# 公交网络拓扑结构分析

为了分析公交网络的拓扑性质，本文以青岛市和潍坊市为例，选取青岛427条公交线路，潍坊140条线路，构建基于站点的无向图。该方法以站点为节点，可达的公交线路为边。本文从节点的度分布，平均路径长度和聚集系数的角度对公交网络的拓扑结构进行分析。

## 公交路线爬取

1. 接口简介：公交路线的爬取依据高德地图提供的API，使用方法参阅：[公交路径规划](https://lbs.amap.com/api/webservice/guide/api/direction" \l "bus)
2. 流程简介：当我们访问高德提供的URL时，服务器会给我们返回一个json格式的字符串，我们所需要的信息都蕴含在字符串中。
3. URL重要参数：

city：城市

keywords：公交线路，例如，1路。

1. 响应结果重要参数：

name：公交线路名字

start\_stop：始发站

end\_stop：终点站

polyline：当前公交路线的坐标集，格式为字符串。

busstops：

name：沿途站名

sequence：当前是第几站

由于潍坊和青岛境内没有公交线路超过1000路，所以我们分别遍历了两个城市前1000路公交，为了确保公交路线的完整性(这里没有考虑小众公交，例如，旅游专线之类的)。

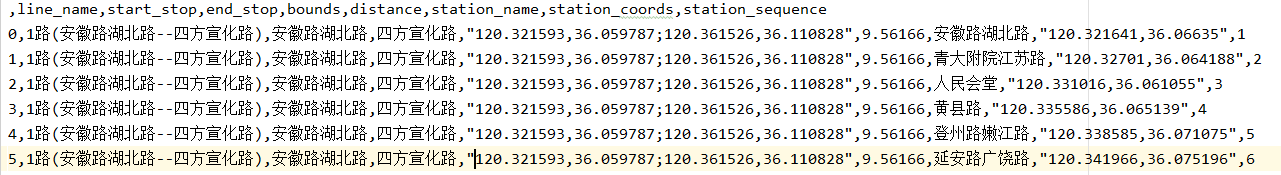


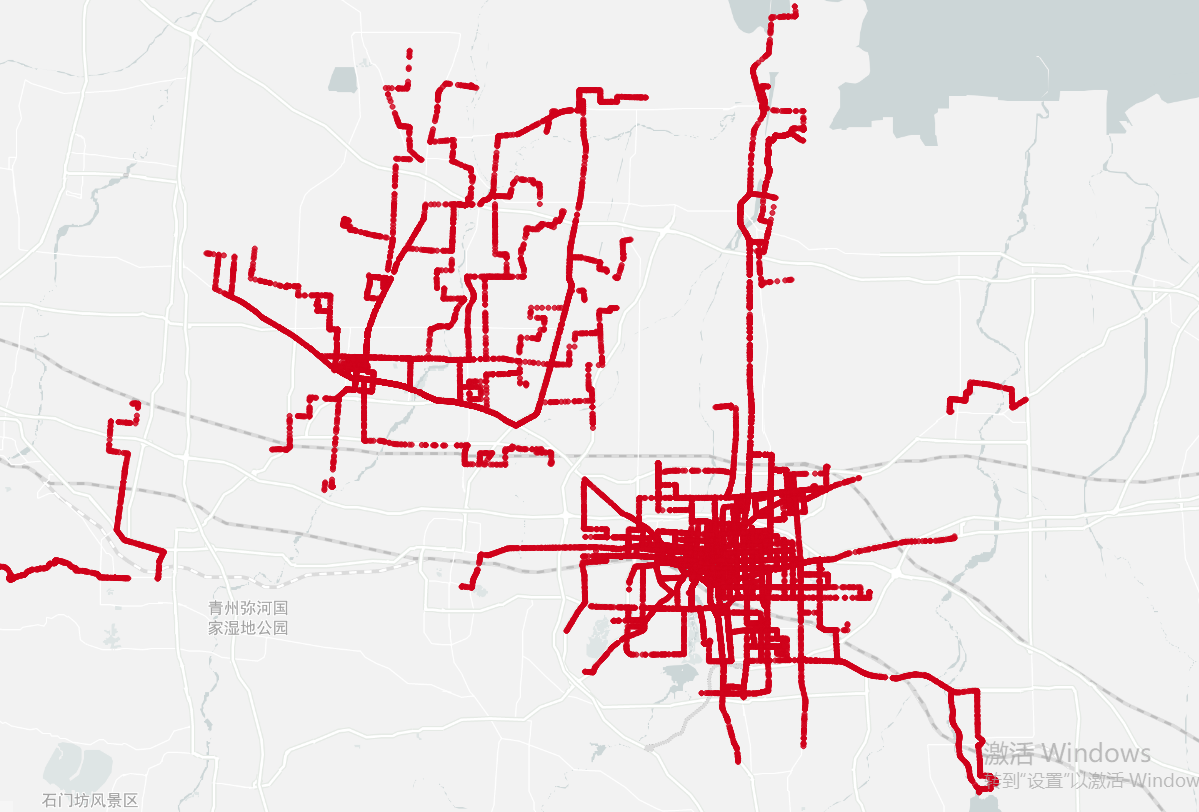
图1-1 公交信息

由于我们基于站点来构建公交网络，所以沿途站点名称至关重要。另外，我们还要区分当前站点属于哪一列公交，对于属于同一列公交的站点，我们要对其顺序连边。

## 公交路线网络的可视化

由于我们把爬取到的公交线路都存储在一个文件中(公交线路太多，不适合分开存储)。所以，可视化公交线路最重要的就是要区分当前站点属于哪一列公交。我们通过记录站点的位序，来区分公交线路。每一条公交线路，对起始站从0(当前站数-1)开始编号，一直到终点站。我们可以通过比较下一行数据的位序是否大于上面一行，如果大于，则属于同一条公交线路；否则，属于另外一条线路。

为了验证爬取的公交信息是属于潍坊市的公交信息，我们利用[高德开放平台| Map Lab](https://maplab.amap.com/dev/mapdata)对潍坊市的轨迹数据进行可视化，青岛市同理。

图2-1 潍坊市部分公交轨迹路线可视化

当我们验证了网络信息的有效性，我们就需要基于站点建立网络结构，为分析网络的拓扑性质做准备。

我们建立的是一个无向无权简单图(没有重边和自环)。虽然在真实的公交路线网络中，两列不同的公交有一段路程是重合的，即网络有重边，这种情况很常见，但是这会加大分析的难度。因为两列公交走同一段路，花费的时间等权重信息可能是不同的。所以我们建立的是一个无向无权简单图。

同时，一个无向无权简单图，易于分析聚类系数，平均路径长度等性质。

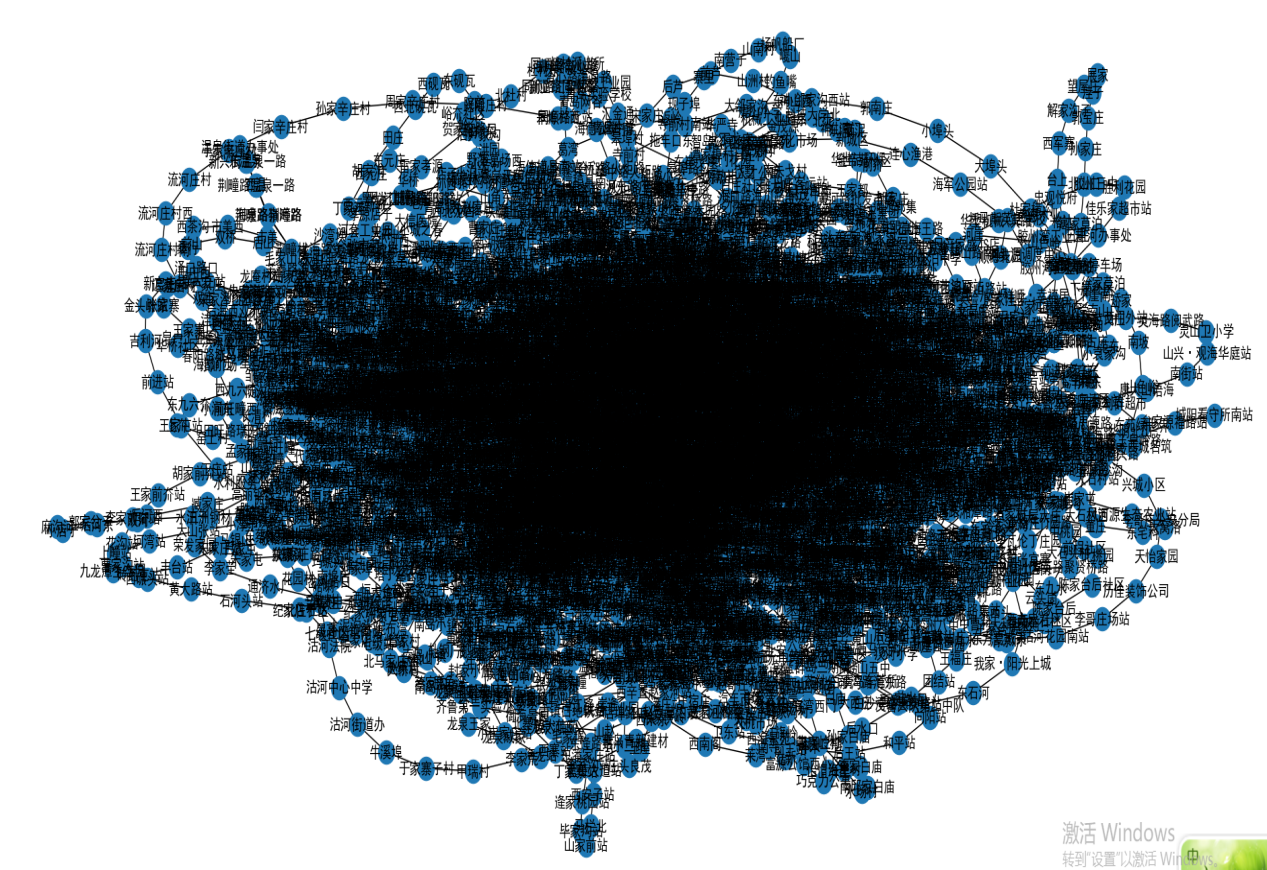


图2-2 青岛市站点网络结构图

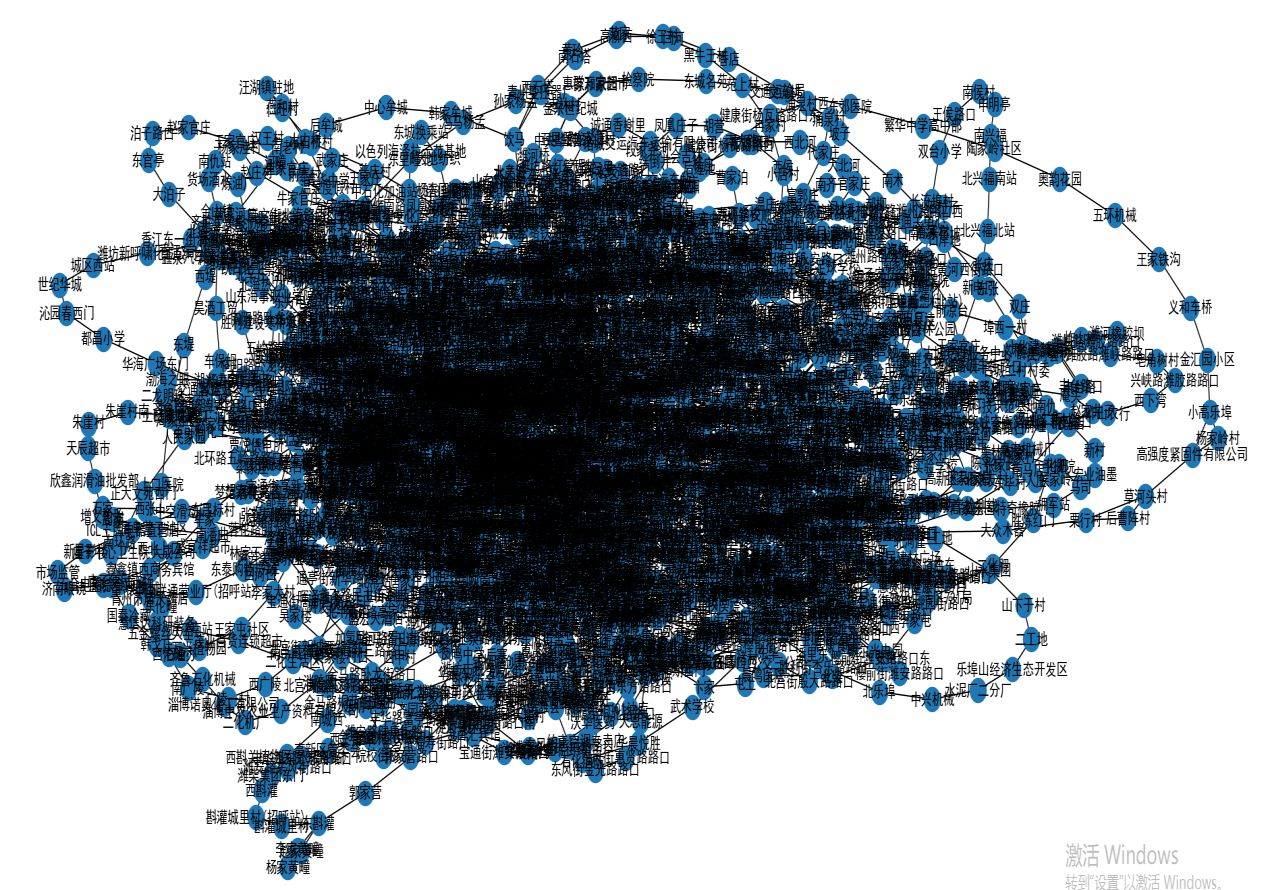


图2-3 潍坊市站点网络结构图

## 公交路线的拓扑结构性质及其实际意义

### 聚类系数

聚集系数是用来度量节点的邻居节点之间的连接紧密程度。

节点i的聚类系数：



Ei是节点i的Ki个邻居之间实际存在的边数。Ki个节点之间最多有Ki(Ki - 1)条边。

网络的聚类系数：



网络的聚类系数越大，说明节点与附近各个节点连接紧密度越高，即当前站点周围的公交线路越密集。站点的聚类系数越大，说明当前节点的载客能力越强。青岛市公交网络的聚集系数约为0.039，潍坊市公交网络的聚集系数约为0.046。实际上，青岛市(427次公交)所拥有的公交车次，远远多于潍坊市(140次公交)的公交车次，并且青岛市(11293平方千米)的占地面积小于潍坊市(15859平方千米)。说明青岛市的公交网络还有很大的优化空间。

### 平均路径长度

节点i和节点j之间的距离dij定义为连接这两个节点的最短路径上的边的数目。网络的平均路径长度L定义为任意两个节点之间的距离的平均值：



N为节点数。

青岛市公交网络中的节点数有13万多，平均路径长度约为11.885；潍坊市公交网络中的节点数有4万多，平均路径长度约为10.831。两个网络中的节点都很多，但是平均路径长度却很短，具有明显的小世界特性。同时，数据可以佐证，网络中的节点数越多，即网络越稀疏，平均路径长度越大；反之，网络越稠密，平均路径越小。在一个比较稀疏的网络中，添加少量的边，即稍微稠密化稀疏网络，就能够显著降低平均路径长度。这是因为一条边的添加，对网络的影响是非线性的，它不仅影响这条边直接相连的节点，也会影响到当前节点的近邻，次近邻等。

在基于站点的公交网络中，平均路径长度反映的是任意两个站点之间平均有多少个站点，这是评价公交网络便利性的一个重要指标。

|  |  |  |
| --- | --- | --- |
|  | 青岛 | 潍坊 |
| 聚集系数 | 0.039232082522731686 | 0.046184444028007324 |
| 平均路径长度 | 11.885076329570312 | 10.831521582108738 |

表3-1 聚类系数和平均路径长度

### 度分布

分析得出青岛市公交复杂网络中节点的最大度为24，潍坊市公交复杂网络中节点的最大度为64。虽然两个城市的公交路线从数量到长度等信息完全不同，但是两个城市公交网络的度分布都基本满足幂律分布，满足无标度网络的特性。数量众多的节点分布在1-10度范围之内，只有极少的节点拥有很大的度。说明大部分的站点与小于10个站点进行连接，也就意味着，有少于5次公交经过这些站点。只有极少数的站点，度比较大。站点的度越大，说明站点有更好的连通性，这些站点可能是交通枢纽。极少数的节点发挥着重要的连接作用，从直观上讲，这在我们日常生活中也是合理的。

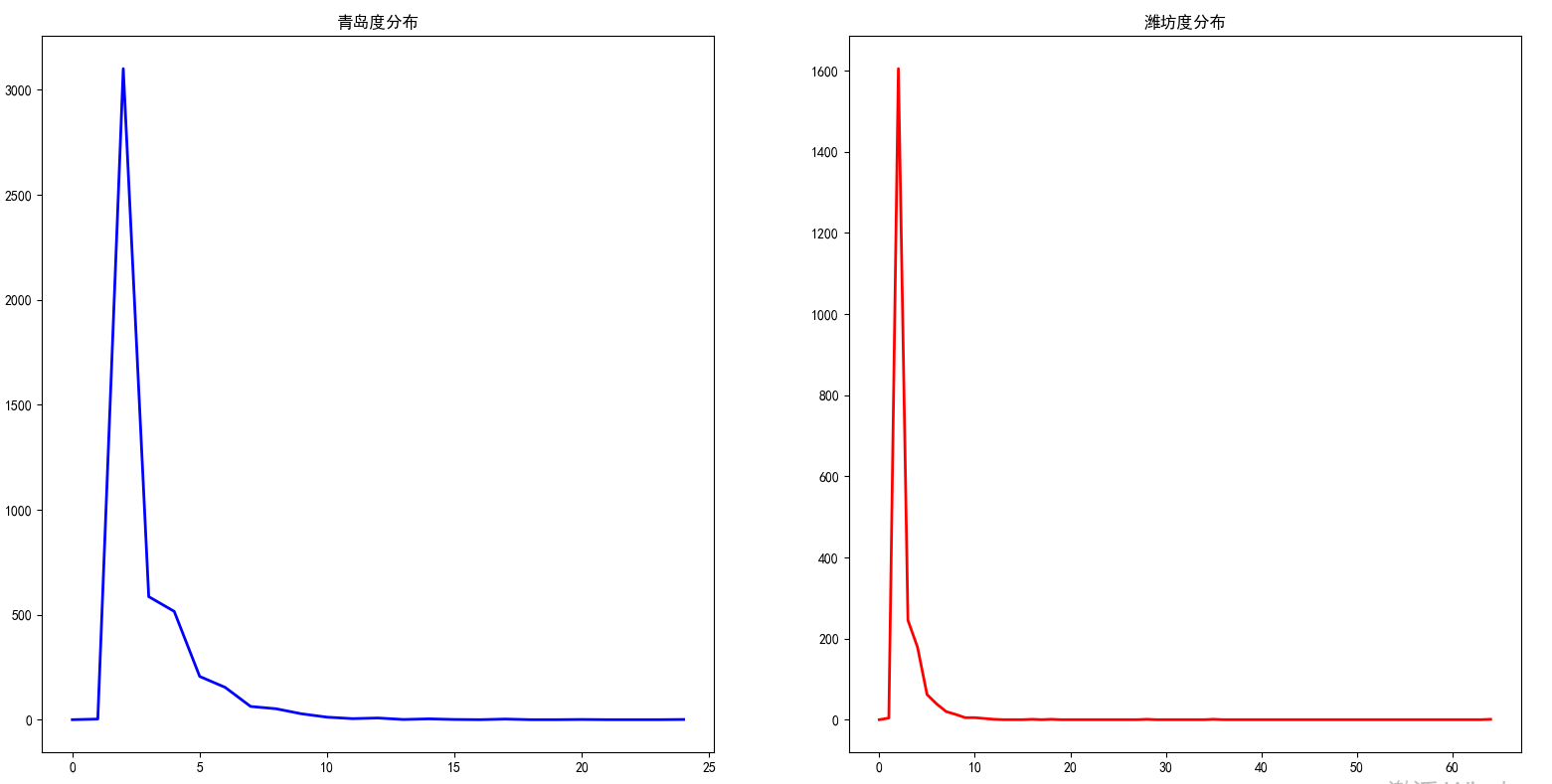


图3-1 度分布

## 总结

本文以青岛市和潍坊市为例，建立了基于站点的公交网络模型。研究表明，在基于站点的公交网络中，度分布极为不均，大部分节点的度比较小，只有极少数的节点度比较大，满足幂律分布，网络呈现出无标度网络特性。虽然网络中的节点数很多，但是平均路径长度较短，呈现出明显的小世界特性。公交网络作为一个复杂系统，还有很大的优化空间。复杂网路的优化，将对人们的出行需求、城市规划、道路分级等产生重大影响。