

به نام خدا



پروژه اول درس شبکه پیچیده پویا

استاد: دکتر فاطمی

علی سلیمی

## فهرست مطالب

3.....	دیتاست
4.....	شبکه استاتیک
5.....	شبکه پیچیده پویا
7.....	شبکه پیچیده پویا تجمیع شده
7.....	جنبه‌ی پیچیدگی (Complexity)
7.....	جنبه‌ی پویایی (Dynamics)
8.....	خروجی :
12.....	تحلیل نهایی

## دیتاست

لینک (<https://openmobilitydata.org/p/cdta>) به مجموعه داده‌های CDTA (Capital District Transportation Authority) در پلتفرم OpenMobilityData.org اشاره دارد. CDTA نهاد مسئول حملونقل عمومی در منطقه پایتخت ایالت نیویورک (شهرهای آلبانی، تروی، شنکتدی و مناطق اطراف) است. این مجموعه داده، اطلاعات مربوط به برنامه‌ها، مسیرها و ایستگاههای حملونقل عمومی (اتوبوسها و احتمالاً سایر خدمات) را در قالب GTFS (General Transit Feed Specification) ارائه میدهد.

جزئیات اصلی این دیتاست GTFS : قالب استاندارد برای اشتراک‌گذاری اطلاعات زمانی و مکانی سیستمهای حملونقل عمومی.

**محتوا:** مسیرهای اتوبوس و خطوط حملونقل. زمانبندی حرکت وسایل نقلیه. موقعیت ایستگاهها و توقفها. اطلاعات جغرافیایی (مختصات ایستگاهها). احتمالی: اطلاعات تعرفهها، تغییرات موقت مسیرها و غیره.

در پوشه gtfs فایل دیتاست دانلود شده موجود است.

ستون‌های کلیدی	شرح	فایل
stop_id, stop_name, stop_lat, stop_lon, location_type	فهرست همه‌ی ایستگاه‌ها (ترمینال‌ها) با مشخصات مکانی و توضیحات	stops.txt
trip_id, route_id, service_id, direction_id	معرفی هر سفر (trip) به همراه دوره‌ی سرویس‌دهی و نوع سرویس (weekday, saturday, sunday)	trips.txt
trip_id, stop_id, arrival_time, departure_time	زمان‌های ورود و خروج هر سفر به ایستگاه‌ها (برای مدل‌سازی بعد زمانی)	stop_times.txt
service_id, monday...sunday, start_date, end_date	تقویم سرویس‌های هفتگی: مشخص می‌کند هر service_id در کدام روزهای هفته فعال است و بازه‌ی تاریخی	calendar.txt

## شبکه استاتیک

نودها: (Nodes)

هر نود نشاندهنده یک ایستگاه حملونقل عمومی (مثلاً ایستگاه اتوبوس) است.

یالها: (Edges)

هر یال نشاندهنده اتصال مستقیم بین دو ایستگاه در یک مسیر (Trip) است. اگر دو ایستگاه در یک سفر (trip\_id) و بهصورت متوالی (stop\_sequence) قرار داشته باشند، بین آنها یال ایجاد میشود. یالها جهتدار نیستند (گراف غیرجهتی است).

ادغام دادهها: دادههای stop\_times (زمانبندی توقفها) و trips (اطلاعات سفرها) برای شناسایی توالی ایستگاهها در هر سفر ادغام میشوند. برای هر سفر (trip\_id)، ایستگاهها بر اساس stop\_sequence مرتب شده و یالها بین ایستگاههای متوالی ایجاد میشوند. توقفهای تکراری در یک سفر حذف میشوند (مثلاً اگر یک اتوبوس در یک ایستگاه دوباره توقف کند).

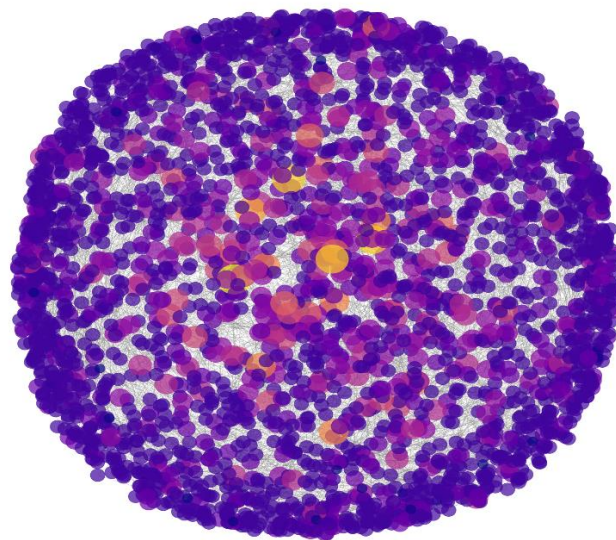
بهینهسازی: یالهای تکراری حذف میشوند (مثلاً اگر دو سفر مختلف بین دو ایستگاه مشترک باشند، فقط یک یال ایجاد میشود). گراف نهایی با استفاده از کتابخانه networkx ساخته میشود.

محدودیت‌های شبکه جهت حرکت: گراف غیرجهتی است، بنابراین جهت حرکت وسایل نقلیه (مثلاً رفت و برگشت) در نظر گرفته نشده است.

وزن یالها: یالها وزن ندارند (مثلاً فاصله بین ایستگاهها یا زمان سفر در نظر گرفته نشده است). در واقعیت، برخی اتصالات ممکنه مهمتر باشند (مثلاً یالهای با تردد بیشتر).

دادههای زمانبندی: این شبکه تنها اتصالات فیزیکی را نشان میدهد، نه فرکانس یا زمان سفر.

دانش‌مندیوب شریامون - مجموعه لایزن ولیم - کپ‌ش



شکل ۱

## شبکه پیچیده پویا

داده‌ها از فایل‌های calendar.txt, trips.txt, stop\_times.txt, stops.txt بارگذاری می‌شن.

هر سفر (trip) در stop\_times.txt به ترتیب ایستگاه‌ها بررسی میشه.

با توجه به زمان خروج از هر ایستگاه، ساعت (۰ تا ۲۳) و نوع روز weekday, saturday, Sunday مشخص میشه.

به ازای هر (روز، ساعت) یک گراف مجزا (snapshot) ساخته میشه که فقط شامل یال‌هایی هست که در اون ساعت فعال بودن.

یال‌ها دارای اطلاعات وزن هستند:

trip\_count: تعداد دفعاتی که سفر در اون مسیر در اون ساعت اتفاق افتاده.

time\_sum: مجموع زمان سفر بین اون دو ایستگاه در اون ساعت.

avg\_travel\_time: متوسط زمان سفر در یال) محاسبه شده بعد از ساخت کامل snapshot ها.

در نهایت، برای هر snapshot یک فایل graphml ذخیره میشه که قابل باز کردن در ابزارهایی مثل Gephi یا Cytoscape هست.

خروجی‌ها

فایل‌های graphml.

به ازای هر ترکیب از روز و ساعت، یک فایل گراف تولید میشه:

gtfs\_snapshot\_weekday\_hour\_07.graphml

gtfs\_snapshot\_saturday\_hour\_10.graphml

gtfs\_snapshot\_sunday\_hour\_15.graphml

...

درون هر فایل:

نودها (ایستگاه‌ها):

lon, lat, stop\_name, stop\_id

یال‌ها (سفر بین ایستگاه‌ها):

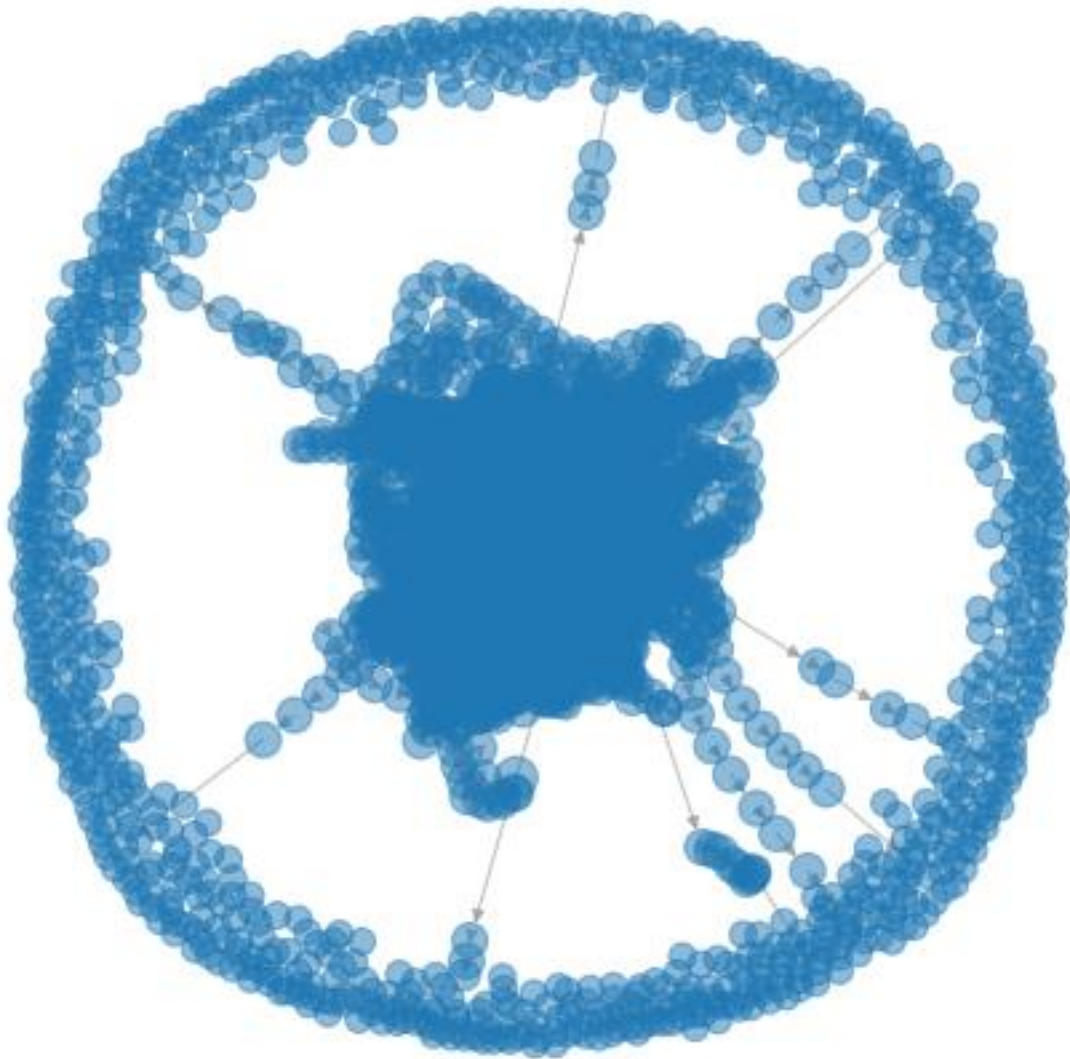
trip\_count: وزن یال بر اساس تعداد تکرار سفر

time\_sum: مجموع زمان سفرها

avg\_travel\_time: میانگین زمان سفر در آن ساعت

نمونه ای از گراف های تولید شده:

Saturday, Hour 13



شکل ۲

## شبکه پیچیده پویا تجميع شده

این قسمت تجميع همه snapshot ها در یک گراف واحد (agg\_graph) است:

نودها از قالب اولیه کپی می‌شوند. یال‌ها بر اساس trip\_count و time\_sum جمع می‌شوند  
avg\_travel\_time برای گراف کلی محاسبه می‌شود. در نهایت گراف تجميعی رسم شده و در  
gtfs\_aggregated.graphml ذخیره می‌گردد.

با این کار تمام ارتباطات زمانی مختلف در یک ساختار واحد قابل تحلیل هستند.

## جنبه‌ی پیچیدگی (Complexity)

- ساختار نامنظم و بزرگ
    - صدها تا هزاران ایستگاه (نود) و ده‌ها هزار یال داریم؛ توزیع درجه‌ی نودها یکنواخت نیست (عده‌ای ایستگاه «هاب» شده‌اند که تعداد یال زیادی دارند).
  - ویژگی‌های شبکه‌های پیچیده
    - Small-world: : میانگین کوتاه‌ترین مسیر بین هر دو ایستگاه معمولاً کوچک است (چند منقل).
    - Scale-free: : توزیع درجه غالباً پیروی از قانون توان است (چند نود خیلی پُرمرتبط، و اکثریت درجه‌ی کم).
    - خوشه‌بندی : ایستگاه‌هایی که در یک منطقه جغرافیایی یا مسیر مشترک قرار دارند، تمایل به ایجاد مثلث و خوشه‌سازی دارند.
  - وزن‌دار و جهت‌دار
    - یال‌ها جهت حرکت ( $A \rightarrow B$ ) دارند و وزنشان (تعداد سفرها و متوسط زمان سفر) مخزن اطلاعات رفتاری مسافران است.
- این خصوصیات باعث می‌شود گراف، یک شبکه‌ی پیچیده‌ی «واقعی» باشد؛ شبکه‌ای که نه کاملاً تصادفی است و نه کاملاً منظم.

## جنبه‌ی پویایی (Dynamics)

- منشأ پویایی
  - در ابتدا ما گراف‌های اسلایسی (snapshot) برای هر «نوع روز» (weekday/saturday/sunday) و هر «ساعت» (۰-۲۳) ساخته شده است.
  - در هر اسلایس، فقط یال‌هایی که در آن بازه‌ی زمانی فعال بودند (سفرهایی که در آن ساعت شروع می‌شوند) وجود داشتند.

- تجميع نهایی

- وقتی همه‌ی این snapshot ها را در agg\_graph با هم ادغام کردیم،

- تنها مقادیر تجميعی trip\_count و time\_sum روی یال‌ها باقی ماند؛

- بُعد «زمان» به شکل یک ویژگی عددی (تعداد سفرها در کل دوره) خلاصه شد.

- آیا پویایی حفظ شده؟

- خیر و بله

- خیر، چون تفکیک زمانی را از بین برده شده و حالا گراف «ثابت» است. دیگر نمی‌توانیم بگوییم «در ساعت ۸ صبح چه اتصالاتی فعال بودند.»

- بله، به این معنا که هنوز می‌توانیم با بررسی مقدار trip\_count، نواحی و مسیرهایی که در طول روزهای مختلف پرتردد بودند را بشناسیم. اما بُعد «چه روزی؟ چه ساعتی؟» را از دست داده‌ایم.

## خروجی :

Aggregated Graph: Nodes=2892, Edges=3501

--- Network Metrics (Barabasi Chap.2) ---

num\_nodes: 2892

num\_edges: 3475

density: 0.0008312629503591295

num\_components: 2

size\_largest\_cc: 2671

avg\_shortest\_path: 24.526543524076747

diameter: 94

avg\_clustering: 0.03741405359662618

transitivity: 0.06448461162677088

degree\_distribution: dictionary with 12 items

degree\_assortativity: 0.29580110664891185

degree\_centrality: dictionary with 2892 items

betweenness\_centrality: dictionary with 2892 items

closeness\_centrality: dictionary with 2892 items

eigenvector\_centrality: dictionary with 2892 items Network metrics saved.



تفسیر	مقدار	معیار
حدود ۳ هزار ایستگاه در شبکه (نود) داریم.	2892	(num_nodes) تعداد نودها
کمی بیشتر از ۳ هزار اتصال مستقیم بین ایستگاه‌هاست.	3475	(num_edges) تعداد یال‌ها
بسیار کم است؛ نشان می‌دهد شبکه است (sparse) فوق‌العاده نادر.	0.00083	(density) چگالی
شبکه به دو بخش جدا تقسیم شده است؛ بخش اصلی و یک بخش بسیار کوچک (معمولاً ایستگاه‌های ایزوله).	2	تعداد مؤلفه‌ها (num_components)
حدود ۹۲٪ از نودها در بخش اصلی متصل هستند.	2671	اندازه بزرگ‌ترین مؤلفه (size_largest_cc)
در بخش اصلی، متوسط فاصله بین دو ایستگاه ۲۴ گام (یال) است.	24.53	میانگین طول کوتاه‌ترین مسیر (avg_shortest_path)
طولانی‌ترین کوتاه‌ترین مسیر در بخش اصلی ۹۴ یال می‌باشد.	94	(diameter) قطر
نسبتاً کم؛ یعنی «دو همسایه» (دو ایستگاه) از یک ایستگاه مبدا به ندرت به هم وصل هستند.	0.0374	ضریب خوشه‌بندی متوسط (avg_clustering)
مشابه همان خوشه‌بندی، اما در سطح کل شبکه؛ مقدار پایین است.	0.0645	(transitivity) ترانزیتیو بودن
نشان می‌دهد بیشترین درجه‌ها در محدوده‌ی ۰ تا ۱۰ قرار دارند (در صورت نیاز می‌توان نمودار کشید).	مقدار مختلف 12	توزیع درجه (degree_distribution)

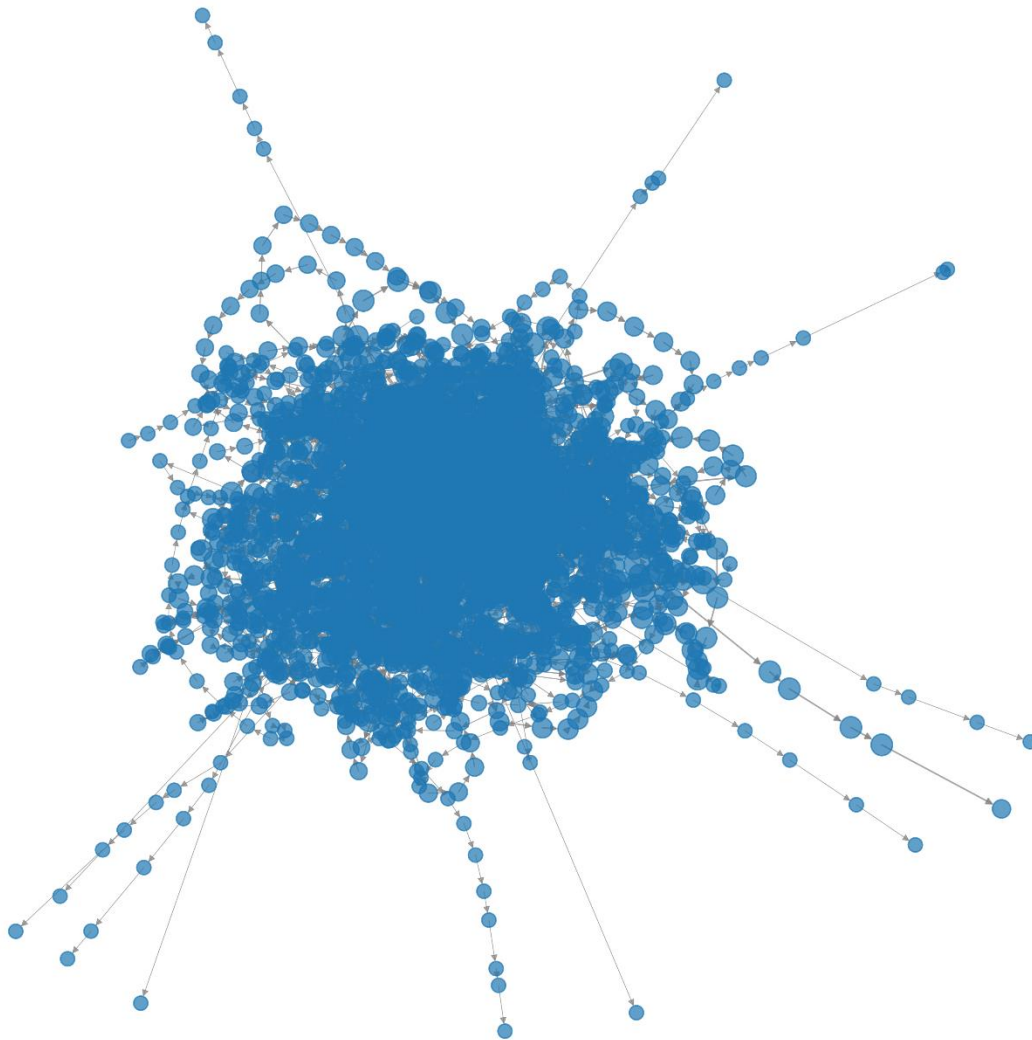
همبستگی درجه (degree_assortativity)	+0.296	مثبت بودنش یعنی ایستگاه‌های پرتراکم تمایل دارند به هم متصل شوند ("assortative mixing").
مرکزیت درجه (degree centrality)	2892	معیار پایه که سهم هر نود در کل اتصالات را می‌سنجد
مرکزیت بینابینی (betweenness centrality)	2892	نودهایی که بیشترین مسیرهای کوتاه را میان دو نود دیگر در خود جای می‌دهند
مرکزیت نزدیکی (closeness centrality)	2892	نودهایی که کمترین میانگین فاصله را تا سایر نودها دارند
مرکزیت ویژه‌برداری (eigenvector centrality)	2892	نودهای پیوندخورده به سایر نودهای مهم را امتیاز می‌دهد.

جدول ۲

چگالی بسیار پایین و تعداد اجزاء مثبت، نشان می‌دهد شبکه نامنظم و پراکنده است؛ اما وجود یک مؤلفه‌ی اصلی بزرگ (۹۲٪) و مؤلفه‌ی کوچک (۸٪ باقی) typisch برای شبکه‌های واقعی است. اگرچه میانگین مسیر ۲۴ یال کمی بالاست (به نسبت شبکه‌های شهری)، اما با توجه به تعداد نود، هنوز مسیرها نسبتاً کوتاه‌اند) درست مثل ویژگی "small-world" مثبت بودن degree\_assortativity نشان می‌دهد که ایستگاه‌های پرفت‌وآمد (هاب) تمایل دارند به هم متصل باشند؛ این هم رفتار متفاوتی نسبت به شبکه‌های کاملاً تصادفی است. مقدار کم خوشه‌بندی، نشان می‌دهد ایستگاه‌ها کمتر در مثلث قرار می‌گیرند—احتمالاً به دلیل هندسه‌ی مسیرها و عدم وجود حلقه‌های محلی زیاد. با استفاده از دیکشنری‌های مرکزیت می‌توان فهرست «تاپ ۱۰ ایستگاه» از زاویه‌های مختلف (degree, betweenness, closeness, eigenvector) استخراج کرد و آن‌ها را بررسی نمود.

شکل شبکه تجمیعی به این صورت است :

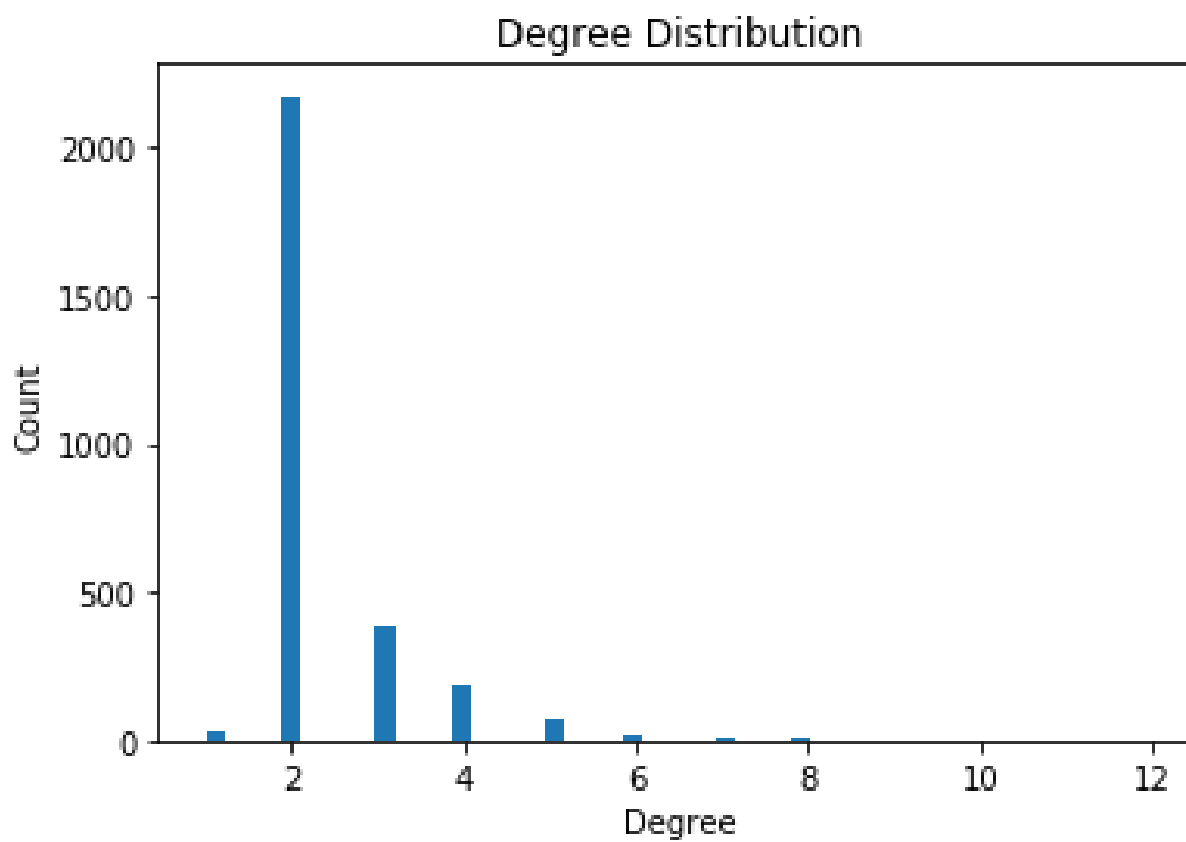
Aggregated GTFS Network (Vector & High DPI)



شکل ۳

## تحلیل نهایی

هیستوگرام توزیع درجه‌ی گراف نهایی :

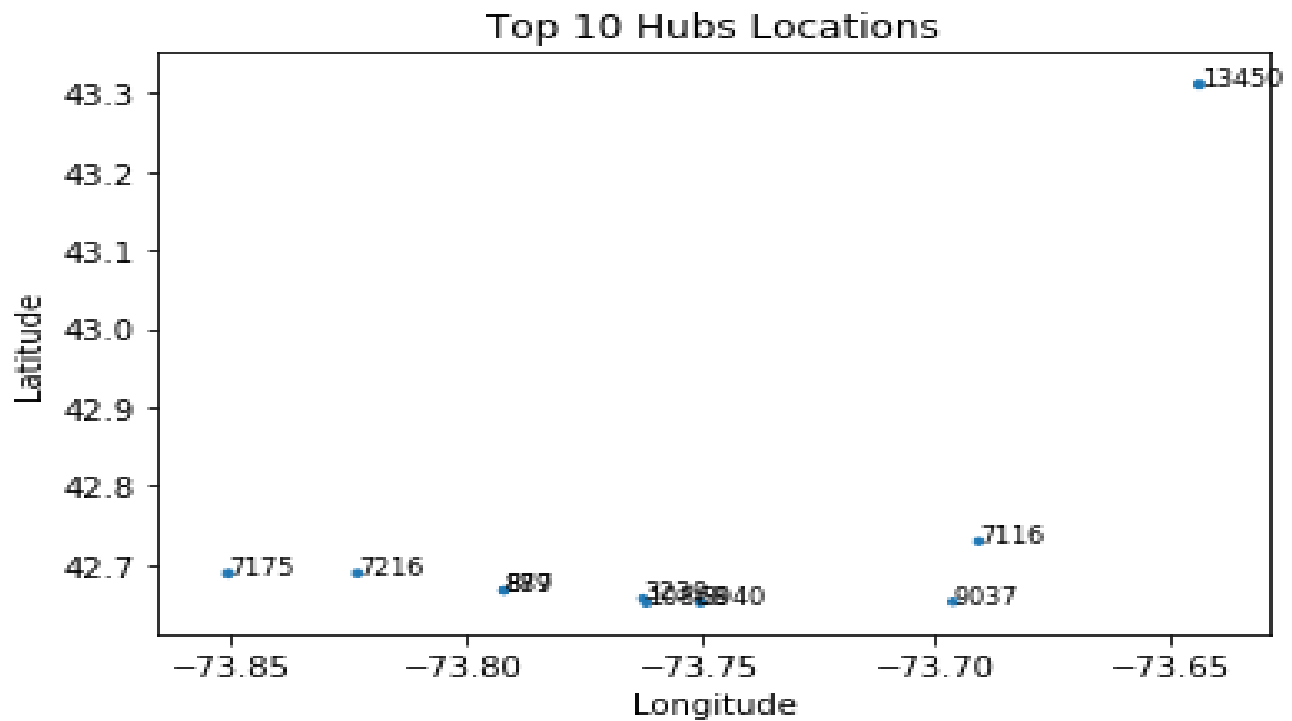


شکل ۴

استخراج ۱۰ ایستگاه hub برتر :

	stop_id	degree centrality	stop_name	stop_lat	stop_lon
0	13450	0.004151	Ridge Street Terminal	43.310816	-73.643825
1	7175	0.003805	Crossgates Mall Station	42.687953	-73.850299
2	3940	0.003459	Broadway & Steuben St	42.651258	-73.749925
3	10868	0.003459	Empire State Plaza Concourse	42.650790	-73.761390
4	7216	0.003113	SUNY Collins Circle	42.688238	-73.822915
5	877	0.002767	Allen/Madison Station - Western Ave & S. Allen St	42.666410	-73.791680
6	889	0.002767	Allen/Madison Station - Western Ave & S. Allen St	42.666560	-73.791700
7	3239	0.002767	Lark/Library Station - Washington Ave & Lark St	42.656161	-73.762142
8	7116	0.002767	Congress Station- 3rd St & Congress St	42.729008	-73.690805
9	9037	0.002767	Defreestville Park & Ride	42.652100	-73.696240

شکل ۵



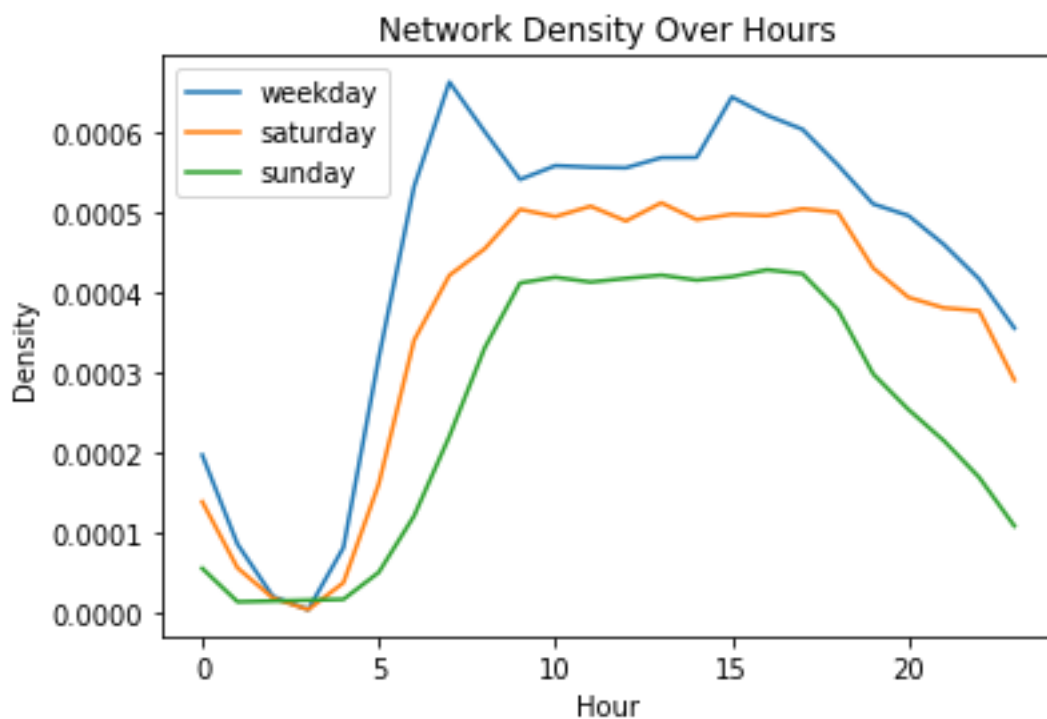
شکل ۶

مقایسه‌ی خلاصه‌ی آماری برای روزهای مختلف

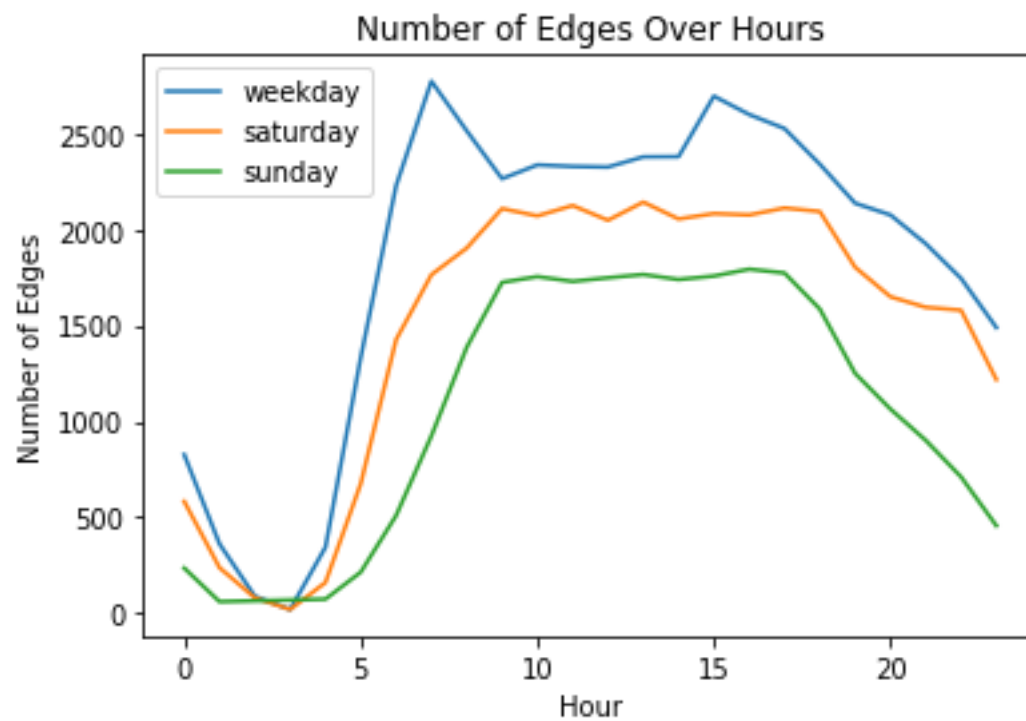
Summary Stats by Day Type:

	day	num_edges		density	
		mean	max	mean	max
0	saturday	1484.041667	2144	0.000355	0.000513
1	sunday	1143.318182	1794	0.000273	0.000429
2	weekday	1835.625000	2775	0.000439	0.000664

شکل ۷



شکل ۸



شکل ۹