امتیاز | نام | امتیاز |
| ۱ | israel | 0.503 | Israel | 0.553 | israel | 364 | benjamin | 277 |
| ۲ | judah | 0.471 | Judah | 0.511 | judah | 254 | elam | 255 |
| ۳ | jerusalem | 0.458 | David | 0.495 | david | 221 | shem | 219 |
| ۴ | david | 0.456 | jerusalem | 0.492 | jerusalem | 202 | shemaiah | 213 |
| ۵ | egypt | 0.429 | Egypt | 0.454 | egypt | 122 | uz | 202 |
| ۶ | ephraim | 0.428 | ephraim | 0.450 | benjamin | 110 | aram | 198 |
| ۷ | manasseh | 0.425 | benjamin | 0.449 | manasseh | 108 | arphaxad | 194 |
| ۸ | benjamin | 0.425 | manasseh | 0.449 | ephraim | 104 | meshech | 193 |
| ۹ | joseph | 0.420 | moses | 0.441 | saul | 103 | lud | 180 |
| ۱۰ | moses | 0.420 | joseph | 0.441 | philistines | 100 | gether | 169 |

نتایج گراف Airport در جدول زیر آورده شده است.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ردیف | Closeness | | Efficiency | | Degree | | Katz | |
| نام | امتیاز | نام | امتیاز | نام | امتیاز | نام | امتیاز |
| ۱ | ATL | 0.498 | ATL | 0.566 | ATL | 314 | LAS | 52 |
| ۲ | LAX | 0.495 | LAX | 0.559 | IAD | 299 | SBN | 47 |
| ۳ | MSP | 0.495 | IAD | 0.557 | ORD | 296 | ECP | 40 |
| ۴ | DEN | 0.490 | MSP | 0.554 | LAX | 292 | ROA | 39 |
| ۵ | IAD | 0.490 | JFK | 0.552 | JFK | 291 | BWI | 38 |
| ۶ | JFK | 0.486 | ORD | 0.552 | DEN | 274 | DTW | 38 |
| ۷ | MCO | 0.485 | DEN | 0.552 | EWR | 273 | CRW | 38 |
| ۸ | ORD | 0.484 | EWR | 0.546 | MSP | 269 | IND | 36 |
| ۹ | EWR | 0.482 | IAH | 0.544 | IAH | 267 | SJU | 35 |
| ۱۰ | IAH | 0.481 | MCO | 0.539 | MIA | 261 | TOL | 32 |

توضیحات مربوط به فرودگاه‌های موجود در جدول قبل در این جدول ذکر شده است:

|  |  |
| --- | --- |
| کد | توضیحات |
| ATL | Atlanta, GA: Hartsfield-Jackson Atlanta International |
| BWI | Baltimore, MD: Baltimore/Washington International Thurgood Marshall |
| CRW | Charleston/Dunbar, WV: Yeager |
| DEN | Denver, CO: Denver International |
| DTW | Detroit, MI: Detroit Metro Wayne County |
| ECP | Panama City, FL: Northwest Florida Beaches International |
| EWR | Newark, NJ: Newark Liberty International |
| IAD | Washington, DC: Washington Dulles International |
| IAH | Houston, TX: George Bush Intercontinental/Houston |
| IND | Indianapolis, IN: Indianapolis International |
| JFK | New York, NY: John F. Kennedy International |
| LAS | Las Vegas, NV: McCarran International |
| LAX | Los Angeles, CA: Los Angeles International |
| MCO | Orlando, FL: Orlando International |
| MIA | Miami, FL: Miami International |
| MSP | Minneapolis, MN: Minneapolis-St Paul International |
| ORD | Chicago, IL: Chicago O'Hare International |
| ROA | Roanoke, VA: Roanoke Blacksburg Regional Woodrum Field |
| SBN | South Bend, IN: South Bend International |
| SJU | San Juan, PR: Luis Munoz Marin International |
| TOL | Toledo, OH: Toledo Express |

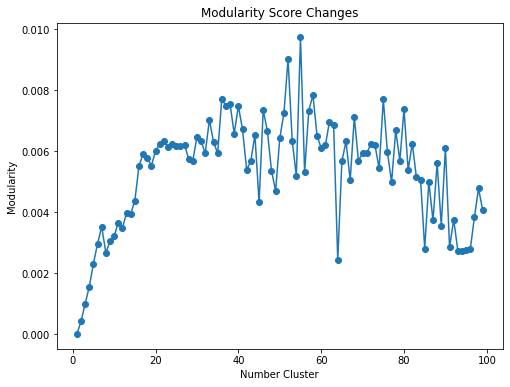
**ه)** آنچه از نتایج قسمت قبل بر می‌آید شباهت بسیار بالای ده گره برتر در سه روش Closeness، Efficiency و Degree و تفاوت آن با روش Katz است. به طوری که به عنوان مثال برای مجموعه‌داده Bible دو معیار اول لیست کاملا یکسان با ترتیب کمی متفاوت از هم دارند. پس به صورت شهودی می‌توان گفت یا سه معیار اول بهتر بوده است یا معیار Katz. از روی اسامی گراف Bible به نظر می‌رسد که اسامی یافت‌شده در سه معیار اول معروف‌تر و احتمالا کلیدی‌تر باشد. بنابراین به نظرم سه معیار اول مناسب هستند و بین آن‌ها من تفاوت خاصی را نمی‌توانم در نظر بگیرم.

در مورد گراف Airport هم نتایج سه معیار اول و هم معیار Katz عمدتا شامل فرودگاه‌های آمریکا است. برای مقایسه فرودگاه ATL به عنوان فرودگاه برتر در سه معیار اول را با فرودگاه LAS به عنوان فرودگاه برتر Katz می‌توان مقایسه کرد. با جستجو در اینترنت متوجه شدم که فرودگاه ATL از LAS بزرگ‌تر است. همچنین در جستجوی دیگر بزرگ‌ترین و شلوغ‌ترین فرودگاه‌ها را پیدا کردم که در این لیست دقیقا اسامی سه معیار اول دیده می‌شد. به بیان دقیق‌تر سه فرودگاه برتر عبارت بود از: ATL، LAX و ORD که توسط سه معیار اول پیشنهاد شده است. پس برای این مجموعه‌داده هم سه معیار اول را نسبت به معیار Katz مناسب‌تر می‌دانم.

سوال ۳

برای این سوال گراف Bible‌ و الگوریتم Kmeans به عنوان الگوریتم خوشه‌بند را در نظر گرفتم.

**ج)** برای تعیین تعداد خوشه مناسب می‌توان مقدار Modularity را در هر گام بررسی کرد. به ازای تعداد خوشه که این مقدار بیشینه شود، بهترین افراز را خواهیم داشت.‌ در تصویر زیر نتایج مربوط به این آزمایش آورده شده است که برای تعداد خوشه ۵۵ بهترین عدد حاصل شده است.



**د)** احتمالا منظور طراح سوال از Min-cut همان Cut بوده است؛ چراکه Min-cut معیاری برای تعیین کیفیت خوشه‌بندی نیست ولی Cut هست. لذا برای این قسمت Cut‌ را محاسبه خواهم کرد. برای ۵۵ خوشه امتیاز Modularity‌ برابر با ۰.۰۰۹۷ و برای Cut برابر با ۵۲۹۱ بدست آمده است.

سوال ۴

برای پیاده‌سازی این سوال از یک رویکرد بازگشتی استفاده کردم. به این شکل که هر بار با کمک روش Modularity Optimization‌ یک گراف را به دو زیر گراف تقسیم می‌کنم. سپس باید بررسی کرد که آیا Modularity حاصل از Modularity‌ پیشین بیشتر بوده است یا خیر. قطعا اگر کل گراف در یک جامعه باشد، مقدار Modularity برابر با صفر است و لذا تقسیم گراف به دو قسمت اگر Modularity را مثبت کند کافی است که این تقسیم انجام گیرد.

در گام بعد برای هر جز به صورت مستقل و بازگشتی تقسیم را می‌توان انجام داد چراکه مقدار ازدیاد یا کاهش ( و نه خالص) Modularity در هر بخش مستقل از بخش دیگر است. لذا در هر بخش اگر گراف اصلی را منحصر به آن بدانیم، باز مقدار اولیه نسبی Modularity‌ برابر با صفر است و اگر تقسیم آن به دو بخش آن را بهبود دهد، قطعا در گراف اصلی و کامل هم همین تاثیر مثبت را خواهد داشت.

مطابق نتایج عملی پس از تقسیم گراف به دو زیربخش شامل ۳۵۴۸ و ۳۵۶۷ گره امکان تقسیم‌های بیشتر برای گراف فراهم نبوده است و پس از آن Modularity دچار مشکل می‌شد. پس همین دو اجتماع را داریم. قاعدتا نمی‌توان نام شش هزار گره را در گزارش ذکر کرد ولی لیست آن‌ها در فایل پیاده‌سازی هست.