**大的背景：**

我们想对思知数据进行领域划分

传统社团划分的模块度，可以理解为每一个社区内部节点的度除以总边数 记作QM

我们引入了一个关于的关系类别的模块度QR，定义为每两个社团关系向量的距离减去对应的平均距离，如下所示：

1: 社团i和j的关系向量的距离

2: 社团i和j的关系向量的平均距离

(距离为标准化的欧氏距离)

所以我们需要计算任意两个社团之间的距离。

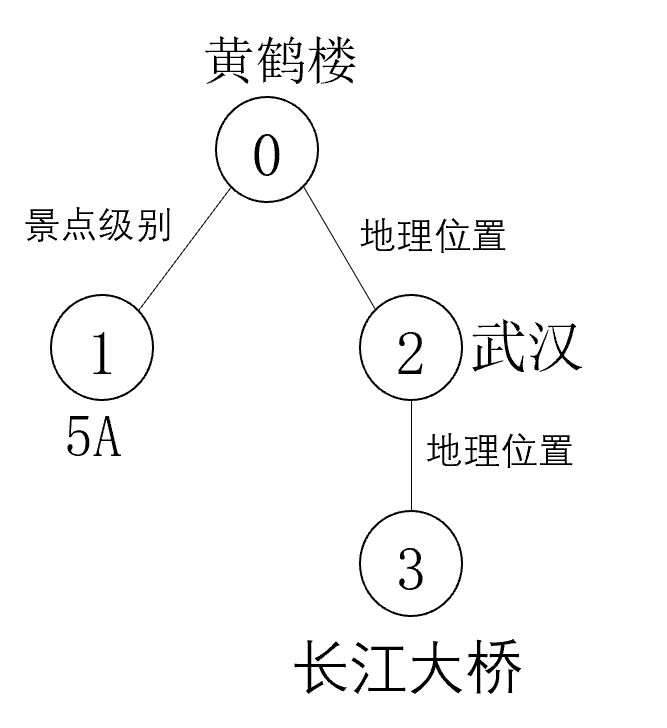
**小的背景：**

即计算QR。

任意一个社团C都对应一个关系向量，用来记录这个社团所包含的边的，以计算

还对应一个平均关系向量，用来记录平均值，以计算

拿下图作为例子，来简要说明一下。



初始领域划分，把每个点作为一个社团

第一个表 记录社团状态

|  |  |
| --- | --- |
| 社团 | 包含点 |
| 0 | 黄鹤楼 |
| 1 | 5A |
| 2 | 武汉 |
| 3 | 长江大桥 |

根据第一个表计算QM

第二个表 记录

|  |  |  |
| --- | --- | --- |
|  | 景点级别 | 地理位置 |
|  | 1 | 1 |
|  | 1 | 0 |
|  | 0 | 2 |
|  | 0 | 1 |

根据第二个表计算，以和为例

即(1, 1)和(1, 0)在这个向量空间的距离

*atan*

第三个表 记录

|  |  |  |
| --- | --- | --- |
|  | 景点级别 | 地理位置 |
|  | 1 | 1 |
|  | 0.5 | 0.5 |
|  | 1 | 1 |
|  | 0.5 | 0.5 |

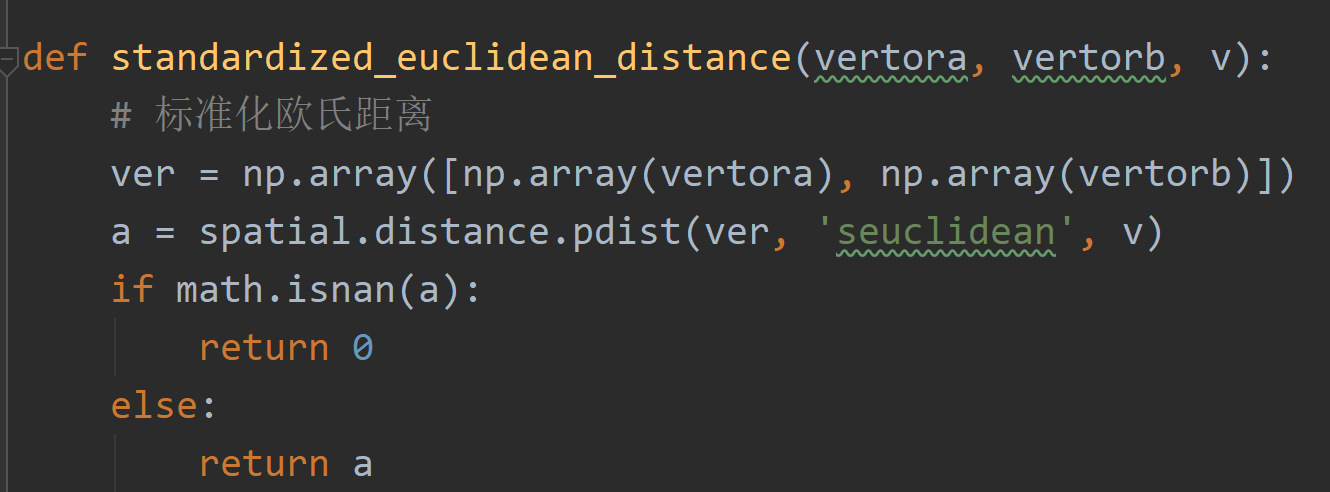
根据第三个表计算，过程同第二个表的计算过程

**存在问题：**

因为这个向量空间可以理解为稀疏矩阵，也就意味着存在某一列值相同(一般是0)，但是这个向量空间中这个情况出现的概率比较低，所以计算出很可能为0(当方差为0时，结果为nan，我们当作0处理)，而不为0所以到时候计算出的QR大概率为负数，实验结果显示 -380、-260等(我也记不太清了)

而QM的范围在(0, 1)， QR的影响过大，需要归一化

下图是我们计算标准化欧氏距离的方法



因为我取的是上三角，所以个数为 (忽略对对角线)，但是得到