

作业四：居民出行 OD 计算

王逸

一、作业要求

1、任务

- (1) 计算居民出行 OD 需求；
- (2) 展示 OD 分布图，开展交通需求分布的初步分析

2、方法

重力模型，相关参数自行设定（可进行简要对比分析）

3、数据

- (1) 深圳市 10 个行政区分布图，含人口（Sum_PF_HOM）和就业岗位（Sum_PF_WOR）统计数据（GIS 格式）；
- (2) 高德/百度在线地图，含道路网和各区府所在地（在线查询）；
- (3) 自行构建各区行政中心点（政府所在地），为点类数据图层，自行查询地图并建立要素类（feature class），并通过高德/百度计算点对之间的最短道路距离；
- (4) 自行构建行政中心点对之间的直线（期望线），并分类显示计算的 OD 量（行政中心点之间的 OD 量通过重力模型计算获得）。

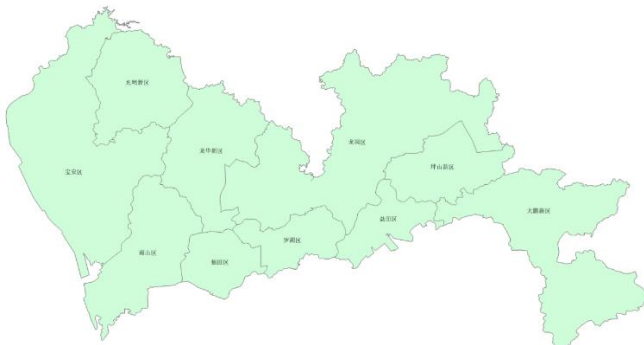


图 1 行政区

DISTRICT	SUM_PF_HOME	SUM_PF_WORK
宝安区	3299524	2087343
大鹏新区	120109	97006
福田区	1373259	1182417
光明新区	621427	385189
龙岗区	3082458	1939281
龙华新区	1875521	1131172
罗湖区	968618	692963
南山区	1547116	1129545
坪山新区	316362	194773
盐田区	218788	152100

图 2 属性表

二、预处理

(1) 数据裁剪与预处理：使用 OpenStreetMap (OSM) 提供的路网数据，通过空间裁剪方法将数据限定在深圳市的行政边界内。对于裁剪过程，使用了 Shapely 库与 Geopandas 工具，将路网数据与深圳市边界进行空间相交操作。

(2) 区政府坐标的获取与最近邻查找：深圳市各区政府的地理位置通过百度地图人工搜索整理成经纬度表。假设每个区政府的坐标为 $P_i = (x_i, y_i)$ ，其中 $i \in [1, n]$ 表示深圳市各个行政区的索引。使用最近邻查找算法将这些坐标点映射到路网的线性结构上，计算每个区政府坐标 P_i 到所有路网边 $e_j \in E$ 的最小距离：

$$d(P_i, e_j) = \min_{p \in e_j} \|P_i - p\|$$

其中， $\|P_i - p\|$ 表示点 P_i 到边 e_j 上点 p 的欧几里得距离。该过程通过 Python 库 Shapely 中提供的 `nearest_points` 函数进行实现，将坐标点 P_i 投影到最近的路网边 e_j 上，形成新的坐标点 $P_{i\text{new}}$ ，即区政府点在路网上的投影。

(3) 最短路径计算与距离矩阵构建：通过上述步骤得到所有区政府坐标点在路网上的投影，可以进一步计算各区政府之间的最短路径。最短路径计算采用 Dijkstra 算法，可以有效解决图 G 中两个顶点之间的最短路径问题。设定每条边的权重为路网边的实际距离，实际距离根据 Haversine 公式计算得到，Haversine 公式用于计算地球上两个经纬度点之间的球面距离，公式如下：

$$d = 2R \cdot \arcsin \left(\sqrt{\sin^2 \left(\frac{\Delta\phi}{2} \right) + \cos(\phi_1) \cdot \cos(\phi_2) \cdot \sin^2 \left(\frac{\Delta\lambda}{2} \right)} \right)$$

其中， d 是两点之间的球面距离， R 是地球的平均半径（取 6371km）， $\Delta\lambda = \lambda_2 - \lambda_1$ 是经度差， $\Delta\phi = \phi_2 - \phi_1$ 是纬度差， $\lambda_2 - \lambda_1$ 是经度， $\phi_2 - \phi_1$ 是纬度。随后使用 Dijkstra 算法求解区政府之间的最短路径矩阵：

$$D_{ij} = \min_{\pi \in \Pi_{ij}} \sum_{e_k \in \pi} w(e_k)$$

其中， π 表示从区政府到区政府的所有路径集， $w(e_k)$ 表示路径中每条边的权重，即路网边的长度。通过对每一对区政府坐标点进行最短路径计算，最终得到一个距离矩阵 D ，其中每个元素 D_{ij} 表示区政府 i 到区政府 j 之间的最短路程。矩阵的大小为 $n \times n$ ，其中 n 为深圳市行政区的数量。

以上步骤通过 Python 中的 Geopandas、Shapely、NetworkX 等库实现。具体操作步骤包括：使用 Geopandas 进行空间数据裁剪；使用 Shapely 进行最近邻查找和点到线的投影操作；使用 NetworkX 库实现图结构并应用 Dijkstra 算法计算最短路径。通过这一方法，我们能够得到深圳市区政府之间的最短路径距离矩阵，为后续的 OD 计算提供了数据支持。

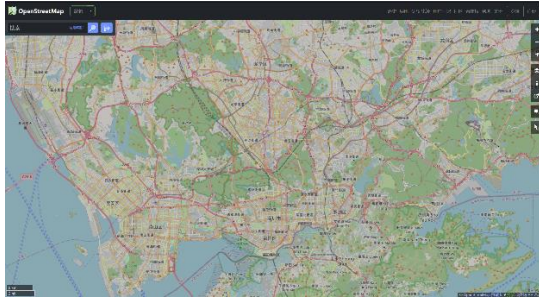


图 3 openstreet 官网界面

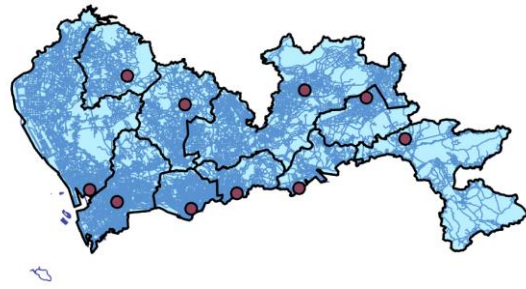


图 4 路网及行政区点效果

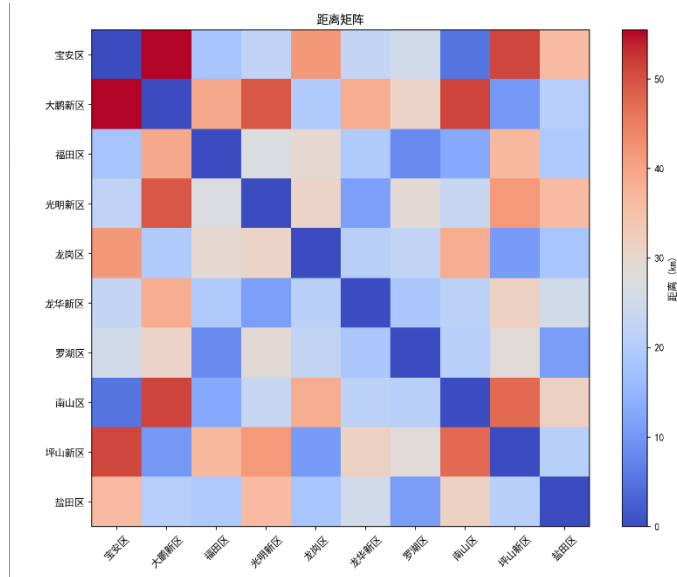


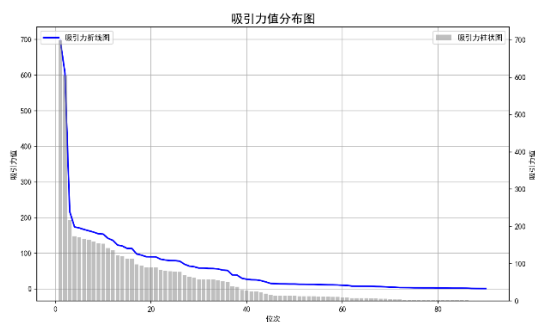
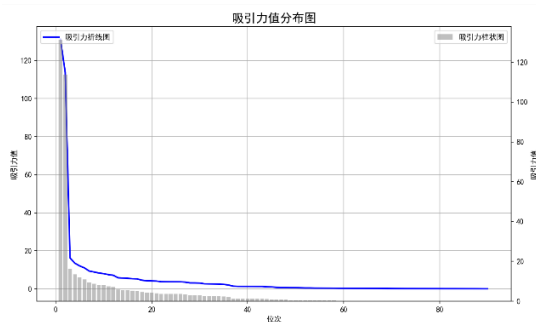
图 5 距离矩阵热图

三、重力模型及其改进型

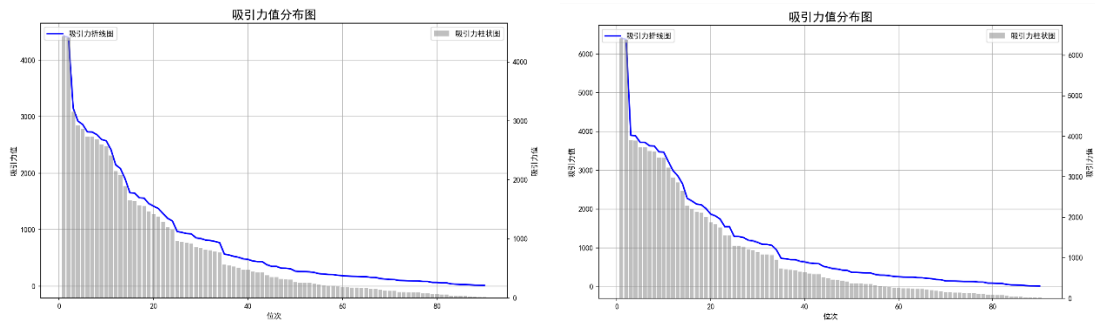
重力模型 (Gravity Model) 是一种基于区域间相互作用强度的预测模型，广泛应用于交通流、物流、贸易和区域规划中。在 OD 图计算中，重力模型用于估计两个区域间的出行流量 (OD 值)。其基本假设类似于物理学中的万有引力定律：两个区域间的相互作用强度与其属性规模成正比，与其距离成反比。重力模型的一般公式为：

$$T_{ij} = k \cdot \frac{P_i^\alpha \cdot Q_j^\beta}{D_{ij}^\gamma}$$

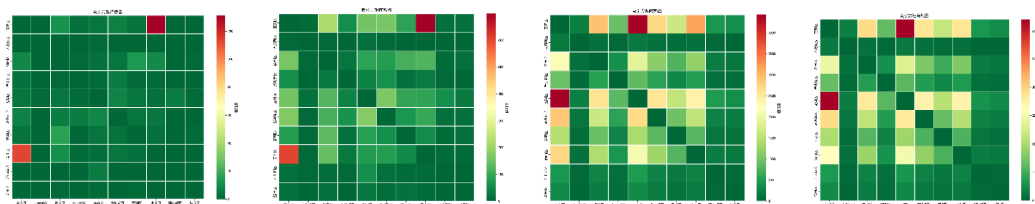
其中， T_{ij} ：区域 i 和区域 j 之间的流量 (即 OD 值)。 P_i ：区域 i 的吸引因子 (如人口)。 Q_j ：区域 j 的吸引因子 (如就业岗位)。 D_{ij} ：区域 i 和区域 j 之间的距离 (如地理距离、时间距离等)。 α 、 β ：分别是 P_i 和 Q_j 的影响系数。 γ ：距离衰减系数。 k ：比例常数。



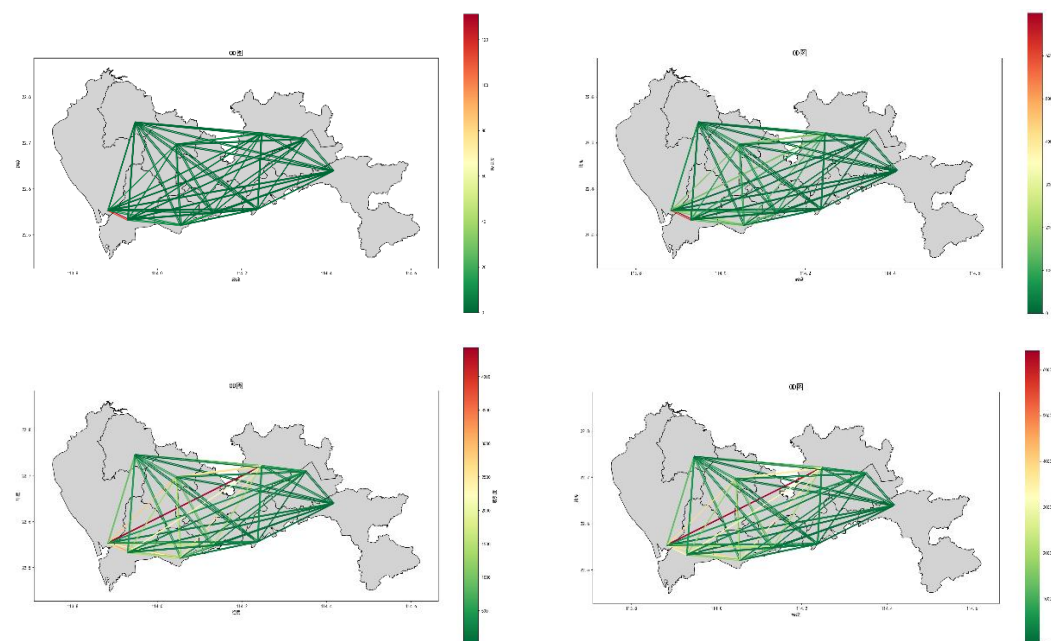
作业四：居民出行 OD 计算
王逸



距离衰减系数在 2、1、0.1、0.01 情况下的吸引力值按位次分布图



距离衰减系数在 2、1、0.1、0.01 情况下的吸引力值矩阵热图



距离衰减系数在 2、1、0.1、0.01 情况下的 OD 图

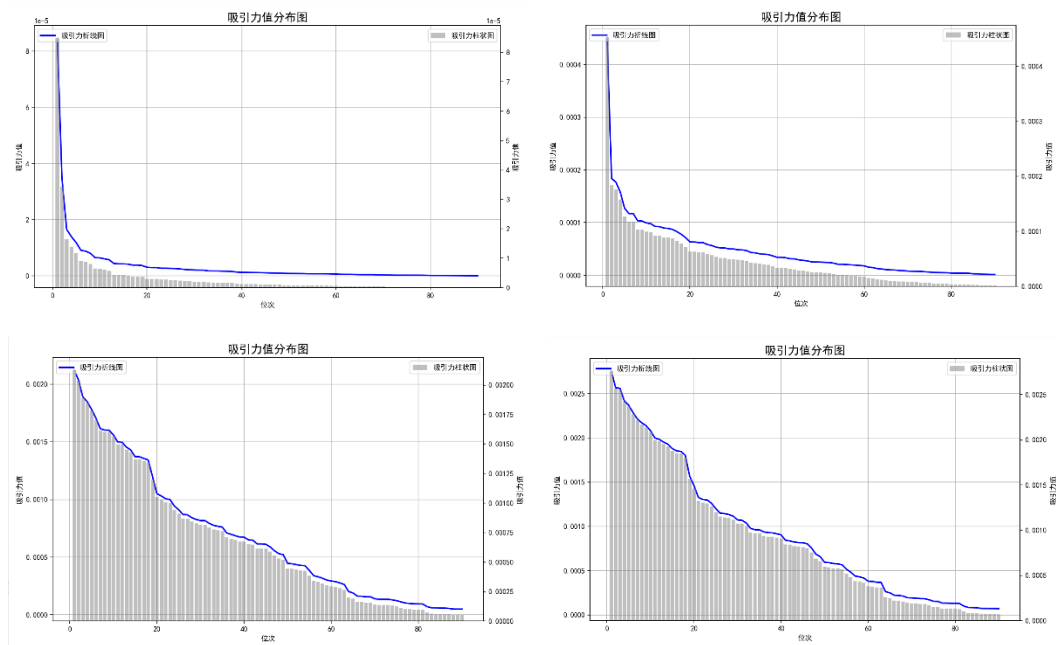
可以调整不同的距离衰减系数观察不同的吸引力数值计算情况。距离衰减系数的大小可以直接决定距离因素对于吸引力的影响，较大的距离衰减系数说明距离影响较大，反之则说明距离影响较小。根据吸引力值按位次分布图，可以发现吸引力数值呈现非常显著的幂律分布规律，大多数的连接都呈现较小的值，而较高的值只有零星的分布，这在下方的吸引力值矩阵热图中也有体现，其中的偏红色的值基本只有几小块。但是在适当减小距离衰减系数之后，幂律分布的显著性有所下降，不同数值的差别也慢慢具有了层次性。根据吸引力值按位次分布图，随着距离衰减系数的下降，重力模型吸引力的差别性逐渐扩大，这在矩阵热图中也有所体现，其中可以发现在距离衰减系数减小的情况下，龙岗区和宝安区的吸引力最大，因为两者的人口和就业岗位都非常多，尽管地理距离较远，但是人口和就业岗位形成的吸引

力影响更大。

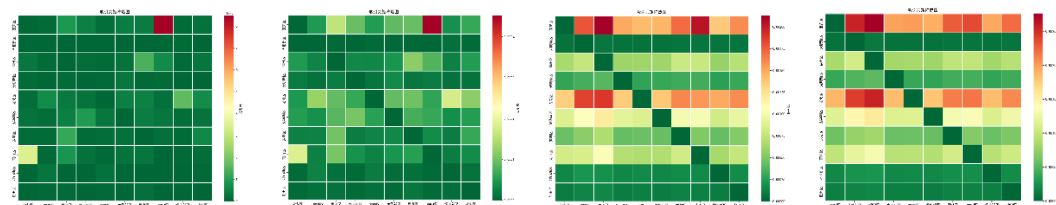
但实际上，这种传统的重力模型只是单一以两者的绝对数量进行计算，忽略了区域人口与岗位的匹配关系。如果一个区域人口本身很多，可能其大部分岗位已经被区域内人口占据，对外部人群的吸引力相对较低。这种现象说明单纯依赖岗位数量作为吸引因子可能会导致模型偏差。因此，存在一种重力模型改进型，这种模型可以比较好地考虑区域人口与岗位的匹配关系。以下是其一般公式：

$$T_{ij} = k \cdot \frac{P_i \cdot \left(\frac{Q_j}{P_j}\right)}{D_{ij}^\gamma}$$

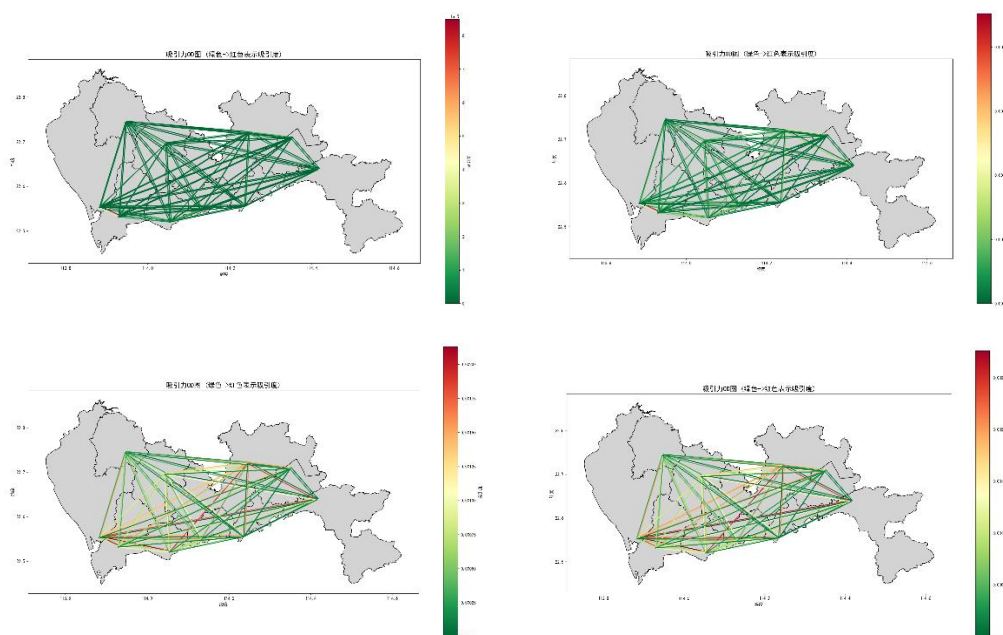
其中， P_i ：区域 i 的人口。 Q_j/P_j ：区域 j 的岗位竞争因子（每人可分配的岗位数量）。 D_{ij} ：区域 i 到 j 的距离。 γ ：距离衰减系数，用于调整距离对流量的影响程度。 k ：归一化常数，用于确保计算结果与实际流量规模一致。



距离衰减系数在 2、1、0.1、0.01 情况下的吸引力值按位次分布图



距离衰减系数在 2、1、0.1、0.01 情况下的吸引力值矩阵热图



距离衰减系数在 2、1、0.1、0.01 情况下的 OD 图

经过改进之后的重力模型相对于原本的重力模型而言，对比程度有了较大提升。其距离衰减系数调整的规律与传统重力模型相似，但是其幂律分布的显著性下降了很多。从吸引力值矩阵来看，在距离衰减系数较大时和传统重力模型差不多，但是在距离衰减系数减小时不同行政区之间的吸引力数值有了比较好的区分度。在最后的 OD 图中，尤其是在距离衰减系数处在 0.1 和 0.01 时 OD 图也能看出不同地区之间的吸引力差别。

根据最新的重力模型结果，存在几个吸引力较高的行政区对，比如福田区与宝安区之间、宝安区与南山区之间、龙岗区与福田区之间，这实际上也比较符合深圳的实际情况。而且还可以发现一些有趣的现象，比如说宝安区和龙岗区容易被其它区吸引，这是因为这些行政区的居住人口非常多，而且邻接周边的行政区，同时这些行政区的就业岗位也不完全满足该行政区居住人口的需求，因此有很多人口被吸引到周边的行政区就业。此外，还可以发现坪山新区、盐田区、大鹏新区这几个区被吸引的程度较低，这几个区普遍距离核心行政区较远，且居住人口比较有限。

五、参考文献

- [1]Maniadakis, D., & Varoutas, D. (2014). Network congestion analysis of gravity generated models. *Physica A: Statistical Mechanics and Its Applications*, 405, 114-127.
- [2]Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., ... & Tomasini, M. (2018). Human mobility: Models and applications. *Physics Reports*, 734, 1-74.
- [3]Haynes, K. E., & Fotheringham, A. S. (2020). Gravity and spatial interaction models.
- [4] 宁鹏飞.(2019).交互式网络地图可视化技术在 OD 数据中的应用研究(硕士学位论文,武汉大学). https://kns.cnki.net/kcms2/article/abstract?v=IX50VqaRjdWkVtw-rvG6MxlVh390yyt5CUZ6IREgXwkOhb7WPG3nMZUC8mAgAaRjryCAqyqqSLYed3JCJeHRPO_WnV6X9QdiiigJoaCJu98JaAw8q_9F7rcPDpDP4GyhbPeaTWAAjBY=&uniplatform=NZKPT&language=CHS

[5]李锐,刘朝辉 & 吴华意.城市人口流动感知与建模方法综述.武汉大学学报(信息科学版)1-28.doi:10.13203/j.whugis20240082.