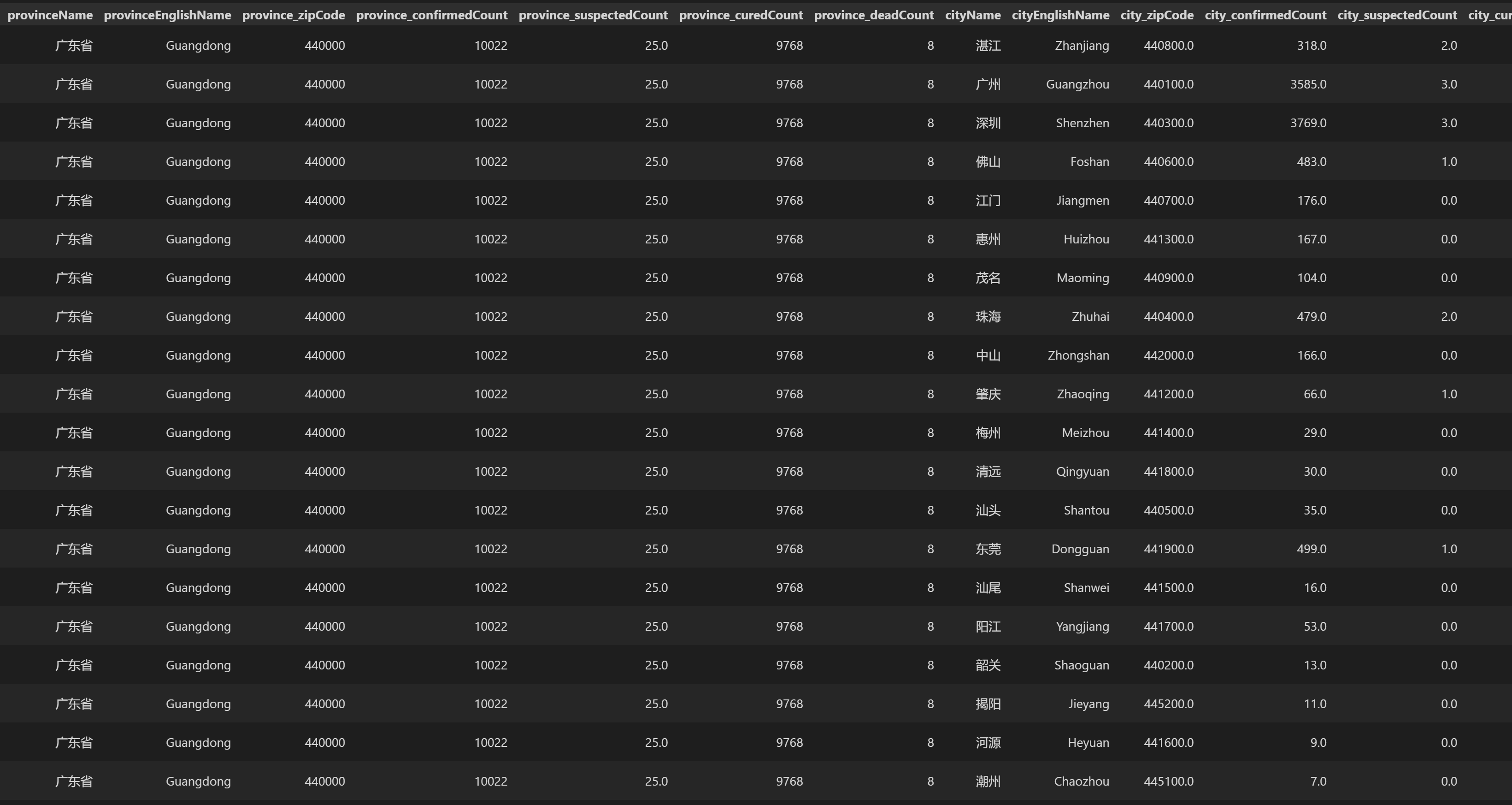
**出发点**：

通n国庆七天全国疫情数据，试图寻找各城市在节假日疫情状况的关联性，对全国城市进行相似度建模。对于关联性高的城市，我们认为其防疫政策与疫情状态存在一定联系。具体而言，根据疫情数据，我们认为每日新增数量以及增长率近似的城市的实际防疫政策以及疫情情况是类似的。同时，我们还考虑城市间的距离关系，但因为现代交通发达，我们给予距离关系较小权重。我们可以猜测，根据人口密度和人口流动量来看，一线城市应该具有较高的疫情情况关联性，这一点也确实吻合现有疫情情况。因此，我们最后会以此验证我们的网络关系的可信度。

**数据来源：**

新型冠状病毒疫情时间序列数据仓库<https://github.com/BlankerL/DXY-COVID-19-Data/releases>

数据示例：



可以看到我们的数据包括全国城市的名字、邮编、累计感染数、累计痊愈数、累计死亡率，城市邮编等。我们可以通过前后两天的疫情数据计算出每日新增数以及增长率。

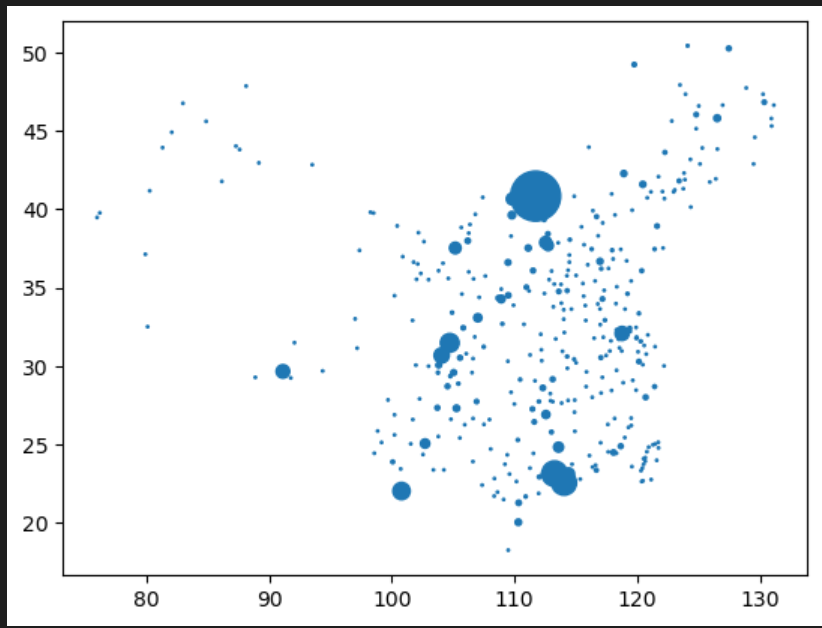
为了更好地可视化全国疫情关系，我们还下载了全国各城市经纬度，可视化全国城市位置。

**相似性计算指标：**

每日新增、增长率、距离。对每两个城市我们使用高斯核函数计算其相似度，公式如下。值得注意的是，因为现代交通发达，节假日内人们前往远距离旅游的可能性远大于前往附近边缘城市，所以我们给距离很小的权重。

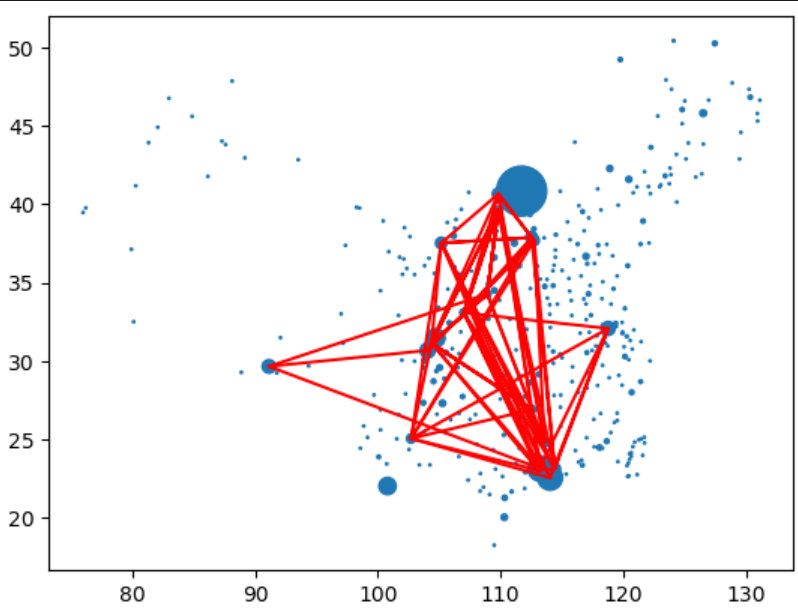
其中A new和A grow rate分别表示城市A的每日新增和增长率，城市B同理。Distance表示城市AB的距离, 是可调节的超参数，用来控制核函数的方差。根据上述相似度计算公式我们计算每两个城市之间的相似度，建立. 随后我们对M进行对称归一化,

对于国庆七天疫情情况，我们根据每一天数据进行迭代计算。为了方便观察，进行如下可视化。

**

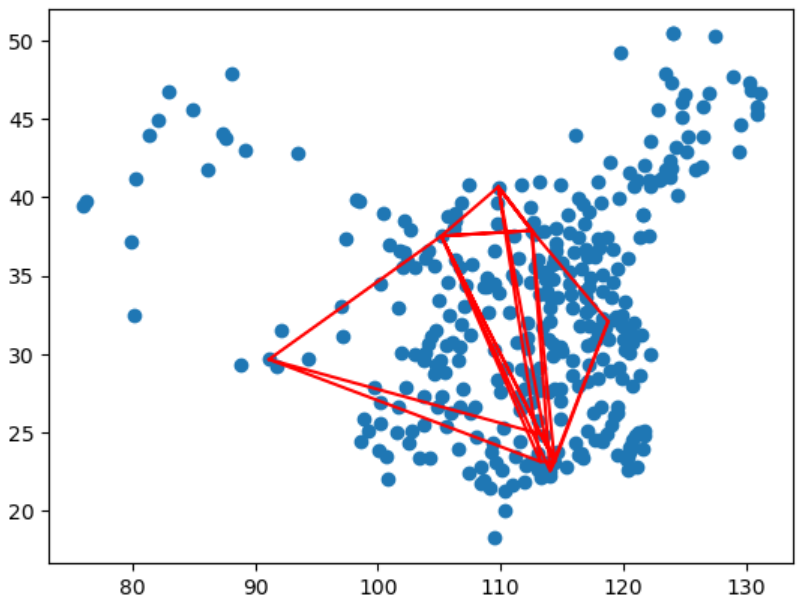
上图为10.1全国新增病例，每个点都代表一个中国城市，点越大代表新增数越多

随后，我们进行第一次相似度计算，且显示高相似度结点之间的边



如图，我们可以看到如深圳、广州等人口流量大的城市出现比较高的相似度。但是因为只计算了一次，网络中的边关系比较杂乱，无法看出其中关联。

随后我们采用的方式更新相似度矩阵，其中 new M是根据第二天数据计算出来的相似度矩阵。在迭代国庆七天所有数据后，我们得到如下所示图（为了方便观察，仅显示边关系）。

**

可以看到，随着相似度矩阵的迭代，各城市间杂乱的关系变得规则，我们有理由相信图中所示具有边关系的城市具有较高的疫情相似度。其中一线城市的广州、深圳、北京、上海都具有较高相似度，符合国庆假日间一线城市人口流动量大的特点。

**未来可行的工作**

随着更多的数据参与计算，以及更合理的相似度计算，我们可以制作出更合理的全国城市疫情相似度图。如此，便可采用条件随机场等经典算法去预测各城市的疫情风险。