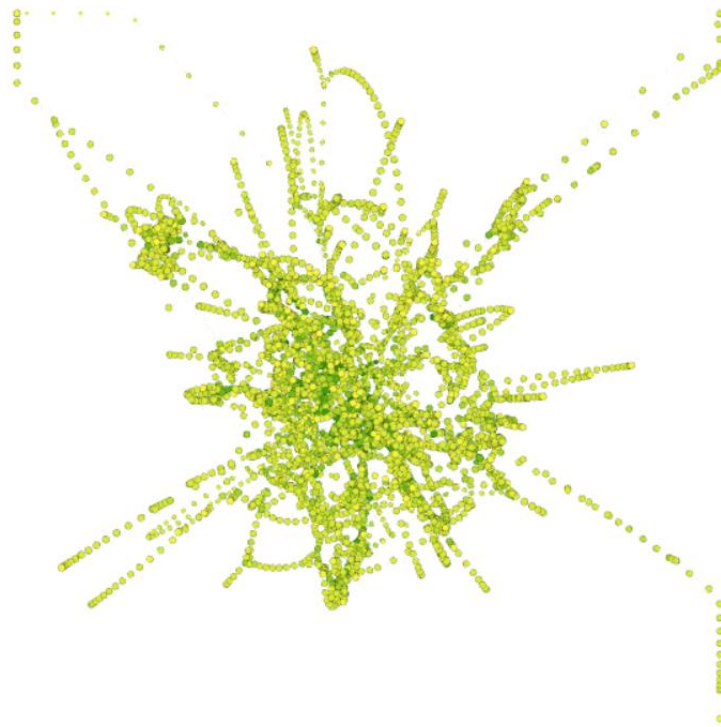


基于复杂网络的天津市公交动态拓扑分析

摘要：基于复杂网络的理论，构建了天津市公交线路的 L-Space 网络模型，对公交站点的地理布局进行了可视化。计算了各个公交站点的平均度、整个公交网络平均最短路径长度、平均聚类系数、中介中心性、紧密中心性、节点离心率、特征向量中心性。研究了天津市公交网络的无标度特性和小世界特性。探讨了天津市公交网络在结构上的特点和属性。

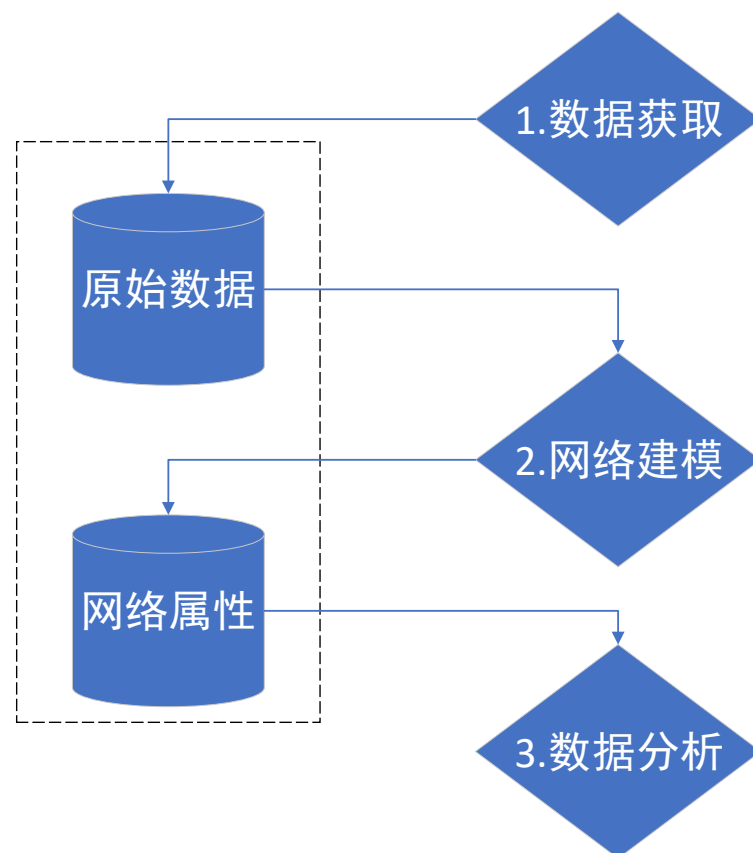


通过 gephi 的 Force Atlas 算法构造的天津市公交网络图

1. 引言

理解天津市公交网络拓扑结构对于城市建设, 交通规划具有很大的参考价值。基于复杂网路研究城市公共交通系统已有很多研究, 如 Latora 和 Marchiori(2002)通过研究波士顿地铁网络发现它是小世界网络, Sen(2002) 等研究了印度铁路网络的小世界特性, 世界航空网络是无标度网络等。研究天津公交网络拓扑特性同样具有高价值和高重要性。

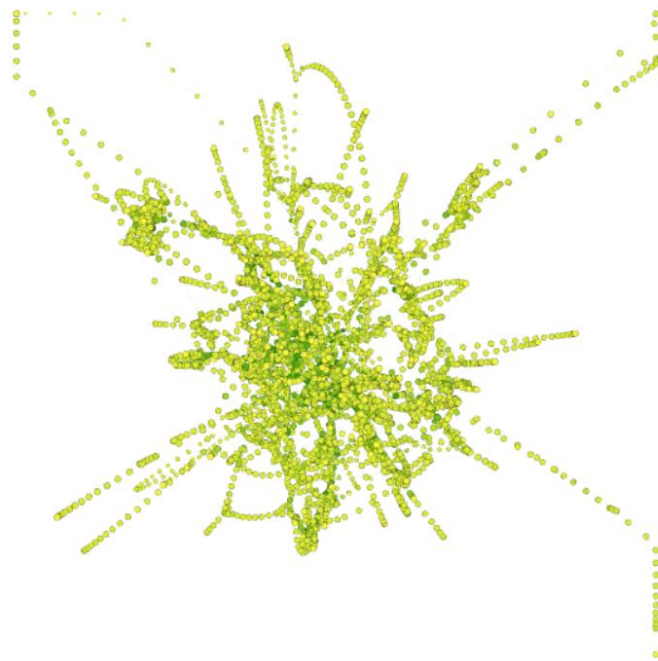
分析流程:



首先我们组通过 Gephi 工具，python 的 Folium 包，对天津市公交站数据进行了一个可视化，首先对公交站点的分布获得了一个大致的印象。下面分别是 Gephi 的 Force Atlas 方法建模和热力图建模。

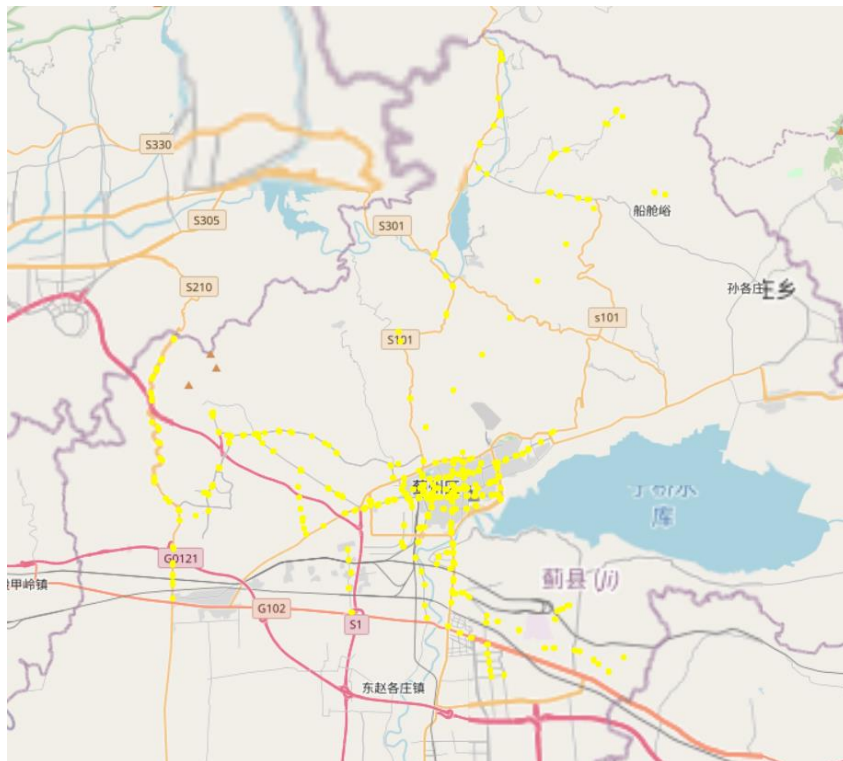
Gephi 的 Force Atlas 建模图：

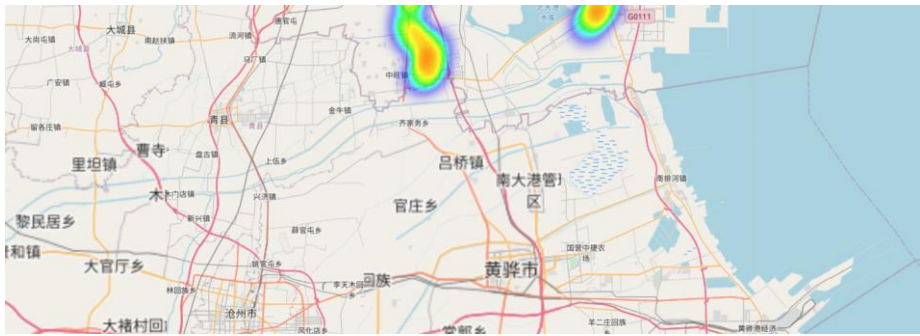
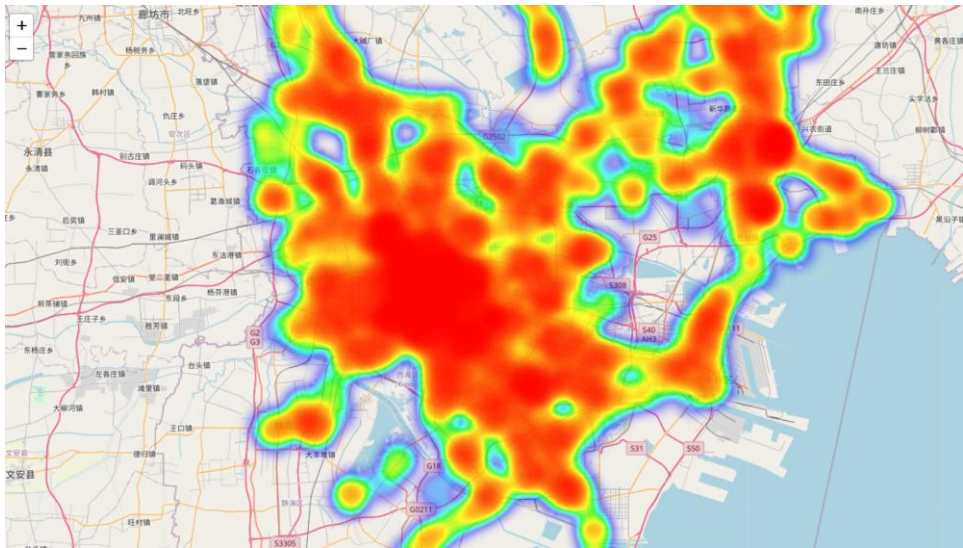
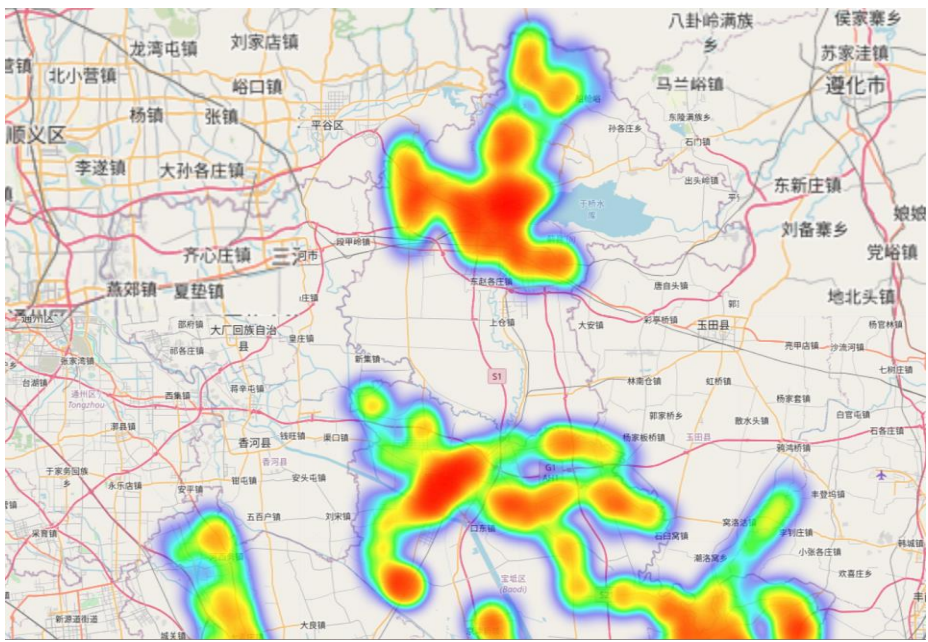
图中每个节点代表现实世界中的一个公交站，颜色越深代表该节点的度数越高
Force Atlas2 为力引导布局[1]，力引导布局方法能够产生相当优美的网络布局，并充分展现 网络的整体结构及其自同构特征。力引导布局即模仿物理世界的引力和斥力，自动布局直到力平衡；Force Atlas 布局使图更紧凑，可读性强，并且显示大于 hub 的中心化权限（吸引力分布选项），自动稳定提高布局的衔接。



通过 Python Folium 库构建的在线地图，分别为站点的地理位置图和站点的分布图

站点地理位置图：





2.1 天津市公交网络的建模方法

前人的大量研究表明现实世界中的大部分网络并非规则网络或者随机网络，而是介于二者之间一种兼具二者属性的网络。故本次采用的建模方法为建立 L-space 网络模型[2]，可以非常直观地反映天津公交网络的各项数据的属性，便于研究复杂网络的性质。L-space 网络空间和现实世界网络一致，现实世界的网络中的一个站点代表 L-space 空间中的一个节点，相邻站点间的路线代表 L-space 空间的一条边。

2.2 天津市公交网络基本属性

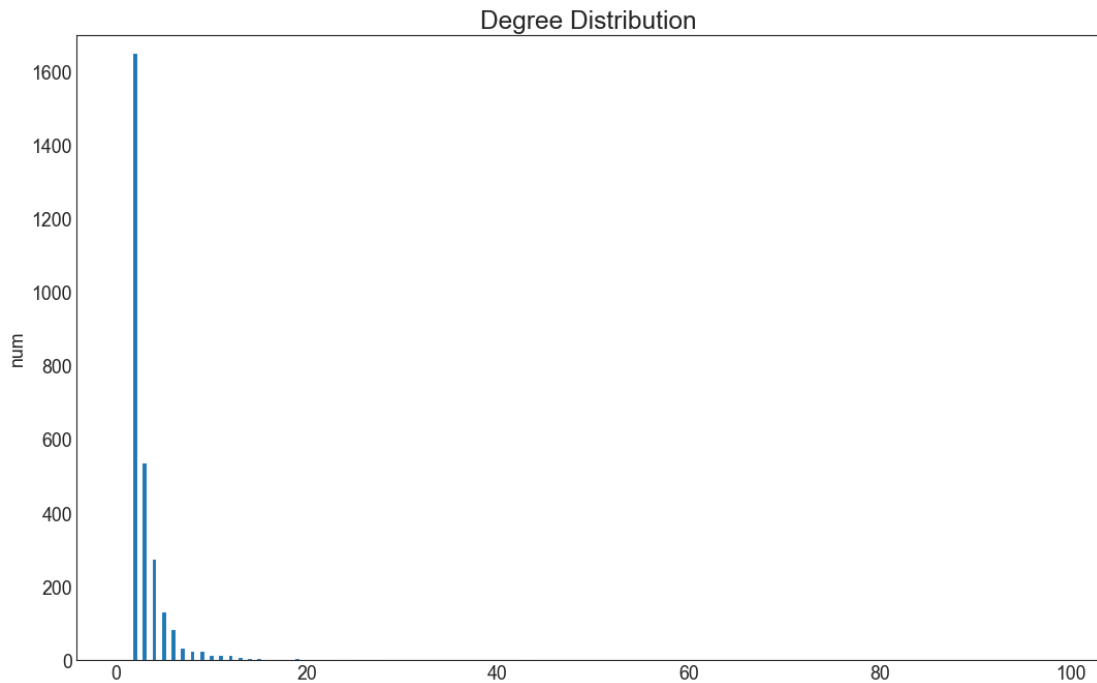
平均度	平均最短路径	平均聚类系数	网络直径
2.822	16.345	0.091	73

网络中一共有 4872 个站点，13751 条边，构建的是有向图。

1) **平均度<k>**. 一个网络由一些节点和连接它们的连结构成。每个节点连出的所有连结的数量就是这个节点的**度**。度分布是对一个网络中节点度数的总体描述。节点 i 的度 k_i 为节点连接边的总数，所有节点 i 的度 k_i 的平均值为这个网络的平均度。在公交网络中，节点的度代表了这个公交站相连的线路数。 [3]

平均度< k >的计算公式，其中 n 为节点总数， k_i 为第 i 个节点的度数

$$\langle k \rangle = \frac{1}{n} \sum_{i=1}^n k_i$$



从图上可以发现，绝大多数站点度数为 2，其次是 3、4、5 等等。

根据天津市公交网络的计算数据，节点的平均度 $\langle k \rangle$ 是 2.822，即平均每个公交站有 2.8 条线路经过，即一条往返线路外加 0.8 条非本趟次公交线路。拥有最大的度的站台为“鼓楼东”，度数为 40，最小度数为单线终点站，度数为 1；

2.2 平均路径长度

网络的平均路径长度 $\langle l \rangle$ 定义为任意两个节点之间距离的平均值，即[4]

$$\langle l \rangle = \frac{2}{n(n-1)} \sum_{i \leq j \leq n} d_{ij}$$

公交网络的最短路径长度是从一个站点到另一个站点所要经过的站点数目的最小值。公交网络的平均路径长度反应了，从一个站点到另一个站点平均需要经过多少站。分析发现天津市公交网络的平均路径长度是 16.345，即平均要经过 15 站才能到达目的地。

2.3 聚类系数

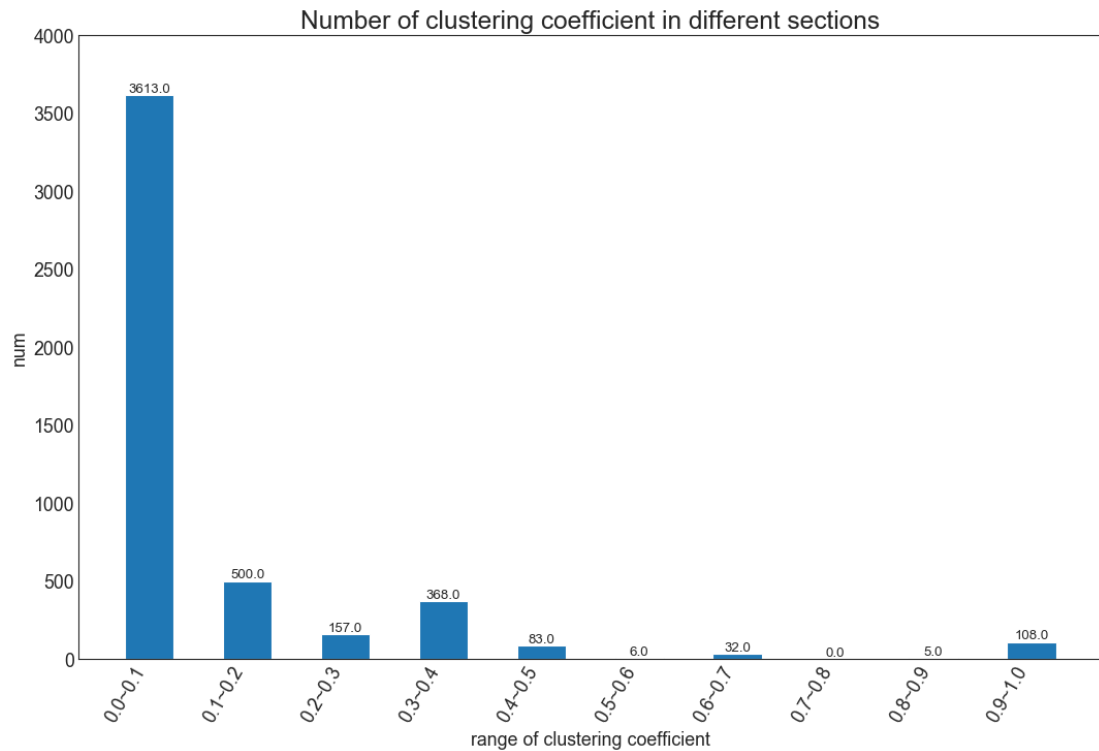
假设网络中节点 i 的度数为 k_i ，那么它有 k_i 个邻居，那么如果这 k_i 个邻居也两两互为邻居，那么，在这些邻居节点之间就存在 $k_i(k_i-1)/2$ 条边。然而实际中没有这么多邻居，所以设实际边数为 e_i ，那么某个节点的 i 的聚类系数 C_i 为[5]

$$C_i = \frac{2e_i}{k_i(k_i-1)}$$

那么平均聚类系数 CC 为

$$CC = \frac{1}{n} \sum_{i=1}^n C_i$$

通过计算可得，网络的平均聚类系数为 0.091



从图中可以发现，绝大部分节点的聚类系数处在 0~0.1 之间。说明公交线路相互交汇的情况很少，大部分公交站点都是存在一个往返线路的情况，很少出现三条线路相交的情况。

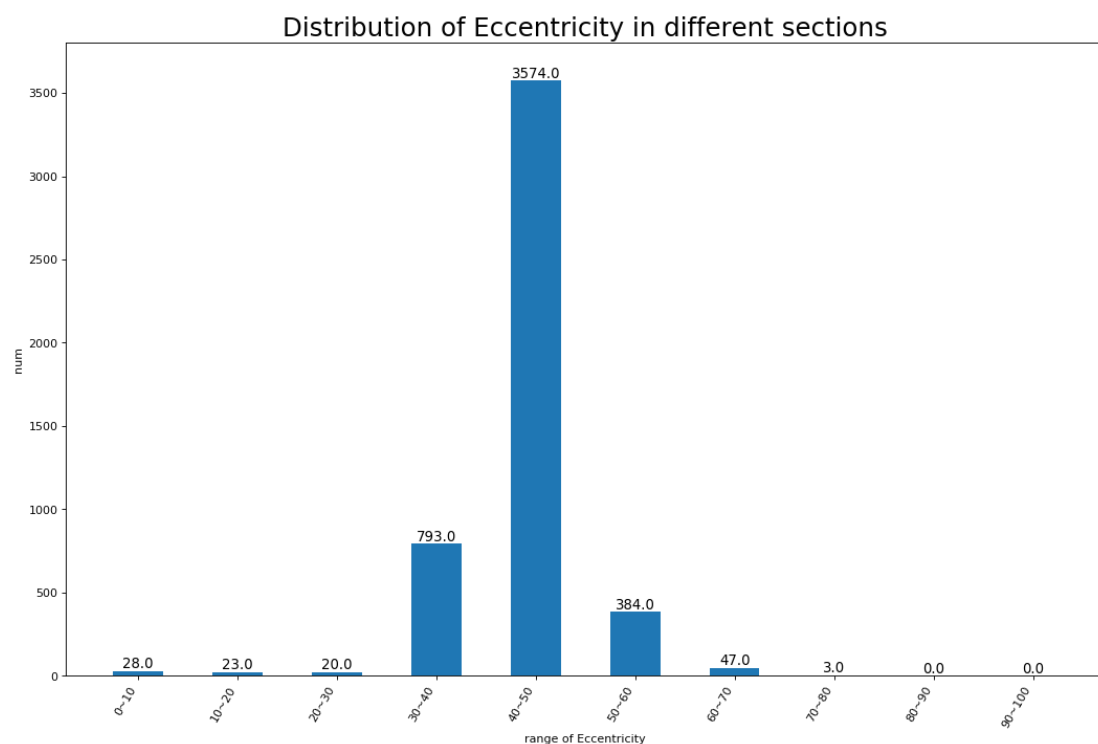
2.4 网络直径

网络中任意两个节点之间的距离的最大值称为网络的直径,记为 $D[6]$,即

$$D = \max d_{ij}$$

经过计算得出，网络直径为 73。实际意义表示公交网络中存在的某一个站点到另一个站点需要经过的最大站点数目。

从网络直径中可以引申出另一个已有概念，即节点 i 的**离心率**（eccentricity），用来表示复杂网络中的顶点 i 到网络中其它节点之间的最大距离。网络直径实际上是最大离心率，在公交网络中即是从一个站点到另一个站点要经过站点数目最多的情况。

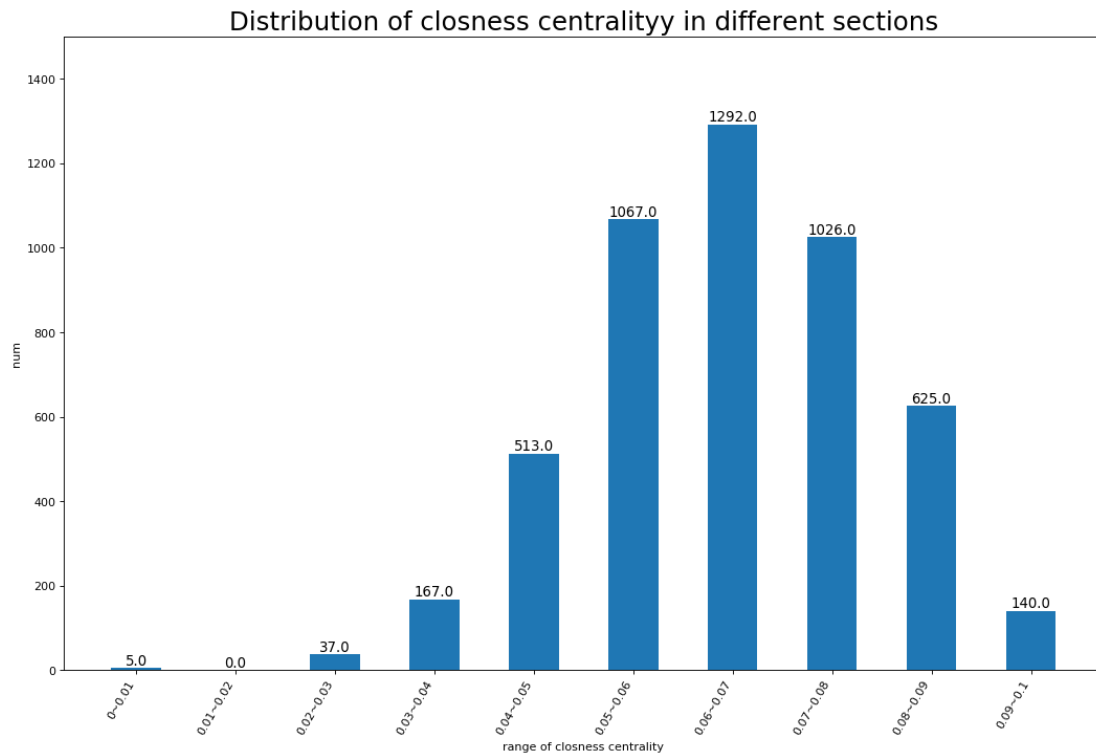


离心率可以用来反映网络的大小规模，从图中可以看出公交线路的普遍长度在 40~50 个站点之间。

2.5 接近中心度，计算的是一个点到其他所有点的距离的总和，这个总和越小就说明这个点到其他所有点的路径越短，也就说明这个点距离其他所有点越近。接近中心度体现的是一个点与其他点的近邻程度 [7]，定义为

$$C(x) = \frac{1}{\sum_y d(y, x)}$$

一个具有高接近中心度的点，说明这个点距离任何其他点都最近，在空间上也体现在中心位置上。



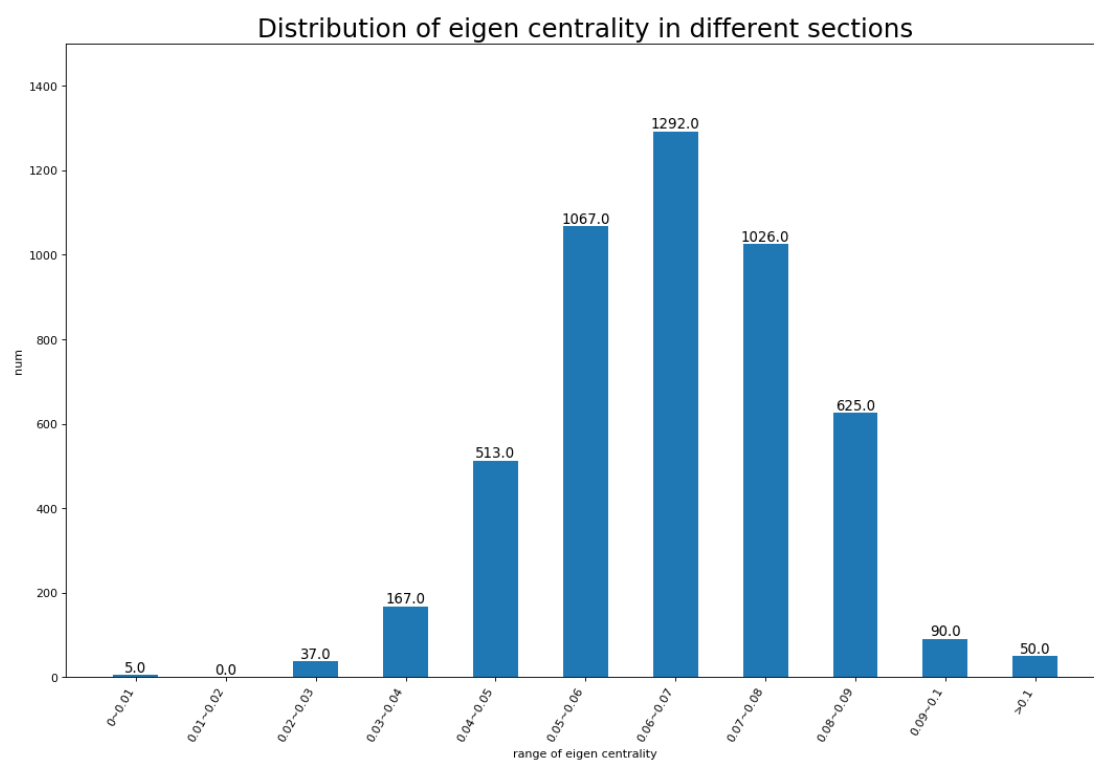
2.6 特征向量中心性的基本想法是：一个节点的重要性既取决于其邻居节点的数量，也取决于其邻居节点的重要性，记 x_i 为节点 v_i 的重要性度量值[8]

$$EC(i) = x_i = c \sum_{j=1}^n a_{ij} x_j,$$

其中， c 为一个比例常数，记 $x = [x_1, x_2, x_3, \dots, x_n]^T$ ，经过多次迭代到达稳态时，可以写成如下的矩阵形式：

$$x = cAx.$$

特征向量中心性值得分布



3.1 网络的种类

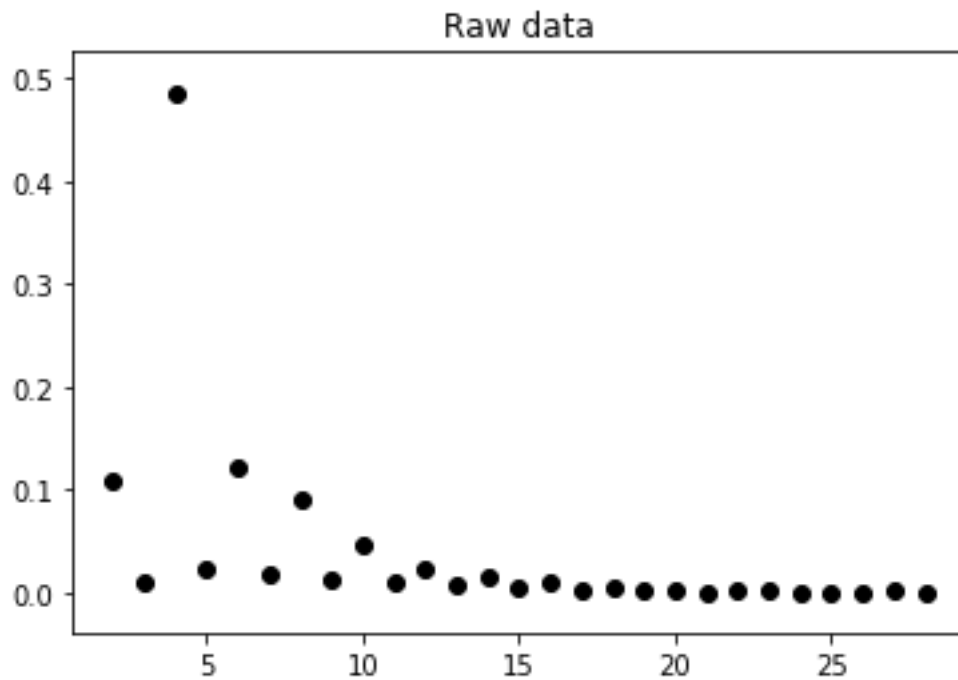
无标度网络中，极少数的点度数极高，而大多数点度数较低，度数在分布上符合幂律分布。[9]

$$p(k) = Ak^{-\gamma}$$

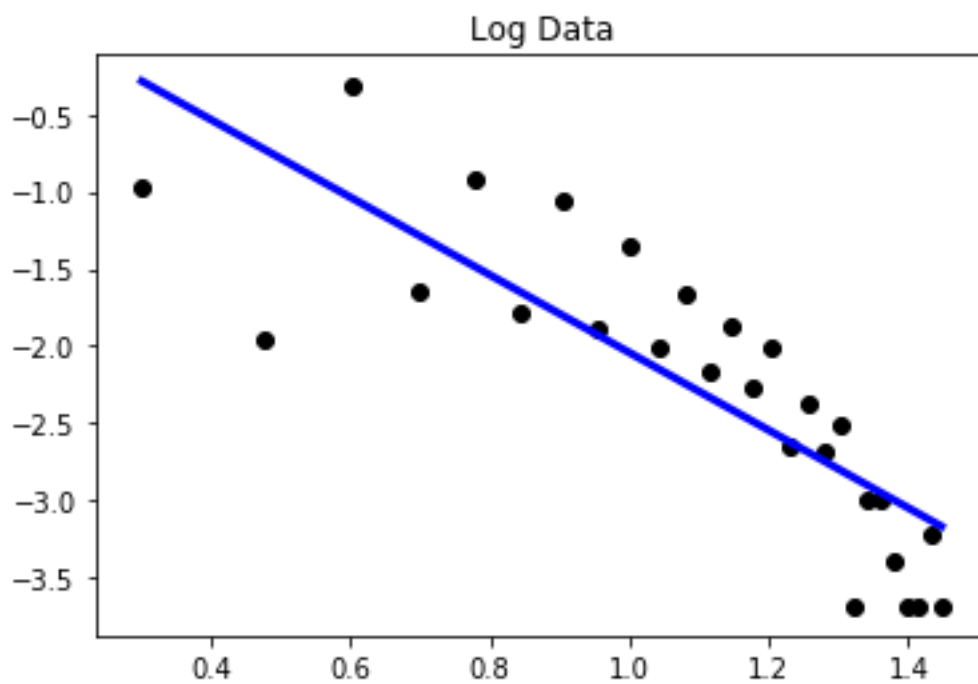
$$\ln p(k) = -\gamma \ln k + c$$

在本次分析中对度分布进行了一个拟合：

首先是度分布散点图



幂律拟合

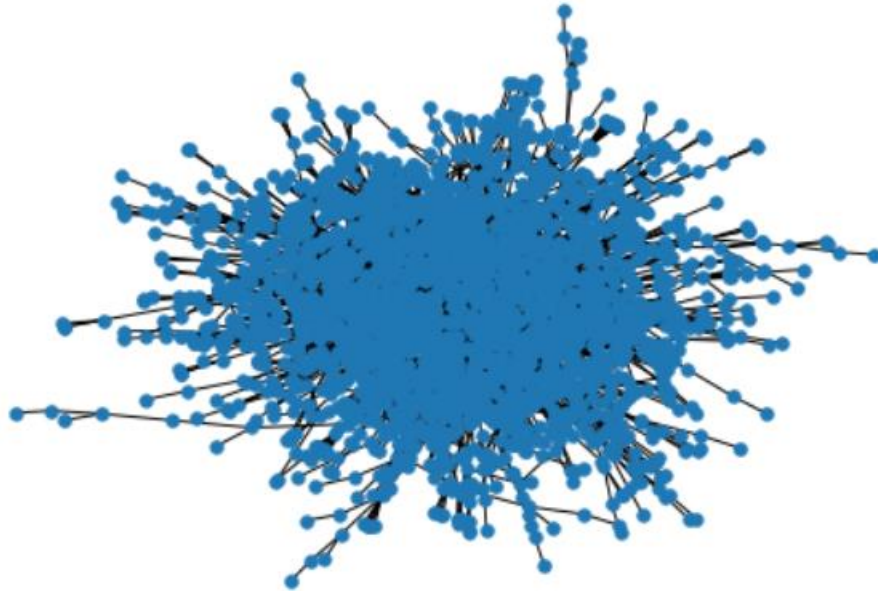


Coefficients:
[-2.51513549]
Intercept:
[0.47312275]

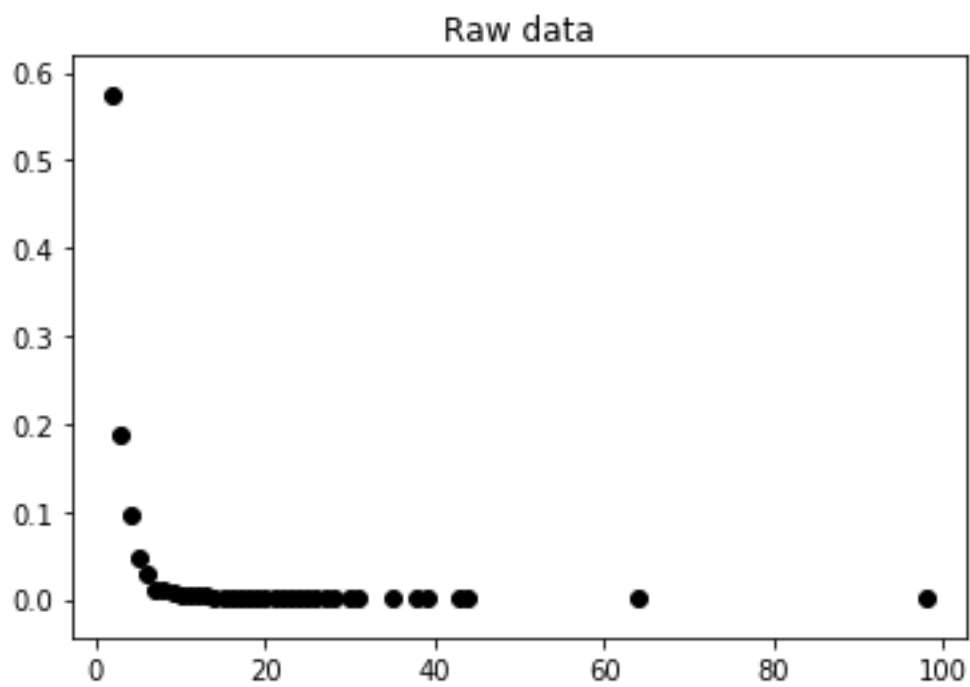
较为符合一般无标度网络的特点，Coefficients 在-2 到-3 之间。下面通过 python 的 networkx 生成一个和天津公交网络的具有相同节点数的

无标度网络来进行验证。

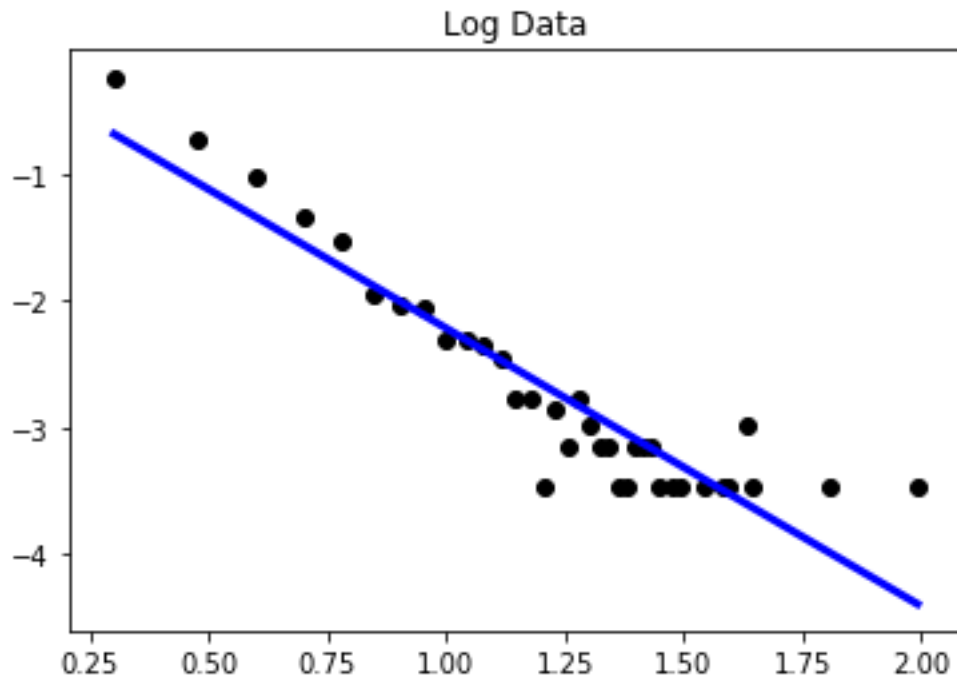
Networkx 生成的无标度网络



生成网络的度数分布



生成网络的拟合结果



Coefficients:

[-2.19284402]

Intercept:

[-0.02269761]

发现生成的无标度网络具有和天津市公交网络相似的特性，充分说明天津市公交网络确实具有无标度特性。

小世界网络特性. 小世界网络可以是高度聚类的（拥有很高的聚类系数）同时具有一个较小的平均路径长度[10]。而从 **L-space** 空间模型的聚类系数（**0.091**）和平均路径长度（**16.345**）来看无法判断其并不具有小世界属性。可能通过 **P-space** 构建其公交换乘网络会具有其他属性，但单单从 **L-sapce** 空间模型来看我们认为它不具有小世界属性。

4. 结论

通过构建 L-space 网络空间研究了天津公交网络的拓扑结构及网络的各项属性。通过 networkx 构建类比网络、Folium 库构建站点经纬度图，站点分布热力图等，得到了网络的度分布特征、平均路径长度、聚类系数等各项特征，研究了网络的无标度特性、小世界特性；发现在 L-space 网络空间下，天津市公交网络平均度仅仅为 2.8，聚类系数为 0.91，说明网络结构松散，网络节点间的互相连接非常稀疏。而网络的平均路径长度达到了 16，说明平均下来从一个站点乘坐公交车到达另一个站需要经过 15 个站，相对而言还是十分漫长的。而网络中还存在少数度数非常高的点和大量度数非常低的点。这种情况下网络中的某个节点在受到随机攻击时还不至于瘫痪，而如果度数高的节点被针对攻击的话可能会导致出现很多断线，导致网络瘫痪的情况。由于缺乏客流量方面的数据，如果在求大于供的情况下，建议增加线路，将之前在拓扑结构上没有交集的线路彼此连接，从而降低平均路径长度，分担度数高节点的客运压力，提高网络的鲁棒性。

[1] ForceAtlas2, A Graph Layout Algorithm for Handy Network Visualization

By Jacomy, Mathieu, Heymann, Sebastien 《Plos One》,2011

[2] S. Derrible and C. Kennedy, “Applications of graph theory and network science to transit network design,” *Transport Reviews*, vol. 31, no. 4, pp. 495–519, 2011.

[3] 汪小帆、李翔、陈关荣，网络科学导论，高等教育出版社，2012 年 4 月：P87

[4] [3]P91

[5] [3]P97

[6] [3]P92

[7] [3]P161

[8] [3]P165

[9] M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, Oxford, UK, 2010

[10] D. J. Watts, “Collective dynamics of ‘small-world’ networks,” *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.