

# 分时段新增公交线路规划系统

Based on NYC Taxi Trip Duration (2016)

Team: Readable,Writable,Executable

Google Girl Hackathon, Season 4

July 8, 2018

# Outline

- Hypothesis Proposal
- Data Cleaning
- Algorithms Analysis
  - Clustering Algorithm
  - Heuristic Search
- Expected Results

# Hypothesis Proposal

## 假设

- 假设在起始点与目的地之间有直达公交车时我们大概率选择公交车
- 假设公交车站的辐射区域为 1km
- 假设相同距离公交车行驶时间是出租车的两倍

# Data Cleaning

- 去除乘客数为 0 的数据
- 去除行驶距离小于 500m 的数据
- 去除行驶时间超过 5h 的数据

# Density Clustering

- 将纽约市地图按照  $100\text{m} \times 100\text{m}$  划分小 cell，每个打车地点归于某个 cell
- 按照每个 cell 里的人数由大到小排序
- 选择第一个未归类的 cell，设为关键点
- 找到它周围所有与关键点距离不超过  $X$ （此处设为  $1000\text{m}$ ）的点，归为该类
- 在归类的同时不断的调整计算关键点的位置
- 直到所有的点都被选择
- 对所有的 cell 进行分割，使其变成直径不超过  $Y$ （此处设为  $1\text{km}$ ）的 cluster

# Partial Result

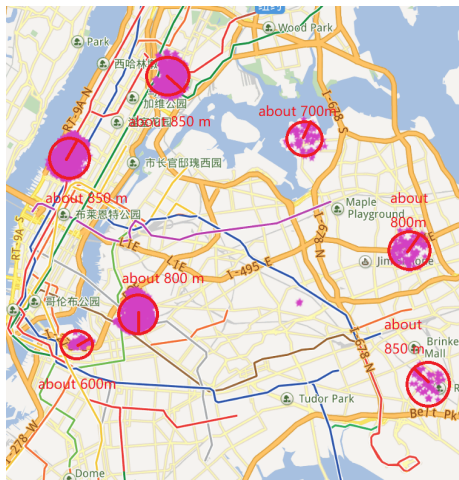


Figure: Seven Clusters from the dataset

# Density Clustering

密度聚类算法 (DBSCAN) 优于 k-means clustering 的地方

- 抗噪声, 能对簇的形状和大小进行约束
- 传统 k-means 需要人为确定要划分的聚类个数
- k-means 可能会导致选出的公交车站点在河里楼里不可能的地方

# Route Planning

## Main Purpose

得到了公交站点经筛选后的候选建造位置后，在可接受的时间内提供一个在结构合理的前提下最大化受益人数的新增公交线路方案。

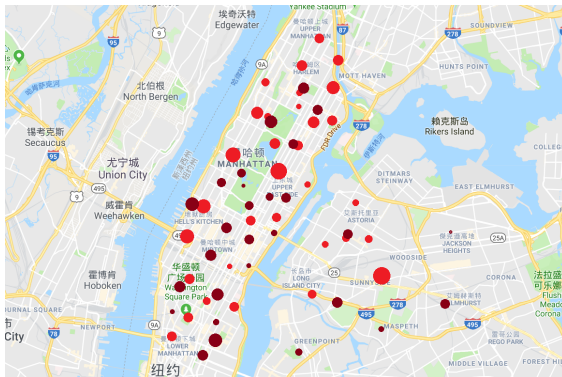
## We want to:

- 设计出的站台数量合理
- 两两站台间距离合理
- 公交线路总长度在一定范围内
- 线路的非直线系数不应过大
- 最大化需求被新增线路覆盖的人数



# Determination of the starting point

## Route Planning



用  $P_i$  表示与团  $i$  相邻的边的边权和，以  $\frac{P_i}{\sum_{j=1}^n P_j}$  的概率选到点  $i$  作为起点，随机选取  $k$  个团作为起点，分别跑确定起点的线路寻找算法。

# Candidate Bus Route Generation

## Probability based Spreading Algorithm

每生成一条路线，都遵循下面的流程：

- 令  $B$  表示已选择的站点的编号队列，初始时， $|B| = 1, B_1 = s$ 。
- 每次用队列中最后一个点寻找下一个站点，确定后加入  $B$  队列。
- 重复上面的步骤，直到无法加入新的站点或者已有的线路过长。
- 将新生成的线路与目前最优线路进行比较，并更新。

# Choose Next Station

## Roulette Wheel Selection Algorithm

令  $u = |B|$ ,  $v$  表示待确定的下一个站点,  $v$  能被作为一个候选后继站点所需要满足的条件有:

- $Dist(B_u, v) < MaxDist$
- $Dist(B_u, s) < Dist(v, s)$
- $Dist(B_i, v) \geq Dist(B_u, v)$ , for  $i \in [1, u)$

# Choose Next Station

## Roulette Wheel Selection Algorithm

获得了一个候选后继站点集合  $G$  后，我们按照每个点的累计客流量来分配概率随机选择。

令  $pf(x, y)$  表示从站点  $x$  出发到站点  $y$  的客流量。

$$P(v) = \frac{\sum_{j=1}^u pf(B_j, v)}{\sum_{i=1}^{|G|} \sum_{j=1}^u pf(B_j, G_i)}$$

## Expected Results

经过多次验证，在增加一条线路的情况下，约有 16 万人可以获得便利，占给出数据的 10%，效果明显  
期望以后给出更多算法改进，同时获得更多更准确的公交线路推荐

## Address

Newworld.science:8001

## Reference

Chao Chen<sup>†</sup>, Daqing Zhang<sup>†</sup>, Zhi-Hua Zhou<sup>‡</sup>, Nan Li<sup>‡</sup>, , Tülin Atmaca<sup>†</sup>, and Shijian Li, B-Planner: Night Bus Route Planning using Large-scale Taxi GPS Traces