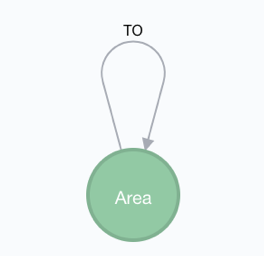
基于Neo4j的订单地域分布关联分析

1. 分析思路，第一阶段，使用小数据集进行分析。完成流程操作。然后对其中存在的一些特征进行进一步分析。
2. 分析过程
   1. 使用R语言做数据预处理:
      1. 提取成都市1月份以来的具有有效经纬度的共2339429个订单，按照区块分配到了403682个关系。（开始点 结束点，订单数）
      2. 为了减小计算难度。过滤掉订单数小4的关系。最终剩下78301个关系。这些关系占总订单的79%。这些关系由5462个区域构成。
   2. 导入到Neo4j中建立图形关系（节点为区域点、边为订单数）关系模型为
      1. 
   3. 使用python调用Neo4j并使用数据分析库进行分析。
      1. betweenness（中介性核心性）分析：
         1. 概念：两个非邻接的成员间的相互作用依赖于网络中的其他成员，特别是位于两成员之间路径上的那些成员。它们对这两个非邻接成员的相互作用具有某种控制和制约作用。Freeman (1979)认为中间成员对路径两端的成员具有“更大的人际关系影响”。该指标的思想是：如果一个成员位于其他成员的多条最短路上，那么该成员就是核心成员，就具有较大的中间度核心性。Anthonisse(1971)[8]和Freeman(1977)[9]首先对该思想进行了量化。假设：①各条线路的权系数相等；②通讯过程中总是走最短路径；③当存在不止一条最短路时，每一条的使用概率相等；中间度核心性的本质是：网络中包含成员i 的所有最短路的条数占所有最短路条数的百分比。它表示成员i 的控制能力，或起着“gatekeeping”作用的概率。也可以理解为：成员j要到达网络中的所有其他成员，对成员i的依赖性的大小，或是成员i 对成员j 与网络中的其它成员进行相互作用的控制能力。
         2. 用途：用于发现城市外围热点区域核心路径点。如下是计算出来的排名靠前的坐标点。



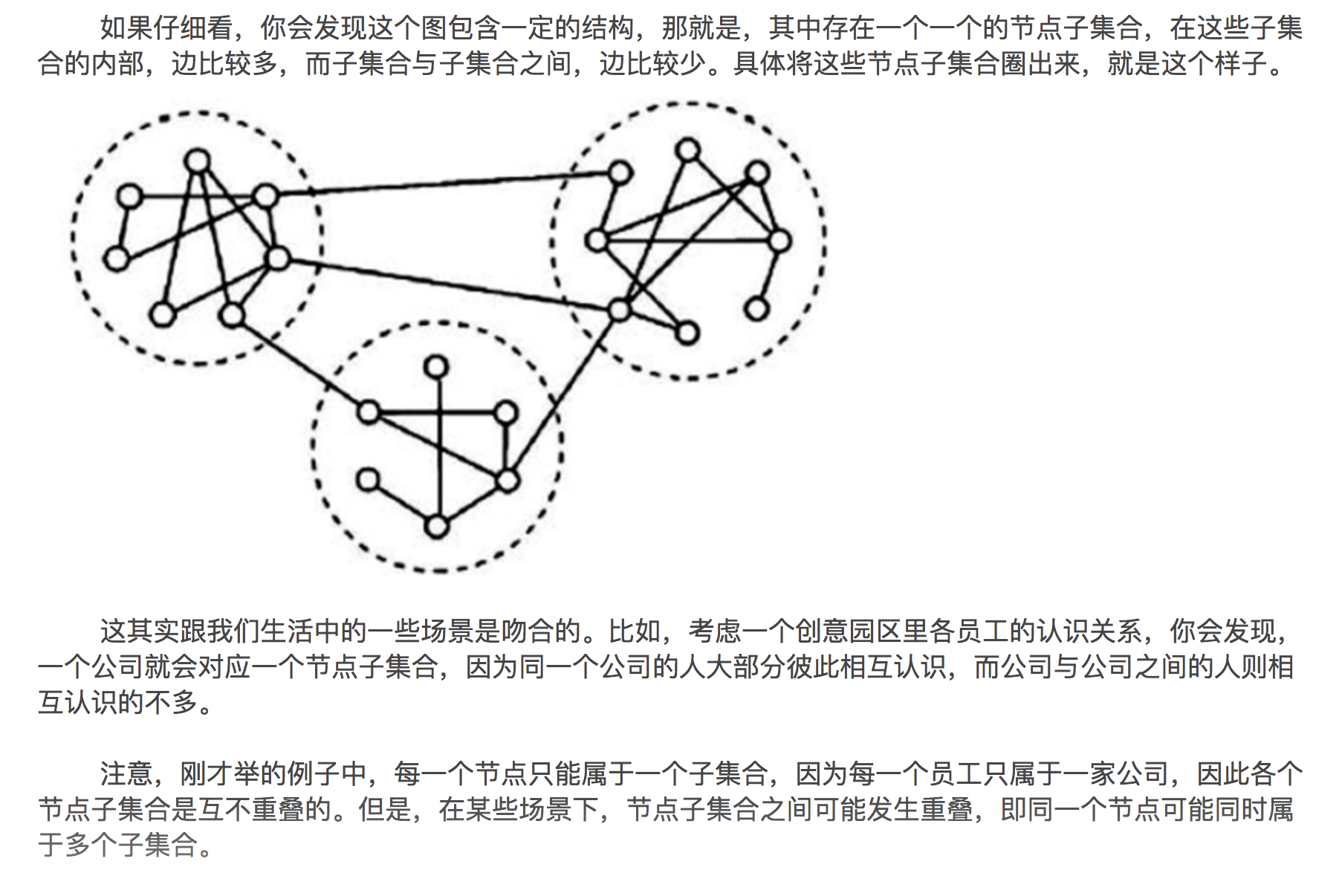
分析结论。通过观察发现。该点多为城市外围的热点区域的主要交通口。特点为周边有学校、医院、等公共设施。集中于地铁口。符合正常的出行场景。这些点。存在着很多的订单。在城市边缘与中心进行交互。属于外围流量的必经入口。通过该方法。可以有效的为各个城市外围热门区域运营提供支持。我们可以在以上点作为停车位。作为推荐停车点的的补充。

* + 1. pagerank分析：

1. 概念：PageRank让链接来"投票"一个页面的“得票数”由所有链向它的页面的重要性来决定，到一个页面的超链接相当于对该页投一票。一个页面的PageRank是由所有链向它的页面（“链入页面”）的重要性经过递归算法得到的。一个有较多链入的页面会有较高的等级，相反如果一个页面没有任何链入页面，那么它没有等级。

2. 用途：由于我们业务场景中也存在着。订单从一个区域流向另一个区域的特点。在以往的分析中。单纯的按区域进行聚合统计。但是忽略了一个比较重要的事实。如果我们的订单流向了一个一个热点区域，那么这个订单，质量就更高。如果流向了一个非热点区域，这个订单，质量就更低。使用这个算法，可以更合理的计算热点区域的质量。（需要对原pagerank进行些调整。）现有的分析结果发现。Pagerank的结果指向市中心的热点区域。

c. 社群发现：

 可以用于将热点进行分区域划分。同区域的热点关联性比较大。可以用来辅助城市分区域调度，管理。下图为用颜色区分的一部分社区图。

