

به نام خدا



دانشگاه صنعتی امیرکبیر  
(پلی تکنیک تهران)

درس تحلیل شبکه‌های پیچیده  
استاد حقیرچهرقانی

تمرین دوم

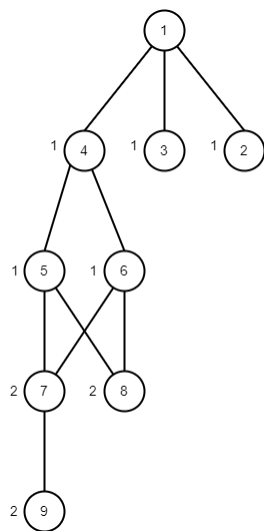
علیرضا مازوچی

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. طبیعتاً 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:



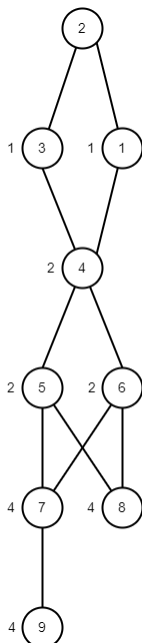
$$\delta_1(9) = 0$$

$$\delta_1(79) = 1, \delta_1(7) = 1, \delta_1(8) = 0$$

$$\delta_1(57) = \delta_1(67) = 1, \delta_1(58) = \delta_1(68) = 0.5, \delta_1(5) = \delta_1(6) = 1.5$$

$$\delta_1(45) = \delta_1(46) = 2.5, \delta_1(4) = 5, \delta_1(3) = \delta_1(2) = 0$$

$$\delta_1(14) = 6, \delta_1(13) = 1, \delta_1(12) = 1, \delta_1(1) = 8$$



$$\delta_2(9) = 0$$

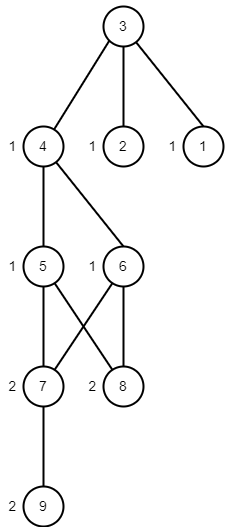
$$\delta_2(79) = 1, \delta_2(7) = 1, \delta_2(8) = 0$$

$$\delta_2(57) = \delta_2(67) = 1, \delta_2(58) = \delta_2(68) = 0.5, \delta_2(5) = \delta_2(6) = 1.5$$

$$\delta_2(45) = \delta_2(46) = 2.5, \delta_2(4) = 5$$

$$\delta_2(34) = \delta_2(14) = 3, \delta_2(3) = \delta_2(1) = 3$$

$$\delta_2(23) = \delta_2(12) = 4, \delta_2(2) = 8$$



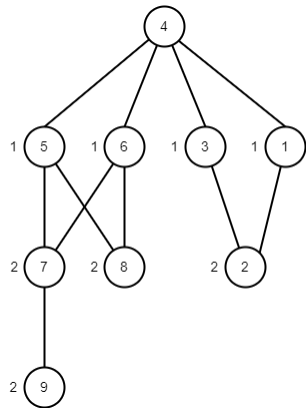
$$\delta_3(9) = 0$$

$$\delta_3(79) = 1, \delta_3(7) = 1, \delta_3(8) = 0$$

$$\delta_3(57) = \delta_3(67) = 1, \delta_3(58) = \delta_3(68) = 0.5, \delta_3(5) = \delta_3(6) = 1.5$$

$$\delta_3(45) = \delta_3(46) = 2.5, \delta_3(4) = 5, \delta_3(2) = \delta_3(1) = 0$$

$$\delta_3(34) = 6, \delta_3(23) = \delta_3(13) = 1, \delta_3(3) = 8$$



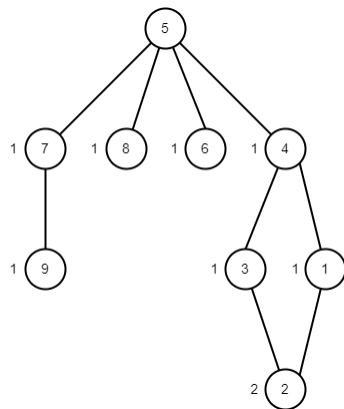
$$\delta_4(9) = 0$$

$$\delta_4(79) = 1, \delta_4(7) = 1, \delta_4(8) = \delta_4(2) = 0$$

$$\delta_4(57) = \delta_4(67) = 1, \delta_4(58) = \delta_4(68) = 0.5, \delta_4(5) = \delta_4(6) = 1.5$$

$$\delta_4(23) = \delta_4(12) = 0.5, \delta_4(1) = \delta_4(3) = 0.5$$

$$\delta_4(45) = \delta_4(46) = 2.5, \delta_4(34) = \delta_4(14) = 1.5, \delta_4(4) = 8$$

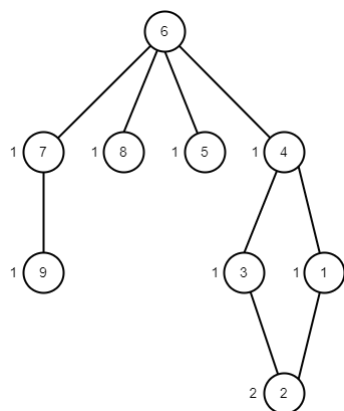


$$\delta_5(2) = 0$$

$$\delta_5(23) = \delta_5(12) = 0.5, \delta_5(1) = \delta_5(3) = 0.5, \delta_5(9) = 0$$

$$\delta_5(79) = 1, \delta_5(34) = \delta_5(14) = 1.5, \delta_5(7) = 1, \delta_5(4) = 3, \delta_5(6) = \delta_5(8) = 0$$

$$\delta_5(57) = 2, \delta_5(56) = \delta_5(58) = 1, \delta_5(45) = 4, \delta_5(5) = 8$$

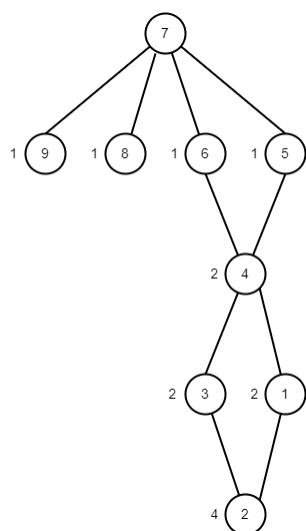


$$\delta_6(2) = 0$$

$$\delta_6(23) = \delta_6(12) = 0.5, \delta_6(1) = \delta_6(3) = 0.5, \delta_6(9) = 0$$

$$\delta_6(79) = 1, \delta_6(34) = \delta_6(14) = 1.5, \delta_6(7) = 1, \delta_6(4) = 3, \delta_6(5) = \delta_6(8) = 0$$

$$\delta_6(67) = 2, \delta_6(56) = \delta_6(68) = 1, \delta_6(46) = 4, \delta_6(6) = 8$$



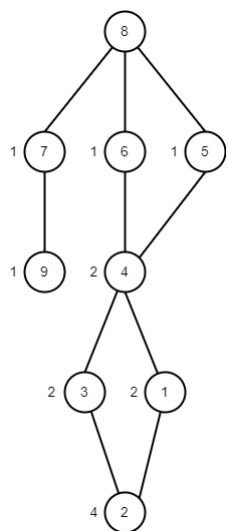
$$\delta_7(2) = 0$$

$$\delta_7(23) = \delta_7(12) = 0.5, \delta_7(1) = \delta_7(3) = 0.5$$

$$\delta_7(34) = \delta_7(14) = 1.5, \delta_7(4) = 3$$

$$\delta_7(46) = \delta_7(45) = 2, \delta_7(5) = \delta_7(6) = 2, \delta_7(8) = \delta_7(9) = 0$$

$$\delta_7(79) = \delta_7(78) = 1, \delta_7(67) = \delta_7(57) = 3, \delta_7(7) = 8$$



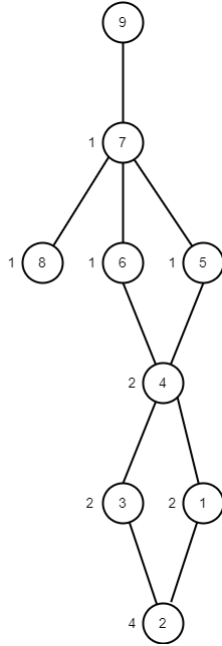
$$\delta_8(2) = 0$$

$$\delta_8(23) = \delta_8(12) = 0.5, \delta_8(1) = \delta_8(3) = 0.5$$

$$\delta_8(34) = \delta_8(14) = 1.5, \delta_8(4) = 3, \delta_8(9) = 0$$

$$\delta_8(46) = \delta_8(45) = 2, \delta_8(5) = \delta_8(6) = 2, \delta_8(79) = 1, \delta_8(7) = 1$$

$$\delta_8(78) = 2, \delta_8(58) = \delta_8(68) = 3, \delta_8(8) = 8$$



$$\delta_9(2) = 0$$

$$\delta_9(23) = \delta_9(12) = 0.5, \delta_9(1) = \delta_9(3) = 0.5$$

$$\delta_9(34) = \delta_9(14) = 1.5, \delta_9(4) = 3$$

$$\delta_9(45) = \delta_9(46) = 2, \delta_9(5) = \delta_9(6) = 2, \delta_9(8) = 0$$

$$\delta_9(78) = 1, \delta_9(57) = \delta_9(67) = 3, \delta_9(7) = 7$$

$$\delta_9(79) = 8, \delta_9(9) = 8$$

با تجميع نتایج خواهیم داشت:

$$bc(e) = \sum_{v \in V} \delta_v(e)$$

$$bc(12) = 1 + 4 + 0 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 = 8$$

$$bc(13) = 1 + 0 + 1 + 0 + 0 + 0 + 0 + 0 + 0 = 2$$

$$bc(14) = 6 + 3 + 0 + 1.5 + 1.5 + 1.5 + 1.5 + 1.5 + 1.5 = 18$$

$$bc(23) = 0 + 4 + 1 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 = 8$$

$$bc(34) = 0 + 3 + 6 + 1.5 + 1.5 + 1.5 + 1.5 + 1.5 + 1.5 = 18$$

$$bc(45) = 2.5 + 2.5 + 2.5 + 2.5 + 4 + 0 + 2 + 2 + 2 = 20$$

$$bc(46) = 2.5 + 2.5 + 2.5 + 2.5 + 0 + 4 + 2 + 2 + 2 = 20$$

$$bc(56) = 0 + 0 + 0 + 0 + 1 + 1 + 0 + 0 + 0 = 2$$

$$bc(57) = 1 + 1 + 1 + 1 + 2 + 0 + 3 + 0 + 3 = 12$$

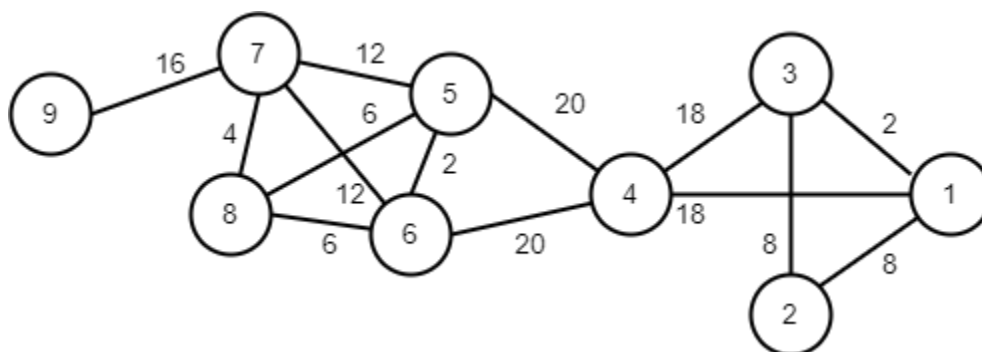
$$bc(58) = 0.5 + 0.5 + 0.5 + 0.5 + 1 + 0 + 0 + 3 + 0 = 6$$

$$bc(67) = 1 + 1 + 1 + 1 + 0 + 2 + 3 + 0 + 3 = 12$$

$$bc(68) = 0.5 + 0.5 + 0.5 + 0.5 + 0 + 1 + 0 + 3 + 0 = 6$$

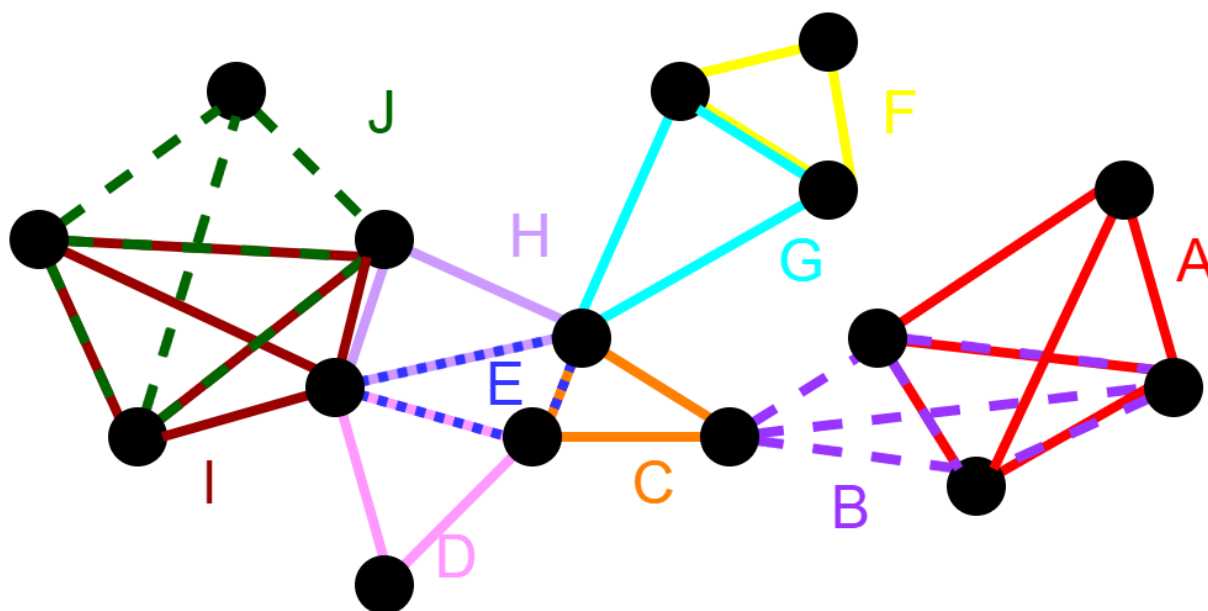
$$bc(78) = 0 + 0 + 0 + 0 + 0 + 0 + 1 + 2 + 1 = 4$$

$$bc(79) = 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 8 = 16$$



حال اگر به عنوان مثال حد آستانه را ۲۰ در نظر بگیریم، یک اجتماع با گره‌های ۱ تا ۴ و یک اجتماع 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:

J	I	H	G	F	E	D	C	B	A	
○	○	○	○	○	○	○	○	۳	۴	A
○	○	○	○	○	○	○	۱	۴	۳	B
○	○	۱	۱	○	۲	۱	۳	۱	○	C
○	۱	۱	○	○	۲	۳	۱	○	○	D
○	۱	۲	۱	○	۳	۲	۲	○	○	E
○	○	○	۲	۳	○	○	○	○	○	F
○	○	۱	۳	۲	۱	○	۱	○	○	G
۱	۲	۳	۱	○	۲	۱	۱	○	○	H
۳	۴	۲	○	○	۱	۱	○	○	○	I
۴	۳	۱	○	○	○	○	○	○	○	J

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:

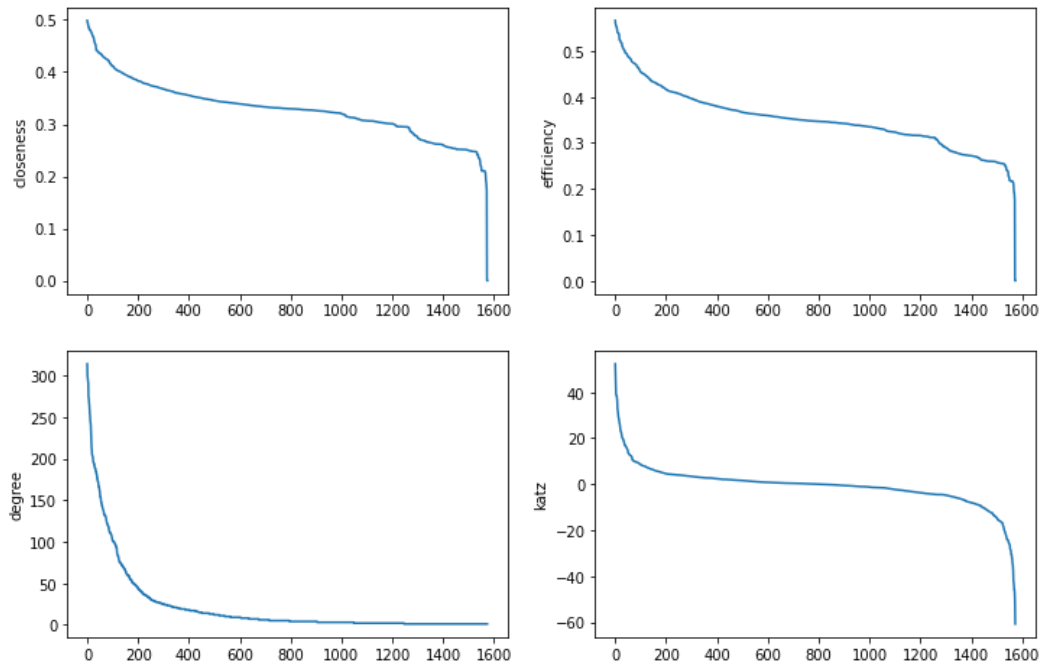
J	I	H	G	F	E	D	C	B	A	
○	○	○	○	○	○	○	○	۱	۱	A
○	○	○	○	○	○	○	○	۱	۱	B
○	○	○	○	○	۱	○	۱	○	○	C
○	○	○	○	○	۱	۱	○	○	○	D
○	○	۱	○	○	۱	۱	۱	○	○	E
○	○	○	۱	۱	○	○	○	○	○	F
○	○	○	۱	۱	○	○	○	○	○	G
○	۱	۱	○	○	۱	○	○	○	○	H
۱	۱	۱	○	○	○	○	○	○	○	I
۱	۱	○	○	○	○	○	○	○	○	J

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.

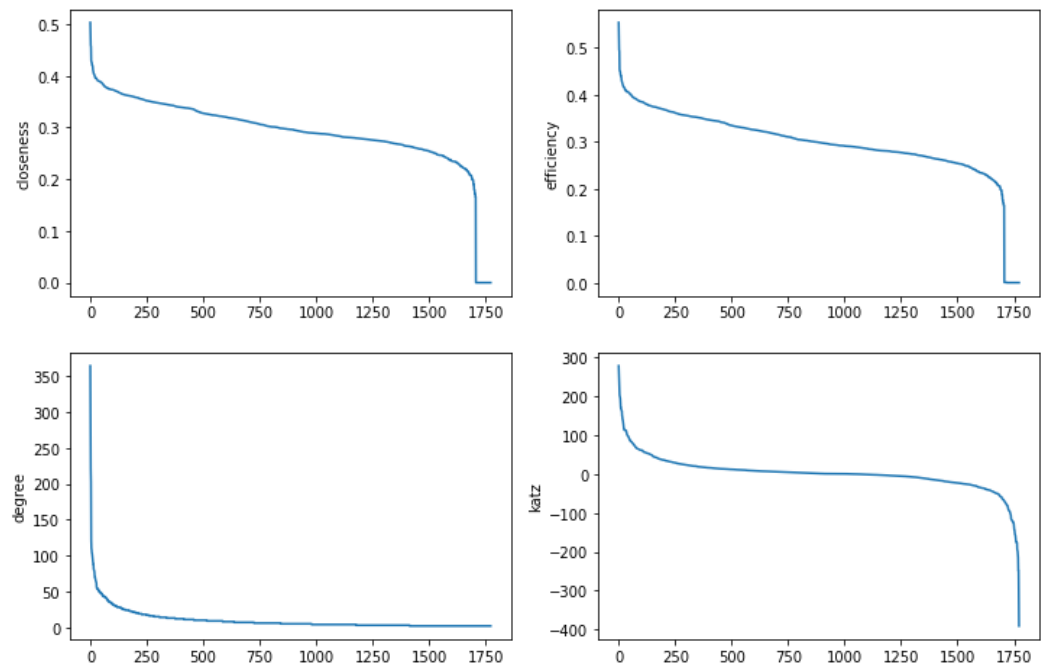
## سوال ۲

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

Graph Airport



Graph Bible





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	ephrain	0.428	ephrain	0.450	benjamin	110	aram	198
۷	manasseh	0.425	benjamin	0.449	manasseh	108	arphaxad	194
۸	benjamin	0.425	manasseh	0.449	ephrain	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.

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

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

توضیحات	کد
Atlanta, GA: Hartsfield-Jackson Atlanta International	ATL
Baltimore, MD: Baltimore/Washington International Thurgood Marshall	BWI
Charleston/Dunbar, WV: Yeager	CRW
Denver, CO: Denver International	DEN
Detroit, MI: Detroit Metro Wayne County	DTW
Panama City, FL: Northwest Florida Beaches International	ECP
Newark, NJ: Newark Liberty International	EWR
Washington, DC: Washington Dulles International	IAD
Houston, TX: George Bush Intercontinental/Houston	IAH
Indianapolis, IN: Indianapolis International	IND
New York, NY: John F. Kennedy International	JFK
Las Vegas, NV: McCarran International	LAS
Los Angeles, CA: Los Angeles International	LAX
Orlando, FL: Orlando International	MCO
Miami, FL: Miami International	MIA
Minneapolis, MN: Minneapolis-St Paul International	MSP

Chicago, IL: Chicago O'Hare International	ORD
Roanoke, VA: Roanoke Blacksburg Regional Woodrum Field	ROA
South Bend, IN: South Bend International	SBN
San Juan, PR: Luis Munoz Marin International	SJU
Toledo, OH: Toledo Express	TOL

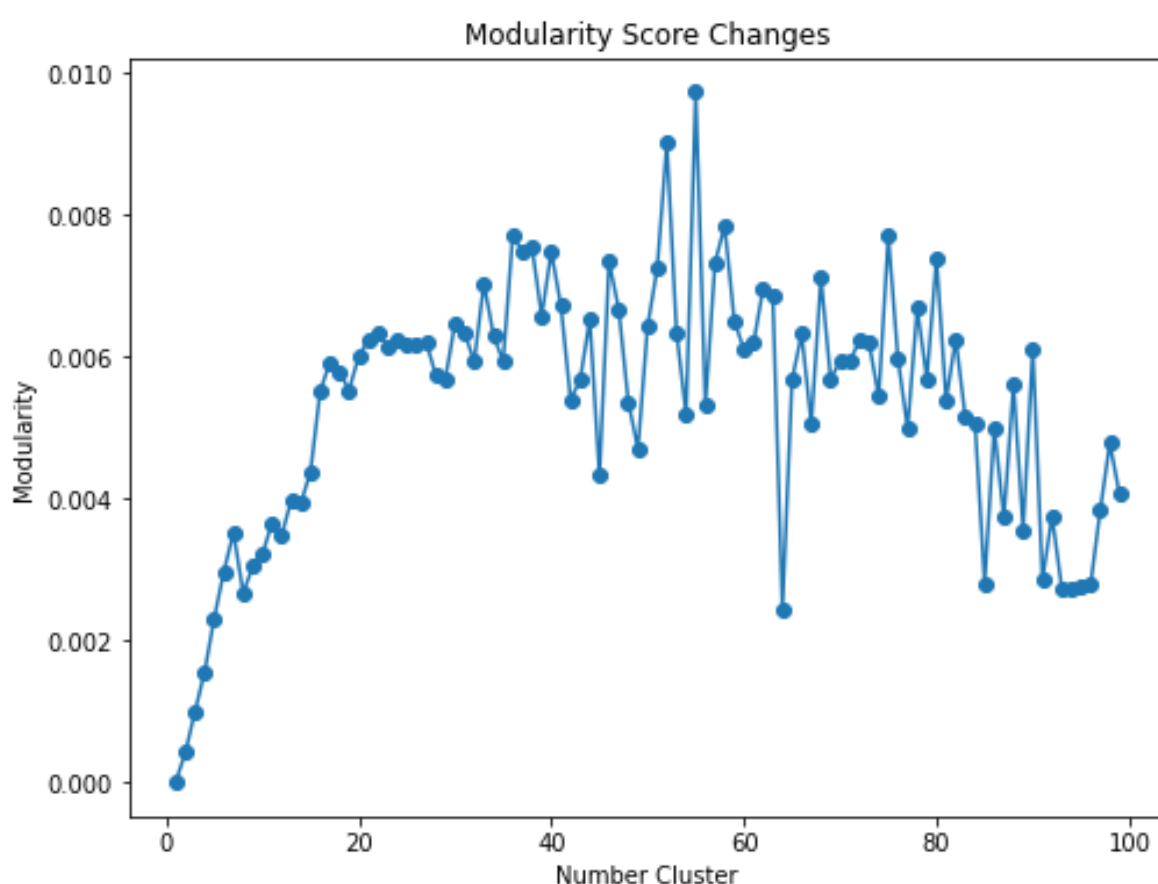
ه) آنچه از نتایج قسمت قبل بر می‌آید شباهت بسیار بالای ده گره برتر در سه روش Closeness، Efficiency و Degree و تفاوت آن با روش Katz است. به طوری که به عنوان مثال برای مجموعه داده Bible دو معیار اول لیست کاملاً یکسان با ترتیب کمی متفاوت از هم دارند. پس به صورت شهودی می‌توان گفت یا سه معیار اول بهتر بوده است یا معیار Katz. از روی اسامی گراف Bible به نظر می‌رسد که اسامی یافت‌شده در سه معیار اول معروف‌تر و احتمالاً 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 عمدتاً 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 دقیقاً 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 مجموعه داده هم سه معیار اول را نسبت به معیار Katz مناسب‌تر می‌دانم.

### سوال ۳

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

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) احتمالاً 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 خواهیم 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.