.

**硕 士 学 位 论 文**

DOCTORAL DISSERTATION

论文题目 关键传播路径下的空气污染物传播

动态分析方法

作者姓名吴瑶

学科专业计算机科学与技术

指导教师黄国言教授

**2018年6月**

中图分类号：xxx 学校代码：10216

UDC：xxx 密级：公开

**工学硕士学位论文**

关键传播路径下的空气污染物传播动态分析方法

|  |  |  |
| --- | --- | --- |
| 硕士研究生 | ： | 吴瑶 |
| 导师 | ： | 黄国言教授 |
| 申请学位 | ： | 工学硕士 |
| 学科专业 | ： | 计算机科学与技术 |
| 所属学院 | ： | 信息科学与工程学院 |
| 答 辩 日 期 | ： | □□□□年□月 |
| 授予学位单位 | ： | 燕山大学 |

A Dissertation in Mechanical Manufacturing and Automation

（Times New Roman 字体4号）

Bio-syncretic Rehabilitation Mechanism theory and Application

（Times New Roman 2号字加粗，题目太长时可用小2号字）

by □□□

（Times New Roman字体4号）

Supervisor: Professor□□□

（Times New Roman字体4号）

**Yanshan University**

（Times New Roman 4号字 加粗）

June,2009

（Times New Roman字体4号）

燕山大学硕士学位论文原创性声明

本人郑重声明：此处所提交的硕士学位论文《 》，是本人在导师指导下，在燕山大学攻读硕士学位期间独立进行研究工作所取得的成果。论文中除已注明部分外不包含他人已发表或撰写过的研究成果。对本文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。本声明的法律结果将完全由本人承担。

作者签字： 日期： 年 月 日

燕山大学硕士学位论文使用授权书

《 》系本人在燕山大学攻读硕士学位期间在导师指导下完成的硕士学位论文。本论文的研究成果归燕山大学所有，本论文的研究内容不得以其它单位的名义发表。本人完全了解燕山大学关于保存、使用学位论文的规定，同意学校保留并向有关部门送交论文的复印件和电子版本，允许论文被查阅和借阅。本人授权燕山大学，可以采用影印、缩印或其它复制手段保存论文，可以公布论文的全部或部分内容。

保密□，在 年解密后适用本授权书。

本学位论文属于

不保密□。

(请在以上相应方框内打“√”)

作者签名： 日期： 年 月 日

导师签名： 日期： 年 月 日

摘 要

随着我国经济的的快速发展，中国大部分地区的空气污染状况日趋严重。严重的空气污染会引发一系列的疾病，因此城市空气质量对保护人民群众身体健康，提高人民生活质量，起着关键的作用。但是，由于我国地域辽阔，无法全方位的覆盖空气质量监测站，因此导致了不能及时并全面的反映当前的空气质量状况，以及不能有效的对空气污染做出有效的防治。于是，本文提出一种基于基于污染物传播机理的污染物传播模型，在此基础上挖掘出实时污染物传播路径和一个周期内污染物的关键传播路径，从而以此为依据来决定监测站的选址以及有效的防治大气污染物的传播。

首先，由于无法获取污染源的详细信息，因此无法采用传统的基于污染源的扩散模型来挖掘传播路径。于是，本文充分的分析了影响污染物扩散的因素和机理，提出了一种基于污染物传播代价的传播模型，进而基于此传播模型挖掘出实时的污染物传播路径。

其次，通过每一时刻的污染物传播路径构建成污染物传播网络图，经过分析发现该网络图具有复杂网络的特性，因此本文采用复杂网络的度量方法来分析污染物传播网络的特性。

再次，本文基于污染物传播网络提出一种图矩阵的方法，来存储每条边在每个子图中出现的情况，并通过生成污染物传播网络子图、构建有向图矩阵、构建简化图矩阵来挖掘污染物关键传播路径。

最后，本文以京津冀地区的大气污染物传播关系为实验，验证了基于传播代价的传播模型的正确性，经实验发现京津冀地区的大气污染物传播网络具有复杂网络的特性，

关键词：大气污染物传播；复杂网络；关键路径；图矩阵；

Abstract

The rehabilitation mechanism is the foundation for a rehabilitation robot to realize its motion, and the quality of the rehabilitation mechanism decides the rehabilitation effect of patients with the rehabilitation robot. The research of rehabilitation mechanism is……

**Keywords:** keyword 1;keyword 2; keyword 3; ……; ……

目 录

摘 要 I

Abstract II

目 录 III

第1章 绪 论 1

1.1 课题背景及研究的目的和意义 1

1.2 国内外研究现状 1

1.3 论文研究内容 1

1.4 论文的组织结构 1

第2章 相关理论和基础知识 2

2.1 基本概念 2

2.1.1 空气质量指数 2

2.1.2 污染物传播机理 4

2.2 复杂网络特性 6

2.3 本章小节 6

第3章 大气污染物传播建模 7

3.1 数据预处理 7

3.1.1 球面上两站点之间的距离 7

3.1.2 监测站方向与风向的夹角 7

3.1.3 监测站之间污染物的相关性 8

3.1.4 数据标准化 8

3.2 污染物传播模型构建 9

3.3 实时关键传播路径挖掘 9

3.3.1 传播约束条件分析 10

3.3.2 关键传播路径挖掘算法 11

3.3.3 算法实例 14

3.4 实验结果与分析 15

3.4.1 实验环境配置 15

3.4.2 实验数据 16

3.4.3 实验结果分析 17

3.4.4 实验结果验证 20

3.5 总结 22

第4章 基于复杂网络的关键传播路径挖掘 24

4.1 引言 24

4.2 大气污染物传播网络构建 24

4.2.1 基本概念与定义 25

4.3 基于图矩阵的污染物关键传播路径挖掘 26

4.3.1 生成大气污染物传播网络子图 27

4.3.2 构建有向图矩阵 30

4.3.3 构建简化图矩阵 31

4.3.4 挖掘污染物关键传播路径 32

4.3.5 实例说明 34

4.4 实验结果分析 38

4.4.1 实验环境配置 38

4.4.2 实验数据 38

4.4.3 实验结果分析 38

4.4本章小结 44

第5章 生物融合式康复机构型综合研究 45

5.1 节标题 45

5.4 本章小结 45

结 论 46

参考文献 47

攻读博士学位期间承担的科研任务与主要成果 48

致 谢 49

第1章 绪 论

## 课题背景及研究的目的和意义

## 1.2 国内外研究现状

## 1.3 论文研究内容

## 1.4 论文的组织结构

第2章 相关理论和基础知识

2.1 基本概念

### 2.1.1 空气质量指数

空气质量指数(Air Quality Index，简称AQI)是环保部门向公众展示当前空气污染程度的一组数据，是一个用来定量描述空气质量水平的数值。AQI的取值范围最低为0最高至500，AQI指数上升，那么身体受伤害的人口比例就会上升。我国将空气质量指数分为染以优、良、轻度污染、中度污染、重度污及严重污染六个等级，具体分类和标准如表2-1。而面向群众的空气质量通报，对每个污染等级都设定了固定颜色与其对应，这样只需要根据报告的AQI值，甚至只看颜色，即可直观判断空气质量水平。





现阶段国内监测的空气污染物的种类主要包括六种，包括SO2、NO2、CO、O3和悬浮颗粒物。悬浮颗粒物中，直径小于等于10的称为PM10，直径小于等于2.5的称为PM2.5。国内城市对空气污染影响较大的是PM2.5。在2016年我国新增加了对PM2.5的监测要求并规定了浓度限值。利用监测网络对所测得的各种污染物浓度数据，计算成AQI值，可以对空气污染程度有更明确的度量。

根据表2-2计算空气质量分指数公式如下：

|  |  |  |
| --- | --- | --- |
|  |  | (2-1) |

式中 ——污染物项目的空气质量分指数；

——污染物项目的质量浓度值；

——表2-2中与相近的污染物浓度限值的高位值；

——表2-2中与相近的污染物浓度限值的低位值；

——表2-2中与对应的空气质量分指数；

——表2-2中与对应的空气质量分指数。

空气质量指数计算方法如式(2-2)：

|  |  |  |
| --- | --- | --- |
|  |  | (2-2) |

式中 ——空气质量分指数；

——污染物项目。

当时，最大的污染物为首要污染物。若最大的污染物为两项或两项以上时，并列为首要污染物。

表2-2 空气质量分指数及对应的污染物项目浓度限值



### 2.1.2 污染物传播机理

空气污染物传播是一个十分复杂的过程，在这个传播过程中会受到多方面因素的影响。其中主要的三个方面是地形因素、气象因素以及污染物因素。地形因素包括：两个地区之间的水平距离 、两个地区之间是否有山脉、两个地区的海拔。气象因素包括：风速、风向、温度、湿度、降雨、降雪、雾、沙尘等。污染物因素包括：两地之间的浓度差。

水平距离：由于污染物传播具有远距离传输的特性，所以在一定范围的的地区，常常会与受污染的区域有着同步增长和减少的趋势，因此这些地区的相关性就会很高，这样就会导致当一个地区受到严重的污染时，与其相邻的其他地区的污染物浓度也会同步增加。

山脉：山脉对污染物的传播既有抑制作用，也有促进作用。当山脉位于两地之间时，这时山脉就会对污染物的传播起到一个阻挡和削弱的作用。当两个地区位于山脉的同侧时，污染物会沿着山脉的走势进行传播，此时山脉对污染物的传播就会起到促进的作用。

海拔高度：当两个地区的海拔高度相差不大，并且之间不存在山脉时，此时污染物的传播主要受气象因素的影响。如果两地之间的海拔高度差十分大时，低海拔地区对高海拔地区的影响是十分小的，而低海拔地区由于地势较低，污染物就会产生汇聚现象，导致该地区的污染物浓度很高。

风向、风速：风是影响一个地区大气污染物运动和分布的重要气象因子。风向决定了污染物的传播方向，风向决定了污染物的传播速度、传播的范围的大小以及局地地区污染物浓度的大小。当风速较大时，会使局地地区的空气变得稀薄，污染物的沉降效果降低，因此局地地区的空气质量会很好，但是这样污染物影响的范围会扩大。当风速很低时，不仅使污染物很难得到扩散，而且对污染物的聚集起到了促进作用，很容易会造成局部严重污染。

温度：逆温是指地面温度比较冷、上层温度比较热的“下冷上热”，形成比较稳定的形势，同时高空的暖温区与地面辐射降温相配合，形成逆温层，不利于污染物的对流和垂直扩散。

湿度：如果一个城市长时间无降雨，造成空气质量不佳，之后虽降了一场雨，但其效果只是“湿润空气”。同时由于空气湿度加大，让累积悬浮的污染物如同穿上一层水衣，给扩散加重负担。如果没有风，“重装”污染物基本上停滞于空气中，就会形成烟雨蒙蒙的效果，使空气质量不佳。在湿度较高的情况下，空气中的污染物会“吸湿增长”，导致其质量增加，污染物浓度也会相应提高，这样一来，空气质量反而会变差。只有在雨量较大，一般是中到大雨的情况下，才容易通过物理沉降的方式降低空气中污染物的浓度，起到净化空气的作用。

降水：降水对大气中的污染物能起净化和再分布的作用。降水对大气的净化作用包括云滴的吸收和降水（雨、雪）对污染物颗粒的冲刷、与对污染物气体的溶解作用。降雨的净化作用与降雨量和降雨持续时间相关，雨水量大，持续时间长，净化作用就越明显，反之，则没有净化作用，甚至会加重污染。

降雪：冬季降雪对污染物稀释有一定的作用，但作用不很明显, 几乎不能通过信度检验。而且随雪量增加反而污染浓度越重的出现概率越大。原因是冬季降水基本以固态冰晶形式出现，其清洗系数比降雨来的弱一些，表现在降雪强度上要比雨强小一个数量级，密度小其沉降速度比雨滴要小的多，雪粒子与空气粒子间隙大，云和降水卷夹凝聚和吸附作用远不如降雨，因此降雪对污染物粒子的捕获作用比降雨小的多，雪洗系数值就低。即使有较大的降雪，它的稀释作用抵消不了气象条件对空气污染的影响。

沙尘：沙尘天气会使空气污染加剧，强沙尘暴过程会使污染物浓度在非常短的时间内提高3倍左右，造成严重的颗粒物污染。空气中高浓度的悬浮颗粒物，在静风、逆温等特殊的气象条件相互作用下，易导致雾霾天气的发生，具体表现为区域性或大范围内的空气质量恶化。

污染物浓度差：一个地区的污染物浓度十分高时，它对周围地区污染物浓度有着很大的影响，根据气体扩散原理，污染物浓度高的地区的污染物会向污染物浓度低的地方进行扩散，到这整个区域受到污染，另外受到风等气象因素的影响，会导致整个区域污染程度的加剧，因此，两个地点之间的污染物浓度差也是影响空气污染物扩散的一个重要因素。

2.2 复杂网络特性

2.3 本章小节

第3章 大气污染物传播建模

3.1 数据预处理

数据去噪：由于数据中存在一些值为空值，这些空值将会对后边的计算产生一定的误差，考虑到样本数据十分的多，因此本文采用了直接去除空值的手法。即如果某条数据中的一项为空值，则舍弃该条数据。

### 3.1.1 球面上两站点之间的距离

地球是一个近乎标准的椭球体，平均半径为6356.755km。假设监测站A的经纬度为(lonA,latA)，监测站B的经纬度为(lonB,latB)。然后根据三角推导，可以得到计算两监测站距离的公式如下：

(1)

(2)

其中，R为平均半径6356.755km，pi为3.1415。

### 3.1.2 监测站方向与风向的夹角

根据向量乘积的原理，假设存在向量和，因此存在如下等式：

(3)

其中和代表向量的横纵坐标，和代表向量的横纵坐标，和表示向量和的模，为两个向量之间的夹角。

假设风向的单位向量为，从监测站A到监测站B方向的向量为，如图1所示。

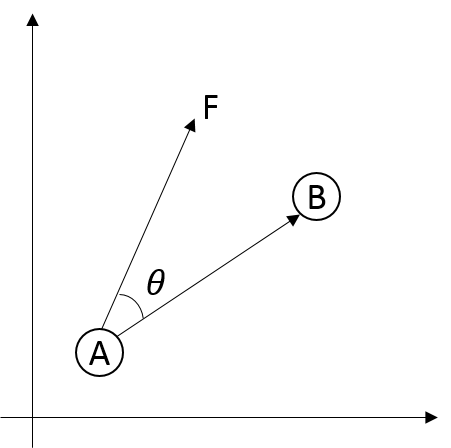


图1风向夹角示意图

因此可将上述向量乘积的原理应用到计算两个监测站方向与风向之间夹角的计算中，公式如下：

(4)

于是可以求出夹角为：

(5)

其中为风向的弧度，为向量的模。

### 3.1.3 监测站之间污染物的相关性

假设监测站A在t时刻的污染物浓度为，监测站B在t时刻的污染物浓度为，监测站A上污染物浓度均值为，监测站B上污染物浓度均值为。因此，A、B监测站之间污染物的相关性为可根据公式(6)求得：

(6)

### 3.1.4 数据标准化

数据的标准化是将数据按比例缩放，使之落入一个小的特定区间。在某些比较和评价的指标处理中经常会用到，去除数据的单位限制，将其转化为无量纲的纯数值，便于不同单位或量级的指标能够进行比较和加权。其中最典型的就是数据的归一化处理，即将数据统一映射到区间上。

本文采用标准化方法，对原始数据进行线性转换。公式如下：

(7)

其中max为样本数据的最大值，min为样本数据的最小值。

3.2 污染物传播模型构建

空气污染物传播起点的气象因素，如：温度(T)、湿度(H)、风向(WD)、风速(WS)等都会影响空气污染物的向外扩散，并且空气污染物的传播难易程度不仅受本地气象因素的影响，也会受到传播目的地的影响。其中，两地之间的距离(D)，两地之间的地形(hill)，两地之间的海拔差)，监测站污染物的相关性()，两地之间的污染物浓度差()，以及两地之间的气象因素的差异都会影响到空气污染物的传播。根据以上描述以及气体扩散理论，t时刻空气污染物从站点A传播到站点B的传播代价为：

(8)

如式中表示起始点A在t时刻的湿度，表示起始点B在t时刻的温度，表示站点A与站点B之间的水平距离，表示站点A与站点B之间时刻的差值，表示站点A与站点B之间的相关性，表示站点A与站点B之间t时刻的风力系数差值，表示传播代价的波动值，表示校正系数，，取值范围均为。

通过污染物在站点A与站点B之间的传播代价可以衡量出污染物传播的难易程度，污染物传播代价越大，代表污染物越是不易传播，反之，越容易传播。该公式不仅考虑了污染物起始地的气象因素，也考虑到了污染物目的地的气象因素，以及两地之间的地形因素。而不是像其它文献中的模型只考虑风向风速等气象因素。

3.3 实时关键传播路径挖掘

由于地形、地貌等一些自然条件的限制，和气象因素的影响，导致在一些地区出现一些污染物的传播通道，例如：以京津冀为例，其西侧、北侧靠山、东临渤海，导致在偏西南风气流的影响下，污染物会沿着太行山系中的洋河河谷和燕山山脉向京、津、冀地区传播，导致大面积、区域性的污染。因此研究污染物的关键传播路径是十分重要的，不仅可以有效的对污染进行区域联防，而且可以及时阻断污染物的传播。并且由于京津冀地区十分广阔，不能实现监测站的全方位的覆盖，因此寻找污染物的关键路径，并合理的在关键路径上设置空气质量监测站，不仅减少了监测站的建设成本，也对整个地区起到了有限的防控作用。

### 3.3.1 传播约束条件分析

水平距离(D)：根据数据集station.csv中的经纬度以及公式(1-2)计算出站点之间的距离。如果监测站和监测站的实际水平距离小于或者大于50km，则断开监测站和监测站之间的路径。因为当距离小于10km时，基本上两个站点所有状态以及数据都是相同以及同步的，这样就会导致在一个局地地区存在大量的路径，使网络中存在大量的环，这样对研究污染物的区域性传播十分不利。从数据可知，大多数县级区域有且至少有一个监测站点，并且距离都在50km以内，又因为空气污染物的传播具有远距离传输特性，并且传输距离远大于50km，因此如果两监测站距离如果大于50km，就断开连接。

山脉(hill)：当两个监测站之间存在山脉时，则认为污染物是无法在这两个地区之间进行传播的。（可以根据山的高度来设定污染物递减的等级）

海拔差()：当海拔差超过一定数值时，则认为污染物无法从低海拔的地区传播到高海拔的地区。

相关性()：当两个在适当范围内的站点上的污染物浓度出现同增或者同减的现象时，往往是因为一个站点上的污染物浓度对另一个站点产生了影响。因此，根据公式(6)对网络中的任意两个站点进行污染物浓度的相关性计算。当时表示站点之间有紧密依赖关系，也就是说一个站点的污染物浓度很容易就会影响到另一个站点，也就说明了这两个站点之间存在一条污染物传播通道。反之不存在传播通道。

风力系数差()：设站点A到站点B的方向为，站点A和站点B在t时刻时的风速风向分别为、，则分别与站点A和站点B风向的夹角为，根据三角形勾股定理可知，风速、在方向上的分量分别为，如图2所示。

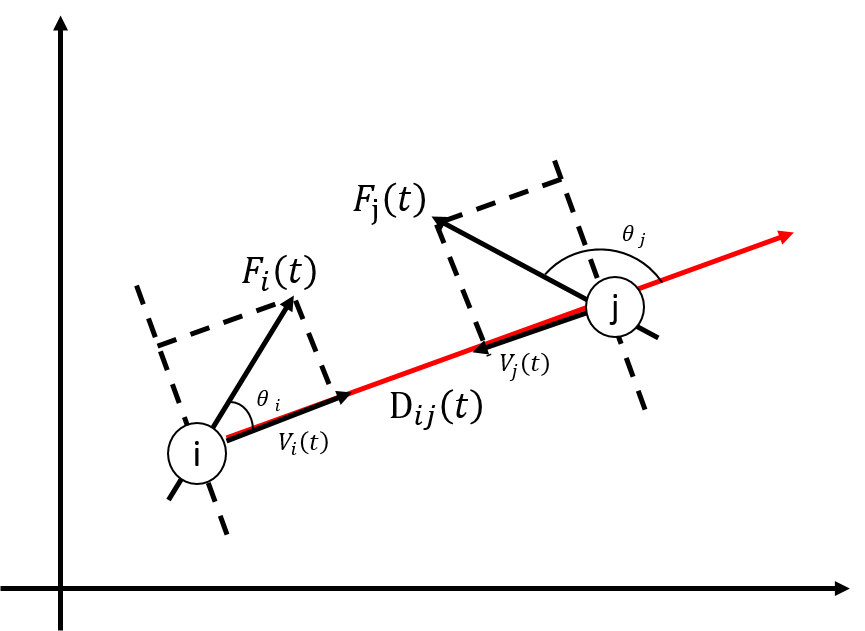


图2 监测站上风力系数示意图

因此可以求出站点A与站点B之间t时刻的风力系数差值为：

(9)

(10)

(11)

(12)

其中可根据公式(4-5)计算得出。

如上式，当站点A处的风向与A、B站点方向的夹角超过时，则认为此时站点A处的空气污染物不会对站点B处的空气质量造成影响，或者当小于等于0时，表示站点A在方向的风力系数小于等于站点B在方向上的风力系数，即站点A处的空气污染物是无法对站点B处造成影响的。

### 3.3.2 关键传播路径挖掘算法

当污染物从一个地区向相邻地区进行扩散传播时，往往不只是一条路径进行传播，而是存在多条不同的传播路径。由于地形因素以及气象素的影响，会导致其中一些路径对污染物的传播有极小的抑制作用，这样就使得污染物在这条路径上很容易被传播，而有些路径则对污染物的传播有极大的抑制作用，导致污染物很难在这条路径传播，其中这种难易程度本文称之为传播代价(Cost)。因此，从传播原理来看，传播代价越小的路径越是容易形成污染物的传播通道，所以，将传播代价最小的路径定义为污染物关键传播路径。

根据污染物传播的约束条件以及关键传播路径的定义，可总结出站点A与站点B存在边的条件，如下式：

其中，表示监测点A和监测点B之间的距离，表示监测点A和监测点B在t时刻的风力系数的差值，表示监测点A和B之间污染物的相关系数，代表监测点A和B在t时刻的污染物传播代价。当为0时，表示站点A、B之间在t时刻不存在一条污染物关键传播路径，为1时表示存在一条污染物关键传播路径。

t时刻污染物从站点A传播到站点B的传播路径)：

其中，表示时刻t时路径)内包含边的集合。

t时刻污染物从站点A传播到站点B的传播路径对应传播代价：

其中，表示时刻t时路径)内所包含边的传播代价的集合。

根据污染物传播模型以及污染物传播代价即可计算出t时刻的从任意一点出发的污染物关键传播路径。如算法1所示：

算法1：Search Critical Path(SCP)

输入： ，

，

，

输出：

执行过程：

BEGIN

1. Initialize:
2. zero List nodeList,
3. ;
5. ;







14. ;
16. END

如算法1所示，首先任意选择一个污染物传播的起始点，方法根据公式（1-2）计算起始点与每个监测站的距离。并选择出距离在10km到50km范围内的站点。将这些站点的集合记作，然后接下来的计算将基于集合中的站点进行。由于数据集中风向不是用角度进行表示的，因此需要通过方法进行角度转换。通过方法计算时刻t时中每个站点与的风向夹角，并筛选出夹角在0度到90度的站点，并将这些站点的集合记作AngleList，然后在AngleList的基础上通过方法计算每个站点与的在t时刻时的传播代价，并将符合传播代价小于0.01的站点的集合记作CostList，最后在CostList中找到传播代价最小的站点，因此路径即为在t时刻以为起点的一条最容易传播的路径，这条路径因此称之为污染物关键传播路径。

### 3.3.3 算法实例

下面将通过一个实例展示出算法1的过程。该实例选取井陉县气象局2014年4月26日时的气象数据以及污染物浓度数据来寻找以井陉县气象局为起点的污染物关键传播路径。井陉县气象局的经纬度为。根据方法计算出满足与井陉县气象局距离在10km到50km范围内的监测站，如表1所示。

表1 距离井陉县气象局的监测站

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 监测站 | 监测站 | 监测站 | 监测站 | 监测站 |
| 市区世纪公园 | 市区西南高教 | 市区化工学校 | 市区职工医院 | 市区人民会堂 |
| 市区西北水源 | 市区高新区 | 市区封龙山 | 鹿泉一中 | 正定联通公司 |
| 栾城通讯公司 | 元氏住建局 | 赞皇县政府 | 平山冶河 | 灵寿供水 |

通过方法计算出表1中监测站与井陉县气象局风向的夹角，如表2所示。

表2 各监测站与井陉县气象局风向夹角

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 角度 | 角度 | 角度 | 角度 | 角度 |
| 63 | 135 | 81 | 150 | 165 |
| 45 | 143 | 35 | 132 | 145 |
| 87 | 56 | 156 | 180 | 90 |

由于当夹角为钝角时，视为污染物不能传播到下一个污染物监测站，因此去除夹角为钝角的点，则剩下的监测站为污染物可以影响到的最大范围，其中符合要求的监测站如表3所示。

表3 符合夹角为锐角的监测站

|  |  |
| --- | --- |
| 监测站 | 夹角 |
| 市区世纪公园 | 63 |
| 市区西北水源 | 45 |
| 栾城通讯公司 | 87 |
| 元氏住建局 | 56 |
| 市区化工学校 | 81 |
| 市区封龙山 | 35 |

最后通过方法计算出监测站井陉县气象局到表3各点的传播代价，计算结果如表4所示。

表4 井陉县气象局到各个监测站的传播代价

|  |  |  |
| --- | --- | --- |
| 监测站 | 夹角 | 传播代价 |
| 市区世纪公园 | 63 | 0.002 |
| 市区西北水源 | 45 | 0.0015 |
| 栾城通讯公司 | 87 | 0.04 |
| 元氏住建局 | 56 | 0.025 |
| 市区化工学校 | 81 | 0.035 |
| 市区封龙山 | 35 | 0.015 |

根据表4的传播代价就可找出污染物的关键传播路径，即传播代价最小的站点即为下一个传播起点。其中由井陉县气象局为起点的关键传播路径为：

由表4中的风向夹角和传播代价可以看出，并不是所有的风向夹角的越小的监测站上的传播代价越小。这是因为，由于两个监测站之间污染物的传输，不仅仅受风向影响，风向只是决定了污染物的主要的传播风向，而无法反映出污染物对下一个站点的影响程度，其中地形因素、气象因素，都很大程度上决定了污染物是否能影响到其他站点。因此就需要通过污染物传播代价来衡量污染物的传播难易程度，并以此来找出关键传播路径。

3.4 实验结果与分析

本文的最终目的是找到污染物实时的传播规律，并依据此传播规律来寻找污染物的实时关键传播路径。第二章通过对污染物传播机理的详细分析，找出了影响污染物传播的关键因素，并基于污染物传播机理建立了污染物传播模型。第三章提出了污染物传播的约束条件和寻找污染物传播关键路径的详细步骤以及方法，本章针对上述寻找出的污染物关键路径进行试验，并对实验结果进行分析。

### 3.4.1 实验环境配置

本文采用的实验数据为\*.\*.\*中产生的一个周期内的大气污染物传播网络的邻接矩阵以及京津冀地区2014年5月1日到2015年4月30日的空气质量数据集airquality.csv。

### 3.4.2 实验数据

本文所需的数据集来源于微软亚洲研究院郑宇博士所提供京津冀地区2014年5月1日到2015年4月30日的数据集，数据集集中包含airquality.csv、city.csv、district.csv、meteorology.csv、station.csv五个文件。其中数据集详细解析如下：

表3.1 station.csv

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 站点编号 | 名称 | | 纬度 | | 经度 | 区域编号 |
| 1001 | | 海淀北部新区 | 40.090679 | 116.173553 | | 101 |
| 6010 | | 天津前进道 | 39.092699 | 117.201676 | | 607 |
| 14004 | | 建设大厦 | 39.942 | 119.537 | | 1401 |

station.csv文件中展示了京津冀范围内所有空气质量监测站的详细信息：站点编号、站点名称、站点所在位置的经纬度，站点所属区域编号。

表3.2 city.csv

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 城市编号 | 名称 | | 纬度 | | 经度 | 类别 |
| 001 | | 北京 | 39.904210 | 116.407394 | | 1 |
| 006 | | 天津 | 39.084158 | 117.200982 | | 1 |
| 011 | | 石家庄 | 38.042307 | 114.514860 | | 1 |

city.csv文件中展示了京津冀范围内所有的市级城市的详细信息：城市编号、城市名称、城市所在位置的经纬度、城市类别（1：北方；2：南方）。

表3.3 district.csv

|  |  |  |  |
| --- | --- | --- | --- |
| 区域编号 | 名称 | | 城市编号 |
| 00101 | | 海淀区 | 001 |
| 00102 | | 石景山区 | 001 |
| 00103 | | 丰台区 | 001 |

district.csv文件中展示了京津冀区域内所划分的区域（区、县）的详细信息：区域编号、区域名称、区域所属城市。

表3.4 airquality.csv

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 站点编号 | 日期 | PM2.5 | PM10 | NO2 | CO | O3 | SO2 |
| 1001 | 2014/5/1 00 | 138 | 159.4 | 56.3 | 0.9 | 50.8 | 17.2 |
| 1002 | 2015/4/20 10 | 45 | 41.6 | 56.7 | 0.7 | 39.8 | 2.9 |
| 1003 | 2015/3/10 02 | 11 | 30 | 34 | 0.6 | 37 | 5 |

airquality.csv文件展示了京津冀区域内每个空气质量监测站每个时刻（小时）采集到的污染物浓度信息，其中需要采集的污染物包括：PM25、PM10、NO2、CO、O3、SO2。它们的单位均除了污染物CO为以外，其余均为。由于本文中需要用到的是污染物的日均AQI值进行计算，因此需要将时均浓度值转化为日均值，并按照AQI的计算公式将浓度转化为AQI。

表3.5 meteorology.csv

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 区域编号 | 日期 | 温度 | 压强 | 湿度 | 风速 | 风向 |
| 616 | 2015/4/18 03 | 15 | 1009 | 65 | 1.8 | 1 |
| 1108 | 2015/4/6 10 | 7 | 1027 | 29 | 6.12 | 13 |
| 1405 | 2015/3/30 05 | 5 | 1010 | 85 | 3 | 24 |

meteorology.csv展示了每个区域在每一个时刻（小时）的气象状况。其中包含温度、压强、湿度、风速、风向五个数据。由于本文中风向需要用角度表示，因此需要根据十六风向图将该数据集中的风向转换为角度，其中风向标识与风向角度的对应关系如下表所示：

表 4 风向标识与风向角度对应关系

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 风向标识 | 描述 | 角度 | 风向标识 | 描述 | 角度 |
| 0 | 无风 | Null | 9 | 不稳定 | Null |
| 1 | 东风 |  | 13 | 东南风 |  |
| 2 | 西风 |  | 14 | 东北风 |  |
| 3 | 南风 |  | 23 | 西南风 |  |
| 4 | 北风 |  | 24 | 西北风 |  |

### 3.4.3 实验结果分析

本实验采用2015年2月6日凌晨3点京津冀各个监测站的气象数据、污染物浓度数据，并将算法SCP中的传播代价上限设置为0.01。通过算法SCP的计算，可计算出两个监测站之间是否会存在传播路径，结果如图3所示。

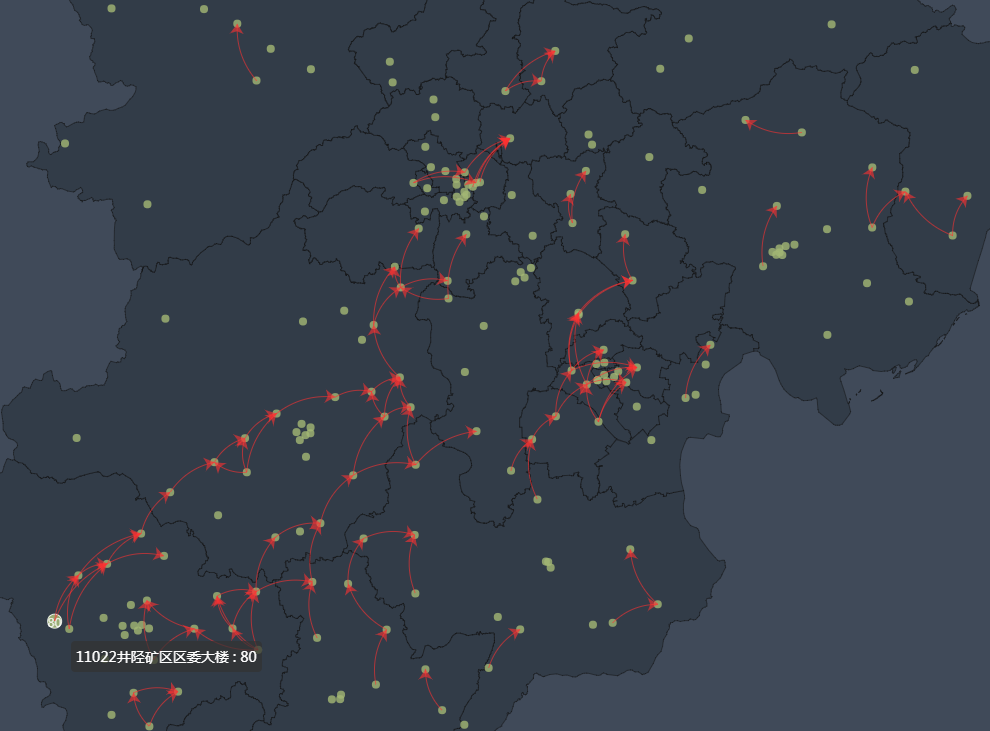


图3 京津冀地区2015年2月6日凌晨3点污染物传播路径

根据第三章所描述的污染物关键路径的搜索方法以及算法实例，本章的实验以2015年2月6日凌晨3点井陉矿区区委大楼监测站点为起点。由于所研究的数据量十分大，因此需要截取其中一部分进行实验分析，截取的路径部分如图4所示。

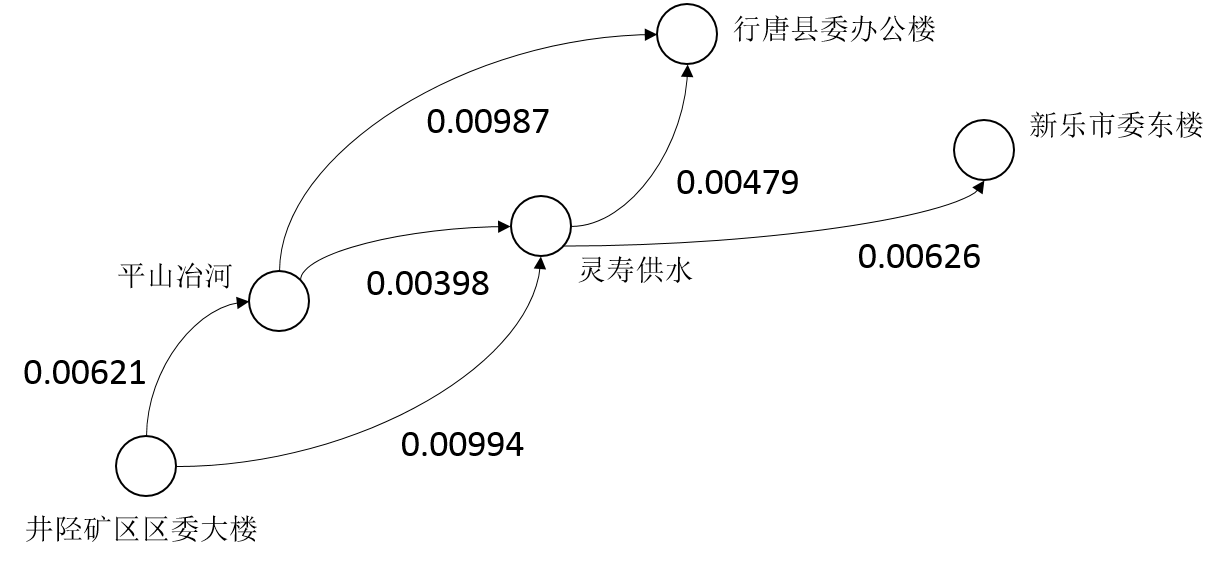


图4 2015年2月6日凌晨3点部分污染物传播路径

由图4可知，以井陉矿区区委大楼为起点的污染物传播路径存在两条，即

其边上的数值为此条路径上污染物的传播代价，根据污染物关键传播路径的定义，选取传播代价最小的一条边为关键传播路径，因此路径为关键路径，根据此方法，接下来以平山冶河监测站为起点，选取传播代价最小的路径为关键路径。然后，以灵寿供水监测站为起点，选取传播代价最小的路径为关键路径。最终，以井陉矿区区委大楼为起点的污染物关键传播路径为：

表5展示出了2015年2月6日凌晨3点以井陉矿区区委大楼为起点的完整的污染物关键传播路径：

表5 2015年2月6日凌晨3点污染物关键传播路径

|  |  |
| --- | --- |
| 监测站名称 | 监测站编号 |
| 井陉矿区区委大楼 | 11022 |
| 平山冶河 | 11023 |
| 灵寿供水 | 11024 |
| 行唐县委办公楼 | 11025 |
| 唐县政府楼  满城税务局  徐水环保局  容城县环境保护局  白沟新城行政中心楼  高碑店环保局  涿州监测站  京南榆垡 | 17017  17007  17009  17015  17022  17012  17011  1035 |

其中这条关键路径的方向是由上至下，即从井陉矿区区委大楼监测站到徐水环保局监测站的方向。由于这条关键传播路径是根据传播机理以及传播代价公式计算所得，因此它表征了污染物的在这个时刻的一个最易扩散的趋势，即当井陉矿区区委大楼的污染物浓度上升时，这条路径所经过的区域将有很大的概率也受到污染。并通过Echarts将关键路径在地图上展现出来，如图3所示。

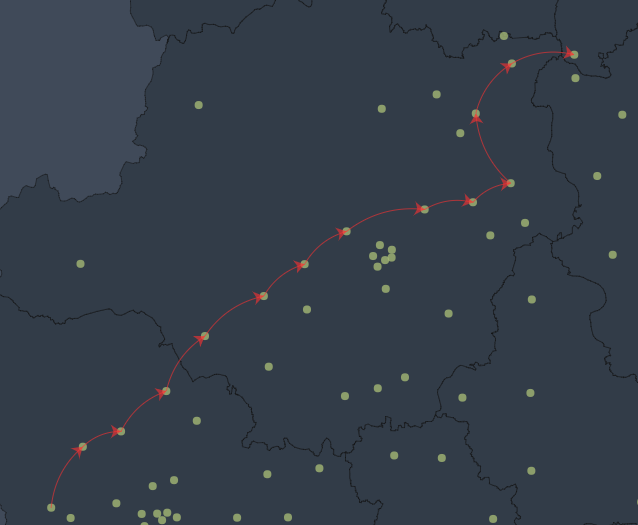


图3 关键传播路径

由图3可以直观的看出污染物传播路径的一个大致的方向，其中，在该路径的左侧是太行山山脉，由于受山势的影响，这里常年盛行西南风，图3所示的路径也恰恰反映出了这一特点，该路径正是沿着太行山山脉的走势，在西南风的影响下，途径保定，将污染物输送至北京，从而对北京的空气污染造成了一定的影响。

### 3.4.4 实验结果验证

接下来以路径为例，来分析验证污染物关键传播路径的科学性以及可行性。为了与这条关键路径形成对比，于是，选择了一个不在这条关键路径但与起始站点井陉矿区区委大楼距离非常接近的站点，即井陉县气象局站点。并根据各个监测站每个时刻的污染物浓度绘制出变化曲线，如图4、表7所示。

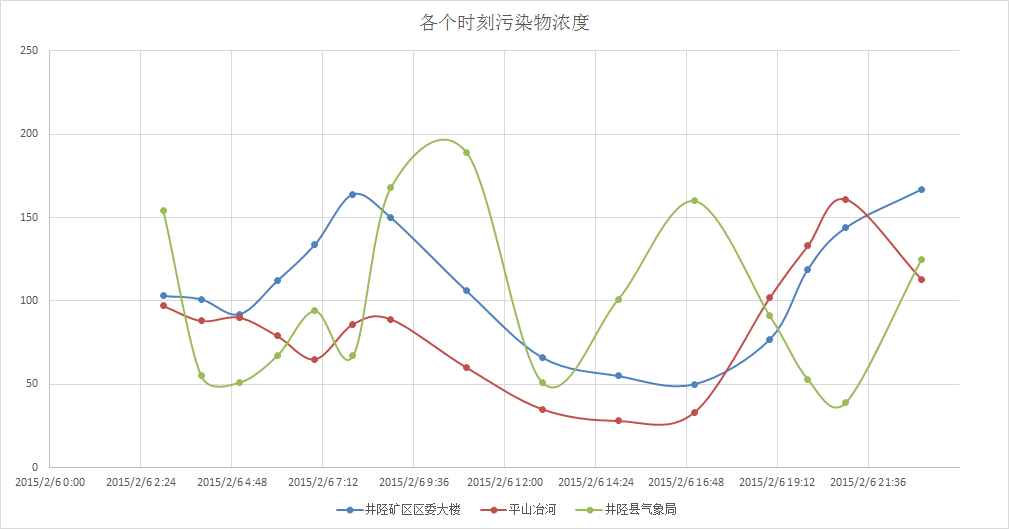


图4 各个时刻污染物浓度

表7 三个监测站各个时刻污染物浓度

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 时刻 | 井陉矿区区委大楼 | | 平山冶河 | 井陉县气象局 |
| 2015-02-06 03:00:00 | 103 | 97 | | 154 |
| 2015-02-06 04:00:00 | 101 | 88 | | 55 |
| 2015-02-06 05:00:00 | 92 | 90 | | 51 |

由于这是一条出现在6日凌晨3点的以井陉矿区区委大楼监测站为起点的关键路径，因此将着重分析凌晨3点及以后的各个时刻的污染物浓度的变化情况。

通过查询数据库中6日凌晨3点井陉矿区区委大楼的气象信息可知当时的风速为6.12m/s，并根据距离公式可计算出井陉矿区区委大楼与平山冶河的距离为24.7km。由于两地距离很近，因此不考虑风速变化的问题，这样就可计算出污染物扩散到平山冶河的大致耗时为1.12小时，即平山冶河监测站的污染物浓度将会在6日凌晨4点07分以后开始出现上升趋势。并且由图4可知，平山冶河监测站的污染物浓度确实在6日凌晨4时之后开始上升，其中就是由于受到了井陉矿区区委大楼污染物扩散的影响。

然后再看并不在这条路径上的井陉县气象局污染物的浓度，它的上升趋势是出现在了6日凌晨4点之后。根据距离公式可计算出井陉县气象局与井陉矿区区委大楼的距离为8.15km。如果路径为6日凌晨3点的污染物关键传播路径，那井陉县气象局污染物的浓度的上升点应该会在凌晨4点之前而不是之后。由此可以证明路径并不是6日凌晨3点以井陉矿区区委大楼监测站为起点的污染物关键传播路径。然后通过表8展现出其余各个子路径的实际情况，如表8所示。

表8 关键路径中子路径实际情况

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 起始点 | 风速(m/s) | 历时(h) | 下一站点 | 下一站点上升点 | 是否符合 |
| 平山冶河 | 7.92 | 0.54 | 灵寿供水 | 2015-02-06 03:32:00 | 是 |
| 灵寿供水 | 8.64 | 0.71 | 行唐县委办公楼 | 2015-02-06 03:43:00 | 否 |
| 行唐县委办公楼 | 6.84 | 2.0 | 唐县政府楼 | 2015-02-06 05:00:00 | 否 |
| 唐县政府楼 | 2.52 | 4.22 | 满城税务局 | 2015-02-06 07:13:00 | 是 |
| 满城税务局 | 6.12 | 1.35 | 徐水环保局 | 2015-02-06 04:21:00 | 是 |
| 徐水环保局 | 3.96 | 1.26 | 容城县环境保护局 | 2015-02-06 04:15:00 | 是 |
| 容城县环境保护局 | 8.28 | 0.52 | 白沟新城行政中心楼 | 2015-02-06 03:31:00 | 是 |
| 白沟新城行政中心楼 | 4.68 | 1.66 | 高碑店环保局 | 2015-02-06 04:39:00 | 是 |
| 高碑店环保局 | 4.68 | 1.32 | 涿州监测站 | 2015-02-06 04:19:00 | 是 |
| 涿州监测站 | 3.6 | 1.77 | 京南榆垡 | 2015-02-06 04:46:00 | 是 |

由表7可以计算出此条关键路径的准确度为：

因此可以说明该路径与实际情况基本吻合，能够准确的反映出在该时刻污染物的一个潜在的传播通道，可以为污染发生提供预警以及及时的防控作用。所以，由此可以证明该传播模型的科学性、准确性和实际应用中的可行性。

## 3.5 总结

为了更加全面的研究大气污染物的传播规律，本文对大气污染物的传播机制进行了深入的研究，分析出了影响大气污染物传播的几个必要条件。一个地区的大气污染不仅仅与本地的污染物排放有关系，而且与该地区的相邻地区的大气污染物的输送有着紧密的联系。影响大气污染物传输的因素有许多，其中风速、风向、降雨降雪、湿度、压强、地面粗糙度、地形等都对大气污染物的传输起着至关重要的作用。因此本文依据大气污染物的传播机制，推导出大气污染物的传播代价，以此来衡量大气污染物的传播的难易程度。并以此为依据，筛选出下一个将要传播至的点。并根据实际的大气污染物浓度变化的趋势对该路径做验证，证明了该路径的准确性，从而也证明了本文提出的污染物传播代价公式的科学性。同时，也为防治大气污染物传播、及时阻断大气污染物传播提供了理论依据。

第4章 基于复杂网络的关键传播路径挖掘

## 4.1 引言

## 4.2 构建大气污染物传播网络

根据第三章提出的大气污染物传播代价公式以及大气污染物传播模型，可以计算出每个时刻（小时）的大气污染物传播路径，由于本文是以年为周期进行的统计分析，因此在一个周期中会产生大量的污染物传播路径，然后将每个时刻的路径存储到一个矩阵当中，并将该位置的值置为1，若在另一个时刻该条路径再次出现时，则在之前的基础上进行加1，以此类推，最终会得到大气污染物在一个周期内传播的邻接矩阵，其中的矩阵中的值表示了该条路径在一个周期内出现的次数。其中，一个周期内大气污染物的传播网络图如下图所示。

在大气污染物传播网络中，节点代表空气污染物监测站点，边代表污染物从一个区域传播到另一个区域的路径，当在一个周期（年）中，大气污染物多次从一个地区传播到另一个地区时，就表明了这两个地区之间很可能存在一条有利于污染物传播的通道，因此，考虑节点之间的依赖关系主要是该路径在一个周期中出现的次数，即边上的权重代表可以用路径在一个周期中出现的次数来表示。

因此，这个大气污染物传播网络图可以表示为三元组，其中，表示网络中点的集合，表示网络中边的集合，表示网络中边上的权重的集合，由于本文中用到的大气污染物传播网络图是一个有向图，因此存在。其中计算权重的算法具体过程如下所示：

计算一个周期（年）内大气污染物传播网络边上的权值

|  |
| --- |
| 算法1 统计周期内大气污染物传播网络边上的权值 |
| 输入：站点的集合stationList和周期内传播路径集合pathList  输出：边出现次数集合martixNum   1. for(int i = 0; i < stationList.size(); i++){ 2. Map<String,Double> map = new LinkedHashMap<String,Double> 3. for(int j = 0; j < stationList.size(); j++){ 4. map.put(stationList.get(j), 0) //存储终点，并初始化为0 5. } 6. matrixNum.put(stationList.get(i), map) //存储起点和终点 7. } 8. for(int i = 0; i < pathList.size(); i++){ 9. for(int j = 0; j < pathList.get(i).size(); j++){ 10. matrixNum的.get(pathList.get(i)).put(pathList.get(i).get(j), matrixNum.get(i).get(j) + 1) 11. matrixVal.get(pathList.get(i)).put(pathList.get(i).get(j),matrixVal.get(i).get(j)+ cost) 12. } 13. } 14. } |

如算法一所示，首先对京津冀内所有的监测站集合进行遍历，并根据这些监测站初始化矩阵martixNum，矩阵中的行表示大气污染物传播的起点，列表示大气污染物传播的重点，矩阵的初始值均设置为0。然后对周期内大气污染物传播路径集合进行遍历，并将每条大气污染物传播路径中的路径片段存储到martixNum矩阵中，并在之前的数值基础上进行累加。最终会得到一个周期内的大气污染物传播网络的邻接矩阵。

### 4.2.1 基本概念与定义

通过上一节获得了大气污染物传播网络有向图，为了获得大气污染物的关键传播路径，本文提出了基于图序列的挖掘方法，此方法在挖掘的过程中需要对访问过的边和节点进行记录。因此，为了记录网络中节点或者边的访问情况，接下来定义了大气污染物传播网络有向加权标识图。

定义一 染物传播网络有向加权标识图：图是一个有向加权标识图，其中：是一个有向加权图，其符合大气污染物传播网络有向加权图的定义，是图 ，其作用是表示监测站是否已经被访问到。是图G中边标识集合，是标识边是否已经被访问过。和的值只有true和false两个。是一个映射函数，其作用是为有向图的节点和边分配标识，。

定义二 子图：存在大气污染物传播网络有向加权标识图和，若图和的节点、边以及映射函数满足条件且且,则将图称作图的子图，记作。

定义三 边序列：如果大气污染物传播网络有向加权标识图G中节点之间存在边，则将置为1，否则将置为0。最终将一条大气污染物传播路径映射成一个0/1的序列，并将该序列称之为边序列。

定义四 有向图矩阵APTM：如果已经知道图 G中的边与节点的出现情况，那么假设存在一个有向图矩阵，矩阵中每个位置的值由0/1字符串构成，表示该对节点对应的边的出现情况，将该图序列矩阵称之为大气污染物传播网络有向图矩阵 APTM，简称有向图矩阵 APTM。

定义五 关键路径序列：在大气污染物传播网络有向图 G中，若存在一个频繁连通子图可以通过序列表示，并且该频繁联通子图中所有的边都满足了预先设置的频繁阈值的限制，则称该序列为大气污染物传播的关键路径序列。

## 4.3 基于图矩阵的污染物关键传播路径挖掘

本文通过图矩阵的数据结构存储大气污染物在两个地区之间的传播关系，并使用图序列挖掘方法来挖掘出大气污染物的关键传播路径序列。

首先，构建大气污染物传播网络，根据第三章提出的大气污染物传播机理以及传播代价公式，获取含有节点、边以及传播代价的调用关系的3列矩阵。

其次，根据一个周期内的空气质量监测站的污染物浓度的数据信息，统计出京津冀地区各个城市污染物的超标率，并将这些超标的城市分别作为大气污染物传播的起点，获取若干个污染物传播的路径，可将这些污染物传播路径看成是大气污染物传播网络有向图的一个子图。并根据定义三，将路径中边的出现情况按照子图的顺序依次用0/1表示，构建有向图矩阵APTM。

再次，通过预先设置的频发阈值，根据阈值的大小筛选出满足要求的序列，并将这些满足要求的序列转换成十进制数，这样便于比较。

最后，有向图矩阵APTM的数值相同的边就是频繁图序列，即大气污染物传播网络的关键传播路径。

### 4.3.1 生成污染物传播网络子图

为了生成大气污染物传播网络子图，则需要获取大气污染物传播的起始点，本文采用一个周期（年）内污染物超标率比较严重的地区作为起始点，具体算法过程如算法4所示。然后分别以这些地区为起点对污染物传播网络进行深度优先遍历，从获得若干条大气污染物传播路径序列，这些序列即为大气污染物传播网络子图，具体算法过程如算法4所示。

算法4 统计一个周期内京津冀各城市的污染物超标率

输入：

输出：

执行过程：

1. //对每行数据按逗号进行分割,获取每个时刻的污染物浓度信息
2. ;
3. ;






11. ;
12. ;
13. ;
15. ;

算法4中首先初始化用于存储一个地区的污染物浓度总和、污染天数、污染物超标天数，然后对一个周期（年）内的京津冀大气污染物浓度的信息进行遍历，由于每一条信息是按逗号分隔的，因此将其分割，并取得对应城市的编号以及该地区的污染物浓度值（5-7行），若中不存在对应的地区的值，则初始化该地区的污染物浓度总和、污染天数、污染物超标天数均为0，否则将该地区每一时刻（小时）的污染物浓度进行累加，污染天数进行累加，如果该时刻的污染物浓度超出了阈值，则将污染物超标天数进行累加（9-12行）。最后，按对进行遍历，计算出每个城市一个周期内平均污染物浓度以及污染物超标率（21-23行）。算法结束。

|  |
| --- |
| 算法5 根据传播网络计算传播路径 |
| 输入：传播网络邻接矩阵M、城市编号cityId  输出：传播路径集合path   1. for (int i = 0; i < s.size(); i++) { 2. String[] ss = s.get(i).split(",") 3. if (!map.containsKey(ss[0]+ss[2])) { 4. Set<String> set = new HashSet<>() 5. set.add(ss[3]+ss[5]) 6. map.put(ss[0]+ss[2], set) 7. } else { 8. map.get(ss[0]+ss[2]).add(ss[3]+ss[5]) 9. } 10. } 11. Set<String> keyset = map.keySet() 12. List<String> marked = new ArrayList<String>() 13. for (String key : keyset) { 14. int[] len={0} 15. if(marked.contains(key)) continue 16. list.add(dfs(key, map, marked,new ArrayList<String>(),len)) 17. } 18. return list; 19. public List<String> dfs(String key, Map<String, Set<String>> map, List<String> marked,List<String> slist,int[] len) { 20. if (marked.contains(key)) { 21. if(slist.size()>len[0]){ 22. maxlist=new ArrayList<>(slist); 23. len[0]=slist.size(); 24. } 25. }else { 26. marked.add(key); 27. slist.add(key); 28. Set<String> values = map.get(key); 29. if (values != null) { 30. for (String value : values) { 31. dfs(value, map, marked,slist,len); 32. } 33. }else{ 34. if(slist.size()>len[0]){ 35. maxlist=new ArrayList<>(slist); 36. len[0]=slist.size(); 37. } 38. } 39. slist.remove(key); 40. } 41. return maxlist; 42. } |

代码分析

### 4.3.2 构建有向图矩阵

由算法5获取的大气污染物传播路径即为大气污染物传播网络有向图的子图，然后根据子图中的信息来构建定义4中的有向图矩阵APTM。首先，获取子图中所有的节点。其次，顺序遍历所有子图，将节点编号看作矩阵的行列号，如果子图中有对应边的存在，则在对应子图的位置上置为1，否则置为0。再次，在所有的子图中，可能存在有多个子图都包含某条边，所以将所有子图对应边位置上的0/1值拼接起来形成新的字符串。最后，将刚才得到的字符串序列存储到矩阵的对应位置中，即形成了有向图矩阵APTM。其具体的算法过程如下所示。

算法5 构建有向图矩阵

输入：

输出：APTM

BEGIN

1. ：




7. ;
9. ;
11. }

在算法5中，是所有大气污染物传播网络的子图的集合，是根据中的三列构造的邻接矩阵。首先，初始化矩阵，用来存储所有子图中涉及到的节点（2-8行），根据中的节点来初始化矩阵，节点编号作为行列号。其次，将每个大气污染物传播网络子图转换成邻接矩阵（12-16行），并将每个矩阵中的元素转换成字符，把每个每个矩阵中对应行列位置的字符拼接起来，形成字符串序列（17-18行）。最后，将所有的序列按照对应的行列存储到矩阵中（19行）。算法结束。

### 4.3.3 构建简化图矩阵

对算法5构造出的大气污染物传播网络有向图矩阵按照规则进行搜索判断，从而获得简化图矩阵。首先，将非频繁出现的边剔除，即逐行扫描有向图矩阵中的各边，统计处每条边在子图中出现的次数。其次，根据边中1的数量和预先设定好的阈值进行比较来判断是否是频繁边，如果不是，则将该边置为0。最后，将所有符合条件的0/1序列转换成十进制数，为后续挖掘污染物关键传播路径提升效率做准备。

算法6 构建简化图矩阵

输入：矩阵和频繁阈值

输出：

BEGIN

1. //表示计算中1的个数
2. 置0
4. 置0

在算法6中，是对序列中1的个数的统计方法。首先，对有向图矩阵进行遍历搜索判断，对每个序列中1的个数进行统计判断（2行），若的值小于预先设置的频繁阈值,则将其视为非频繁边，将置为0，并将该序列从向图矩阵中移除。若大于,则判断是否至少t个连续路径中出现1，如果没有，则认为其不符合规则边的要求，将置为0（3-4行）。否则，将序列转换成十进制数并根据对应的行列号存入简化图矩阵中（6行），最后得到简化图矩阵（12行）。算法结束。

### 4.3.4 挖掘污染物关键传播路径

大气污染物的关键传播路径是由多个满足频繁阈值并有周期规律的路径片段组成。通过算法6获得污染物传播网络中的频繁出现的边，并根据这些边生成了简化图矩阵。然后采用深度优先遍历的方法对矩阵进行遍历，搜索数值相同的路径，从而得到该路径片段周期出现的次数。为了更加体现得到的路径片段的重要性以及关键性，本文中预先设定了路径片段周期出现次数的阈值,即由超出阈值的路径片段组成的路径为大气污染物关键传播路径。具体的算法实现如下所示。

算法7 污染物关键传播路径挖掘算法

输入：，

输出；

Begin

5. ;



10. 将中的序列按顺序拼接成字符串
11. output
12. END

算法6 是对矩阵中每一个路径片段进行判断是否呈现出周期性性来获得污染物关键传播路径。首先，对矩阵进行遍历，判断矩阵中的路径片段是否为0，如果为0，则判断下一条路径片段。否则将该路径片段存储到数组a中（2-4行），并使用函数去除矩阵a中的重复路径片段。其次，如果数组a不为空，则对矩阵进行深度优先遍历，判断是否存在有与数组a的数值相同的路径片段。如果存在，则将该路径片段对应的行列号存储到字符串q中，加1，否则将当前值从数组a中删除。然后继续依据上述方法，继续遍历数组a其他的值（7-12行）。最后，计算每个的值，如果小于1，则删除对应的（13-15行）。最后将满足条件的拼接成字符串即为污染物关键传播路径。算法结束。

### 4.3.5 实例说明

接下来对以衡水地区为起点的污染物传播网络进行关键传播路径的挖掘。首先对一个周期内的污染物传播网络图进行子图挖掘，根据算法4挖掘出以衡水地区为起点的所有传播路径，共挖掘出10条传播路径，其中共有30个节点，这些路径组成了污染物传播子图集合。

由于数据量较大，因此本文只展示了其中的4条传播路径，如下表所示。

|  |  |
| --- | --- |
| 序号 | 路径 |
| 1 | 23012景县县委,23011阜城县法院,23006武强县环保局,23009饶阳县环保局,21011肃宁县政府,21009河间市交通局,21006任丘华油八处,22007文安县环保局,22009霸州顺达燃气,22005固安党校,17011涿州监测站,1005房山良乡,1007朝阳奥体中心,1024昌平镇，1027怀柔镇,1028密云镇,1032京东北密云水库 |
| 2 | 23007故城县县政府,23010枣强县环保局,23005冀州环保局,23013武邑县环保局,23006武强县环保局,23009饶阳县环保局,21011肃宁县政府,21009河间市交通局,21006任丘华油八处,22007文安县环保局,22009霸州顺达燃气,22005固安党校,17011涿州监测站,1005房山良乡,1007朝阳奥体中心,1024昌平镇,1027怀柔镇,1028密云镇,1032京东北密云水库 |
| 3 | 23010枣强县环保局,23005冀州环保局,23013武邑县环保局,23006武强县环保局,23009饶阳县环保局,21011肃宁县政府,21009河间市交通局,21006任丘华油八处,22007文安县环保局,22009霸州顺达燃气,22005固安党校,17011涿州监测站,1005房山良乡,1007朝阳奥体中心,1024昌平镇,1027怀柔镇,1028密云镇,1032京东北密云水库 |
| 4 | 23004安平县环保局,17024安国市政府,17027望都县环境监测站,17007满城税务局,17009徐水环保局,17015容城县环境保护局,17022白沟新城行政中心楼,22009霸州顺达燃气,22005固安党校,17011涿州监测站,1005房山良乡,1007朝阳奥体中心,1024昌平镇,1027怀柔镇,1028密云镇,1032京东北密云水库 |

如上表所示，由于每条传播路径的长度各不相同，其中包含的节点也不同，因此需要将每条传播路径转换邻接矩阵，矩阵的大小应该由节点的个数决定。因此，该邻接矩阵是一个大小的矩阵，然后根据每条路径片段在每个子图中的出现情况，进一步生成以石家庄地区为起点的污染物传播网络有向矩阵APTM，矩阵中的每一项表示该路径片段在每个子图中出现的情况。由于矩阵过大，因此只展示部分矩阵中的数据，如表\*\*所示。

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 23011 | 23006 | 23009 | 21011 | 21009 | 21006 | 22007 | 22009 | 22005 | 17011 |
| 23011 | 00000  00000 | 10000  00100 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 |
| 23006 | 00000  00000 | 00000  00000 | 11101  11100 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 |
| 23009 | 00000  00000 | 00000  00000 | 00000  00000 | 11101  11101 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 |
| 21011 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 11101  11101 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 |
| 21009 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 11101  11101 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 |
| 21006 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 11101  11101 | 00000  00000 | 00000  00000 | 00000  00000 |
| 22007 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 11101  11101 | 00000  00000 | 00000  00000 |
| 22009 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 11111  11111 | 00000  00000 |
| 22005 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 11111  11111 |
| 17011 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 | 00000  00000 |

通过算法6来统计有向图矩APTM中每一项里1的个数，并设置频繁阈值为2，通过统计可知，边<23012景县县委，23011阜城县法院局>、<23007故城县县政府，23010枣强县环保局>、<23004安平县环保局，17024安国市政府>、<23008深州市政府，11016深泽供电局>中只出现了1次1，因此将其所在位置只为0，然后将剩余的边序列由二进制转换为十进制数，所生成的矩阵即为简化图矩阵。具体结果如下表所示。

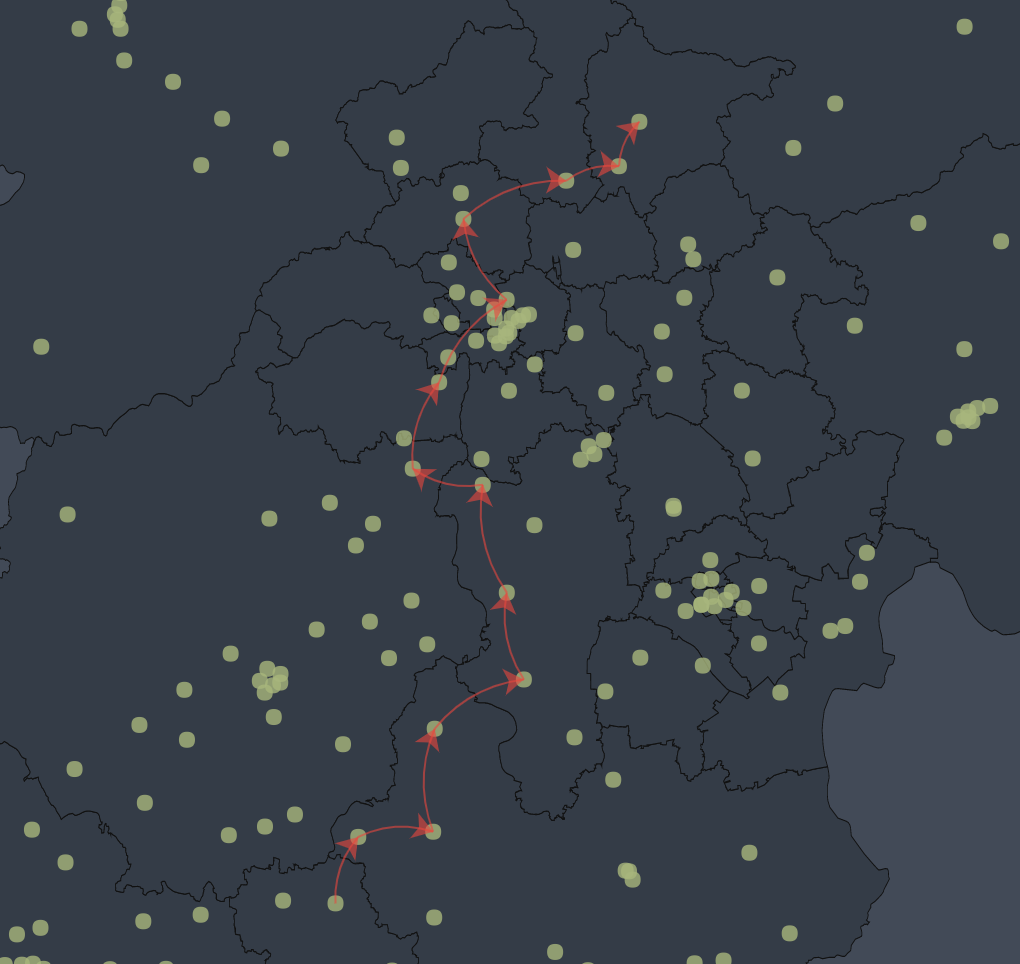
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 23011 | 23006 | 23009 | 21011 | 21009 | 21006 | 22007 | 22009 | 22005 | 17001 |
| 23011 | 0 | 516 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23006 | 0 | 0 | 956 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23009 | 0 | 0 | 0 | 957 | 0 | 0 | 0 | 0 | 0 | 0 |
| 21011 | 0 | 0 | 0 | 0 | 957 | 0 | 0 | 0 | 0 | 0 |
| 21009 | 0 | 0 | 0 | 0 | 0 | 957 | 0 | 0 | 0 | 0 |
| 21006 | 0 | 0 | 0 | 0 | 0 | 0 | 957 | 0 | 0 | 0 |
| 22007 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 957 | 0 | 0 |
| 22009 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1023 | 0 |
| 22005 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1023 |
| 17001 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

在表中，经过频繁阈值的筛选和十进制数转换之后，数据的分布情况即可清晰可见。在简化图矩阵中，经过算法6的筛选之后，存在着相同二进制数的边，其值分别为957、1023、66、516、956、432、384、400。其中各个值对应的边的情况如下表所示

|  |  |
| --- | --- |
| 值 | 路径片段 |
| 957 | 23009饶阳县环保局,21011肃宁县政府  21011肃宁县政府,21009河间市交通局  21009河间市交通局,21006任丘华油八处  21006任丘华油八处,22007文安县环保局  22007文安县环保局,22009霸州顺达燃气 |
| 1023 | 22009霸州顺达燃气,22005固安党校  22005固安党校,17011涿州监测站  17011涿州监测站,1005房山良乡  1005房山良乡,1007朝阳奥体中心  1007朝阳奥体中心,1024昌平镇  1024昌平镇,1027怀柔镇  1027怀柔镇,1028密云镇  1028密云镇,1032京东北密云水库 |
| 66 | 17024安国市政府,17027望都县环境监测站  17027望都县环境监测站,17007满城税务局  17007满城税务局,17009徐水环保局  17009徐水环保局,17015容城县环境保护局  17015容城县环境保护局,17022白沟新城行政中心楼  17022白沟新城行政中心楼,22009霸州顺达燃气 |
| 516 | 23011阜城县法院,23006武强县环保局 |
| 956 | 23006武强县环保局,23009饶阳县环保局 |
| 432 | 23013武邑县环保局,23006武强县环保局 |
| 384 | 23010枣强县环保局,23005冀州环保局 |
| 400 | 23005冀州环保局,23013武邑县环保局 |

由表可知数值为516、956、432、384、400的边只有一条，设,则去除数值为516、956、432、384、400的边。最后以衡水地区的监测站为起点，将路径片段组合起来，结果可得污染物关键传播路径为：

23009饶阳县环保局，21011肃宁县政府，21009河间市交通局，21006任丘华油八处，22007文安县环保局，22009霸州顺达燃气，22005固安党校，17011涿州监测站，1005房山良乡，1007朝阳奥体中心，1024昌平镇，1027怀柔镇，1028密云镇，1032京东北密云水库，该大气污染物的关键传播路径图如图所示。



由图\*可以看出，该污染物关键传播路径由衡水地区出发，途径沧州、廊坊、保定最终到达北京地区。

## 4.4 实验结果分析

### 4.4.1 实验环境配置

实验的系统环境配置为macOS 10.13.2操作系统，硬件环境配置为Intel Core i5，主频为3.10GHz，内存为16 GB 2133 MHz LPDDR3。算法采用Java，Matlab编写。数据展示采用Echarts.js以及Excel，运行环境为IntelliJ IDEA，JDK版本为1.7，采用的数据库为MySQL数据库。

### 4.4.2 实验数据

本文采用的实验数据为\*.\*.\*中产生的一个周期内的大气污染物传播网络的邻接矩阵以及京津冀地区2014年5月1日到2015年4月30日的空气质量数据集airquality.csv。

### 4.4.3 实验结果分析

由于本文中将小时作为一个时刻，将一年作为一个周期。因此，在一个周期中就会有8760个时刻，并且每个时刻中会出现若干条污染物传播路径，因此数据量是十分巨大的，所以本文只展示部分实验数据。通过对京津冀地区2014年5月1日到2015年4月30日期间每个站点之间的污染物传播代价进行计算，并最终计算出了每条边的平均传播代价，如表1所示。

表1 周期内站点间平均传播代价

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 监测站名称 | 海淀北  部新区 | 海淀北京植物园 | | 石景山古城 | 丰台  云岗 | 房山  良乡 | 海淀  万柳 | 朝阳奥  体中心 |
| 海淀北部新区 | 1 | 1 | 1 | | 1 | 1 | 1 | 1 |
| 海淀北京植物园 | 0.6456 | 1 | 1 | | 1 | 1 | 1 | 1 |
| 石景山古城 | 0.7369 | 0.6476 | 1 | | 1 | 1 | 0.6736 | 0.6424 |
| 丰台云岗 | 0.7675 | 0.745 | 0.6779 | | 1 | 1 | 0.7486 | 0.8037 |
| 房山良乡 | 0.9308 | 0.8633 | 0.7853 | | 1 | 1 | 0.8557 | 0.687 |
| 海淀万柳 | 0.7263 | 1 | 1 | | 1 | 1 | 1 | 1 |
| 朝阳奥体中心 | 0.7409 | 0.7536 | 1 | | 1 | 1 | 1 | 1 |

通过表1可以看出，该表不是一个对称结构的表格，也就是说明只有在特定因素条件的制约下，相应的路径才能成为比较有利于污染物传播的一条路径，而在其他的条件下可能不利于污染物的传播。并通过图0展示出周期内京津冀地区污染物传播网络图：

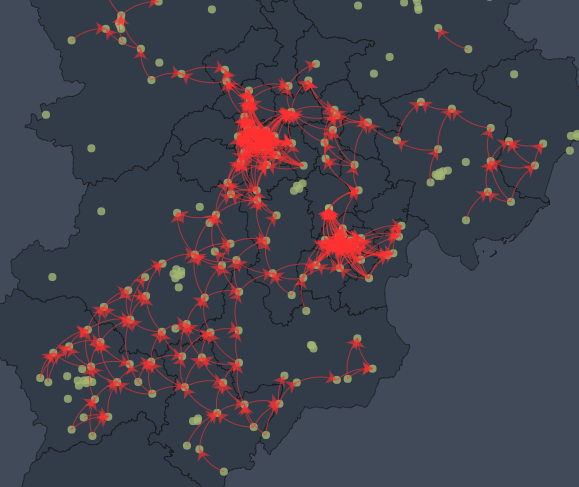


图0 周期内京津冀污染物传播网络图

由图0可知，北京和天津地区城市内部污染物传播路径十分的复杂以及繁多，而其他城市的污染物传播路径则稀疏许多。这是因为这两个城市的经济与其他城市相比比较发达，城市群落比较密集，高层建筑较多，因此地面粗糙度比较高，这使得气流沿地表流过时，同地面的建筑物发生摩擦作用，使风向风速发生变化，因此导致了这两个城市内部形成了复杂的污染物传播网络。

通过算法\*统计出一个周期内京津冀各个城市的污染物超标率，如图1所示。由图可见廊坊、衡水、石家庄、唐山、沧州、保定六个城市的超标率超过了65%。

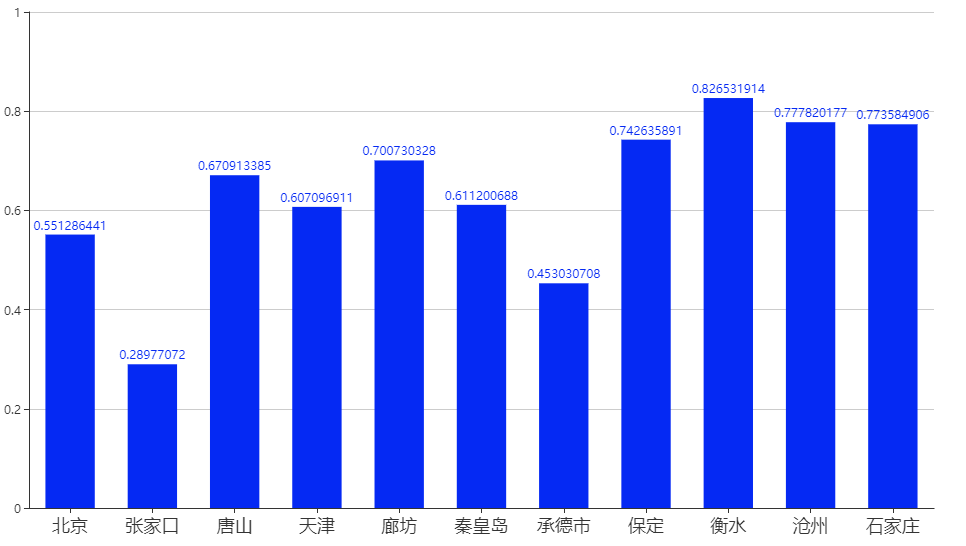


图1 京津冀各城市污染物超标率

由图1可知廊坊、衡水、石家庄、唐山、沧州、保定这几个地区的空气污染比较严重，而秦皇岛、北京、承德、天津空气污染情况相对较轻，只有张家口一个城市污染物超标率是低于30%的，是空气质量最好的城市。这些数据的统计结果与近几年京津冀各城市的空气质量状况十分吻合，由此证明了统计结果的准确性。因此，选取廊坊、衡水、石家庄、唐山、沧州、保定这几个地区中的监测站为起点，对污染物关键传播路径进行挖掘。

首先，根据算法\*计算出以每个地区为起点的污染物传播子图个数以及节点个数。经计算，各地区的污染物传播子图个数以及节点个数情况如下表所示：

其次，根据算法\*和算法\*计算出每个地区污染物传播网络的有向图矩阵以及简化矩阵。但是，由于每个矩阵中的数据量十分巨大，因此本章不再对污染物传播子图、构建有向图矩阵以及构建简化矩阵的数据进行展示，具体的计算过程可见上一章的算法实例说明，本章只对符合频发阈值以及呈周期出现的边进行展示，如表1-6所示。

表 1 以廊坊地区监测站为起点的路径片段表

|  |  |
| --- | --- |
| 值 | 路径片段 |
| 184 | 22010三河交通局,1023顺义新城、1023顺义新城,1024昌平镇、1024昌平镇,1031京西北八达岭、1031京西北八达岭,19009怀来府前小学 |
| 71 | 1024昌平镇,1027怀柔镇、1007朝阳奥体中心,1024昌平镇、1027怀柔镇,1028密云镇、1028密云镇,1032京东北密云水库 |
| 69 | 22005固安党校,17011涿州监测站、17011涿州监测站,1005房山良乡、1005房山良乡,1007朝阳奥体中心 |

表 2 以衡水地区监测站为起点的路径片段表

|  |  |
| --- | --- |
| 值 | 路径片段 |
| 957 | 23009饶阳县环保局,21011肃宁县政府、21011肃宁县政府,21009河间市交通局、21009河间市交通局,21006任丘华油八处、21006任丘华油八处,22007文安县环保局、22007文安县环保局,22009霸州顺达燃气 |
| 1023 | 22009霸州顺达燃气,22005固安党校、22005固安党校,17011涿州监测站、17011涿州监测站,1005房山良乡、1005房山良乡,1007朝阳奥体中心、1007朝阳奥体中心,1024昌平镇、1024昌平镇,1027怀柔镇、1027怀柔镇,1028密云镇、1028密云镇,1032京东北密云水库 |
| 66 | 17024安国市政府,17027望都县环境监测站、17027望都县环境监测站,17007满城税务局、17007满城税务局,17009徐水环保局、17009徐水环保局,17015容城县环境保护局、17015容城县环境保护局,17022白沟新城行政中心楼、17022白沟新城行政中心楼,22009霸州顺达燃气 |

表 3 以石家庄地区监测站为起点的路径片段表

|  |  |
| --- | --- |
| 值 | 路径片段 |
| 400 | 11016深泽供电局,17024安国市政府、17024安国市政府,17027望都县环境监测站 |
| 65535 | 17027望都县环境监测站,17007满城税务局、17007满城税务局,17009徐水环保局、17009徐水环保局,17015容城县环境保护局、17015容城县环境保护局,17022白沟新城行政中心楼、17022白沟新城行政中心楼,22009霸州顺达燃气、22009霸州顺达燃气,22005固安党校、22005固安党校,17011涿州监测站、17011涿州监测站,1005房山良乡、1005房山良乡,1007朝阳奥体中心、1007朝阳奥体中心,1024昌平镇、1024昌平镇,1027怀柔镇、1027怀柔镇,1028密云镇、1028密云镇,1032京东北密云水库 |

表 4 以唐山地区监测站为起点的路径片段表

|  |  |
| --- | --- |
| 值 | 路径片段 |
| 168 | 13018玉田县政府,6015天津东环路、6015天津东环路,1033京东东高村、6015天津东环路,1033京东东高村、6015天津东环路,1033京东东高村、6015天津东环路,1033京东东高村、6015天津东环路,1033京东东高村 |

表 5 以沧州地区监测站为起点的路径片段表

|  |  |
| --- | --- |
| 值 | 路径片段 |
| 1729 | 21006任丘华油八处,22007文安县环保局、22007文安县环保局,22009霸州顺达燃气、22009霸州顺达燃气,22005固安党校、  22005固安党校,17011涿州监测站、17011涿州监测站,1005房山良乡、1005房山良乡,1007朝阳奥体中心、1007朝阳奥体中心,1024昌平镇、1024昌平镇,1027怀柔镇、1027怀柔镇,1028密云镇、1028密云镇,1032京东北密云水库 |

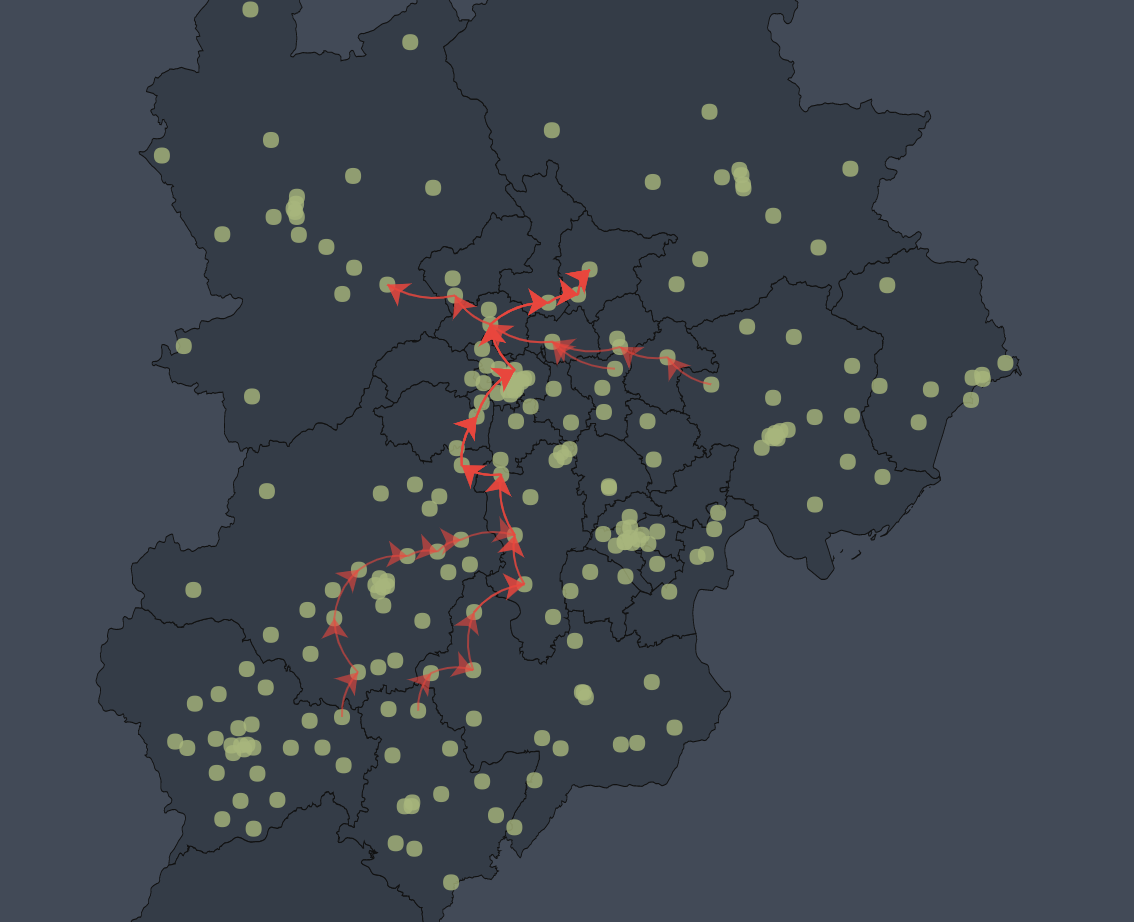
表 6 以保定地区监测站为起点的路径片段表

|  |  |
| --- | --- |
| 值 | 路径片段 |
| 30207 | 22009霸州顺达燃气,22005固安党校、22005固安党校,17011涿州监测站 |
| 32767 | 17011涿州监测站,1005房山良乡、1005房山良乡,1007朝阳奥体中心、1007朝阳奥体中心,1024昌平镇、1024昌平镇,1027怀柔镇、1027怀柔镇,1028密云镇、1028密云镇,1032京东北密云水库 |

根据表1-6中的数据，以每个地区中的监测站为出发点，通过算法7可计算污染物关键传播路径。如下表所示。

|  |  |
| --- | --- |
| 地区 | 关键传播路径 |
| 廊坊 | 22010三河交通局,1023顺义新城,1024昌平镇,1031京西北八达岭,19009怀来府前小学 |
| 22005固安党校,17011涿州监测站, 1005房山良乡,1007朝阳奥体中心, 1024昌平镇,1027怀柔镇, 1028密云镇,1032京东北密云水库 |
| 衡水 | 23009饶阳县环保局,21011肃宁县政府, 21009河间市交通局,21006任丘华油八处, 22007文安县环保局,22009霸州顺达燃气, 22005固安党校,17011涿州监测站, 1005房山良乡,1007朝阳奥体中心, 1024昌平镇,1027怀柔镇, 1028密云镇,1032京东北密云水库 |
| 石家庄 | 11016深泽供电局,17024安国市政府, 17027望都县环境监测站,17007满城税务局, 17009徐水环保局,17015容城县环境保护局, 17022白沟新城行政中心楼,22009霸州顺达燃气, 22005固安党校,17011涿州监测站, 1005房山良乡,1007朝阳奥体中心, 1024昌平镇,1027怀柔镇, 1028密云镇,1032京东北密云水库 |
| 唐山 | 13018玉田县政府,6015天津东环路, 1033京东东高村,1023顺义新城, 1024昌平镇,1031京西北八达岭,19009怀来府前小学 |
| 沧州 | 21006任丘华油八处,22007文安县环保局, 22009霸州顺达燃气,22005固安党校, 17011涿州监测站,1005房山良乡, 1007朝阳奥体中心,1024昌平镇, 1027怀柔镇,1028密云镇,1032京东北密云水库 |
| 保定 | 17011涿州监测站,1005房山良乡, 1007朝阳奥体中心,1024昌平镇, 1027怀柔镇,1028密云镇,1032京东北密云水库 |

最后，通过Echarts将各地区的污染物关键传播路径绘制在地图上，如下图所示。



其中，红色路线为各城市的污染物关键传播路径，其中部分路径存在重叠部分。由上图可以看出，在京津冀地区的西南部，污染物沿着太行山脉的走势由石家庄地区将污染物输送至北京地区，途中与保定、廊坊地区的污染物传播路径想汇合；东南方向污染物由沧州和天津输送至北京地区，途中与廊坊地区的污染物传播路径相汇合。东北方向由唐山地区沿燕山山脉传播至北京地区，途中与天津地区的污染物传播路径相汇合。由此可见，此污染物的关键传播路径与实际中的地理特征以及气象特征基本吻合，即证明了此挖掘关键路径方法的科学性以及准确性。

## 4.4本章小结

本章研究内容是在上一章大气污染物传播机理以及传播代价公式的基础上进行的研究，主要对大气污染物传播网络的关键路径进行了挖掘。首先，构建大气污染物传播网络，并描述了传播网络中的基本概念以及关键传播路径的定义。其次。统计出一个周期内污染物超过超标率的城市，并以这些城市中的监测站为起始点，对传播网络进行深度优先遍历，获得若干个大气污染物传播网络子图。再次，详细介绍了挖掘大气污染物关键传播路径的详细的原理以及过程，并给出了相应的算法实例。最后，对挖掘关键路径的试验结果进行了展示以及分析。

第5章 生物融合式康复机构型综合研究

## 5.1 节标题

## 5.4 本章小结

结 论

学位论文的结论作为论文正文的最后一章单独排写，但不加章标题序号。

结论应是作者在学位论文研究过程中所取得的创新性成果的概要总结，不能与摘要混为一谈。博士学位论文结论应包括论文的主要结果、创新点、展望三部分，在结论中应概括论文的核心观点，明确、客观地指出本研究内容的创新性成果（含新见解、新观点、方法创新、技术创新、理论创新），并指出今后进一步在本研究方向进行研究工作的展望与设想。对所取得的创新性成果应注意从定性和定量两方面给出科学、准确的评价，分（1）、（2）、（3）…条列出，宜用“提出了”、“建立了”等词叙述。

参考文献

1. 戴红. 康复医学[M]. 北京: 人民卫生出版社, 1998: 3-10.
2. Krebs H I., Palazzolo J, Dipietro L, et al. Rehabilitation Robotics: Performance-Based Progressive Robot-Assisted Therapy[J]. Autonomous Robots, 2003, 15(1): 7-20..

攻读博士学位期间承担的科研任务与主要成果

**（一）参与的科研项目**

**（二）发表的学术论文**

1. ×××,×××.并联2-RRR/UPRR踝关节康复机器人机构及其运动学[J]. 机器人, 2010, 32(1):6-12 .(EI收录号: 20101212786168)

致 谢