



北京大学

# 硕士研究生学位论文

题目：面向高速路网的关键路段识别  
研究与实现

姓 名： 刘丹萌

学 号： 1401214385

院 系： 北京大学

专 业： 计算机科学与技术(智能科学与技术)

研究方向： 数据仓库与数据挖掘

导 师： 宋国杰

2017 年 5 月



## 摘要

交通问题是当今世界关注的热点问题。随着人们生活水平的提高、交通系统的发展，社会对交通的需求也日益增大。交通与环境、交通与能源、交通与需求之间的矛盾日益加剧，交通事故和堵塞给人们带来了巨大的效率、能源和生命上的损失。在高速公路的建设中，如何选取关键路段，并对关键路段进行针对性建设已经成为一个热点问题。

在传统交通关键路段研究中，研究方法主要有①针对交通网络拓扑结构的研究；②针对交通网络统计数据进行分析；③在微观层面下，针对路段的特性进行研究。这些方法都有各自的局限性。方法①只针对了路网的拓扑结构，没有考虑节点之间的信息交流；方法②基于统计学方法，该方法只能静态分析路网关键路段，无法分析关键路段随着时间、路网流量的变化规律；方法③主要研究交通的微观领域特性，没有考虑路网的宏观变化规律。在此我们提出一种基于宏观高速公路网络的目标模型。

本文从智能交通的实际应用角度出发，针对现有关键路段挖掘方法的不足，提出两个研究方法：①提出一种基于贪心算法的关键路段挖掘模型，并证明贪心算法的可行性；②提出一种基于复杂网络社群划分的关键路段挖掘模型，实现动态挖掘关键路段。

集成上述成果，实现了一个基于 B/S 架构的高速公路关键路段挖掘系统，并且在真实应用场景下初步验证了原型系统的可靠性和适用性。

**关键词：**复杂网络，智能交通，关键路段，社群划分



# **Research And Implementation Of Key Road Identification For Highway network**

Danmeng Liu (Computer Science)

Directed by Associate Professor. Guojie Song

## **ABSTRACT**

Traffic problem is a hot issue in the world today. With the improvement of people's living standard and the development of traffic system, the demand for traffic is also increasing. The contradiction between traffic and environment, transportation and energy, traffic and demand is becoming more and more serious. Traffic accidents and congestion have brought great efficiency, energy and life losses. In the construction of expressway, how to select the key sections and carry out the targeted construction of key sections has become a hot issue.

In the research section of the traditional traffic key, the main research methods are 1) for research of traffic network topology; 2) according to the analysis of network traffic statistical data; 3) at the micro level, to study the characteristics of road. Each of these methods has its limitations. Method 1) only for the topology of the network, do not consider the exchange of information between nodes; 2) method based on statistical methods, this method can only static analysis of key sections of road network, to analysis key sections changes with time / network flow method; 3) the micro field characteristics of main traffic, without considering the macro changes the road network. Here, we propose a target model based on macro freeway

network.

This paper from the perspective of practical application of intelligent transportation, the existing key sections of mining method, put forward two research methods: 1) put forward a mining model of key sections based on the greedy algorithm, and prove the feasibility of the greedy algorithm; 2) put forward a mining model of key sections of complicated network community based on partitioning, dynamic mining key sections.

Integrating the above results, a freeway critical link mining system based on B/S architecture is implemented, and the reliability and applicability of the prototype system are preliminarily verified in real application scenarios.

**KEYWORDS:** Complex Network, Intelligent Transportation , Key Road, Community Partition

## 目录

<b>第一章 绪论</b>	<b>1</b>
1.1 研究背景 . . . . .	1
1.2 研究内容 . . . . .	4
1.3 研究意义 . . . . .	5
1.4 论文结构 . . . . .	6
<b>第二章 相关研究</b>	<b>7</b>
2.1 高速公路关键路段挖掘相关研究 . . . . .	7
2.2 复杂网络关键路段挖掘相关研究 . . . . .	10
2.2.1 基于特征向量的排序方法 . . . . .	10
2.2.2 基于路段移除和收缩的排序方法 . . . . .	11
2.2.3 路段重要性排序方法的评价标准 . . . . .	12
2.3 复杂网络社群划分相关研究 . . . . .	13
2.3.1 基于划分的社群挖掘 . . . . .	14
2.3.2 基于模块性优化的社群挖掘 . . . . .	14
2.3.3 基于标签传播的社群挖掘 . . . . .	15
2.3.4 基于动力学的社群挖掘 . . . . .	15
2.3.5 基于仿生计算的社群挖掘 . . . . .	16
2.4 本章小结 . . . . .	16
<b>第三章 基于贪心的高速公路关键路段识别模型</b>	<b>17</b>
3.1 引言 . . . . .	17
3.2 问题定义 . . . . .	18
3.2.1 问题定义 . . . . .	18
3.2.2 模型定义 . . . . .	19

3.2.3 次模性分析 . . . . .	20
3.2.4 基于贪心法的关键路段求解 . . . . .	22
3.3 实验及结果 . . . . .	25
3.3.1 实验数据 . . . . .	25
3.3.2 实验结果 . . . . .	26
3.3.3 时间分析 . . . . .	28
3.4 本章小结 . . . . .	29
<b>第四章 基于社群划分的关键路段识别方法</b>	<b>31</b>
4.1 引言 . . . . .	31
4.2 问题定义 . . . . .	32
4.2.1 问题定义 . . . . .	32
4.2.2 模型定义 . . . . .	35
4.2.3 基于社群划分的关键路段求解 . . . . .	37
4.3 实验及结果 . . . . .	39
4.4 本章小结 . . . . .	45
<b>第五章 原型系统的设计与实现</b>	<b>47</b>
5.1 系统功能 . . . . .	47
5.2 系统架构 . . . . .	47
5.3 界面功能展示 . . . . .	48
5.4 本章小结 . . . . .	49
<b>总结与展望</b>	<b>51</b>
5.4.1 主要工作 . . . . .	51
5.4.2 未来工作展望 . . . . .	52
<b>参考文献</b>	<b>53</b>
<b>附录 A 附件</b>	<b>57</b>





## 插图

1.1 中国高速公路网络示意图 . . . . .	1
1.2 国家科技支撑计划 . . . . .	4
2.1 基于统计分析的关键路段识别模型 . . . . .	8
3.1 关键路段挖掘：以 1h 为区间 . . . . .	27
3.2 关键路段挖掘：以 1d 为区间 . . . . .	27
3.3 不同方法求得的关键路段结果图 . . . . .	28
3.4 关键路段挖掘：03: 00 . . . . .	28
3.5 关键路段挖掘：09:00 . . . . .	28
4.1 高速公路车辆跳数分布图 . . . . .	33
4.2 高速公路社群特性 . . . . .	33
4.3 基于模块化函数的社群划分方法 . . . . .	40
4.4 基于模块化函数的社群划分方法 . . . . .	41
4.5 结合物理路网特性的社群挖掘 . . . . .	41
4.6 结合路段距离的变权社群挖掘结果 . . . . .	42
4.7 对比实验：以 1h 为区间 . . . . .	43
4.8 对比实验：以 1d 为区间 . . . . .	43
4.9 枚举求得关键路段 . . . . .	43
4.10 贪心求得关键路段 . . . . .	43
4.11 基于社群划分的关键路段 . . . . .	43
4.12 基于统计学的关键路段 . . . . .	43
5.1 智能高速系统架构 . . . . .	48
5.2 逻辑流程图 . . . . .	49

5.3 系统分群结果图 . . . . .	49
5.4 系统路段选取结果图 . . . . .	49

## 表格

3.1 算法结果集 . . . . .	27
3.2 不同算法的运行时间 . . . . .	29
4.1 不同社群划分方法效果对比 . . . . .	42
4.2 路网通行效率提升量误差分析 . . . . .	44
4.3 关键路段选取误差分析 . . . . .	44
4.4 不同方法运行效率分析 . . . . .	44



# 第一章 绪论

## 1.1 研究背景

高速公路是支撑国家经济发展、服务群众生活、保障国家安全的战略资源和设施。截止 2016 年底，我国公路通车总里程达到 457 万公里，其中高速公路 12 万公里，2017 年将新增 4500 公里，是世界上规模最大的高速公路（如图1.1所示）。



图 1.1 中国高速公路网络示意图

随着中国经济的快速发展，人们生活水平的不断提高，居民的出行和货物运输的数量也在逐渐增加。交通出行是人类活动不可缺少的一部分。据估计，每天平均有 40% 的人口在路上花费至少 1 小时。近几年来，人们变得越来越依赖于交通系统，对于交通管理人员来说，机遇和挑战共存。首先，交通拥堵已成为一个日益严重的问题。全球范围内的道路上的车辆增加，根据调查，截止至 2016 年初，北京共有 544 万辆车，比 2014 年初增加了 50 万辆。这些激增的车辆会对道路系统产生严重的压力，极大的增加拥堵以及拥堵后的损耗。拥堵会导

致燃油消耗增加，空气污染，以及实施公共交通计划的困难。车流流量过多时，交通事故风险与交通运输系统中的膨胀增加，交通事故之后的恢复时间与恢复代价也会急剧增加。在中国，2009年的交通事故死亡人数约有7万人，在2015年达到9万人。美国联邦公路管理局公布的报告显示，发生在城市的交通事故约占所有拥堵延误的50% - 60%。一个国家的技术竞争力、经济实力和生产能力，在很大程度上取决于其交通系统性能。交通管理资源有限，中国高速公路正在逐渐走向免费通行，难以对交通系统进行全面建设；同时，国家运输系统的有效性也依赖于一个国家的处理紧急情况的能力，如交通事故发生以后的大规模疏散。随着高速公路的不断发展，各类维护高速公路的需求也都被一一提出：

1) 人员分配问题。最典型的是今年五一，交通部联合路政大队、高速交警大队、项目部制定组有针对性的应急预案，根据历史的交通信息，负责做好恶劣天气、旅游高峰车流量饱和、突发事件引起交通阻塞的应急处理。安排值班人员，落实机械设备和应急物资准备，一旦发生突发事件，迅速启动，切实做好节日期的保畅工作。

2) 安全管理。面对节日期间可能出现的状况，高速公路养护所需要在节日期前组织开展道路安全隐患检查活动：一是对管段路段进行安全隐患排查，发现问题立即落实整改；二是加强春季防火工作的管理，及时清理桥涵构造物下的易燃物品，对边坡、中央分隔带进行打草、苗木修剪，对匝道圈进行专人看护，安排养护员负责所辖路段的防火报警工作。

3) 基础设施建设。交通的基础设施建设主要包括轨道交通设施、停车场设施建设等。交通基础设施包括为交通系统保障安全正常运营而建设的公路、轨道、隧道、高架道路、车站、通风亭、机电设备、供电系统、通信信号、道路标线等设施。

4) 高新设施建设。目前越来越多的新技术涌现，高新技术层出

不穷 [1]。这些新技术对高速公路稳定性的提高具有重要意义，但是，新技术普遍造价高昂，无法直接大面积使用，需要交通管理者进行合理分配。

上述的交通需求中有一个共性，那就是找出高速公路网络中最重要的路段。针对高速公路中的重要路段，交通管理者可以进行针对性管理。高速公路的变化日新月异，传统的研究方法很难准确的预估关键路段，亟需提出一套适用于大规模路网的高速公路的关键路段挖掘识别系统。在准确识别高速公路关键路段的前提下，我们可以进行有效的人员分配，避免资源浪费；可以针对高危路段进行安全管理，减少事故风险；可以针对核心路段进行基础设施建设，提升高速公路稳定性，节省预算；可以研究道路施工问题，合理分配施工路段，将施工造成的影响减到最小。高速公路关键路段挖掘问题是交通研究领域重要的基础问题。

本文提出面向高速公路的关键路段挖掘模型，以高速公路整体运行效率为评价标准，结合高速公路的网络特性，给出贪心算法的可行性证明，提出基于贪心算法的高速公路关键路段挖掘模型；在贪心算法基础上，提出一种优化策略，建立基于复杂网络社群划分的关键路段挖掘模型，优化算法效率，在静态关键路段挖掘的基础上，实现关键路段动态挖掘。

本文针对高速公路进行关键路段挖掘，主要基于高速公路的收费数据，并得到如下项目的支持：

### 1) 山西/安徽省智能交通系统

实验室开发过面向山西/安徽省的智能交通系统，在这个系统里，我们开发了比较完善的底层数据存取框架，采用 Mysql 数据库结合 Infobright 数据仓库的数据存储格式，Linq To Sql 数据存取方式，集成数据存取逻辑接口，为本研究提供数据基础。

### 2) 国家科技支撑计划：高速公路网运行状态智能检测与安全服

## 务保障关键技术研发及系统集成

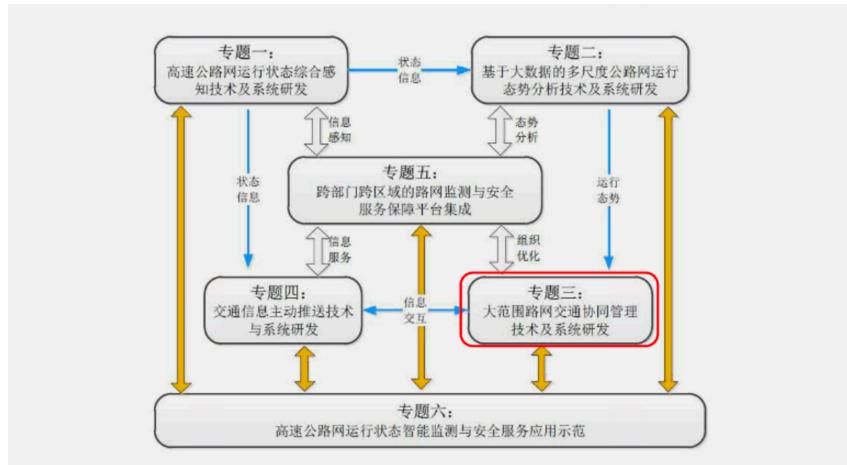


图 1.2 国家科技支撑计划

图1.2为项目专题间关系图，我们主要负责专题三：大范围路网交通协同管理技术以及系统研发。专题一主要研究高速公路的网络监测设施的部署问题，这些设施将用于改善路网运行状态以及获取路网运行数据；专题二主要用于分析和处理这些数据，将这些数据转化为流量、速度、密度和 O-D 信息。专题三负责提出决策模型，基于前面专题获取的数据，输出目前路网状态下最合理的绕行路线和限流建议；专题四负责研究如何设置设施，将这些建议分发出去；专题五和专题六负责将整个系统进行集成。可以看出，本课题中的专题一、专题三、专题四都和高速公路重要路段挖掘方法有关。专题一中的监测设施主要部署在重要路段中，优化监测效果；专题三中的决策模型可以基于路段的重要程度，策划导流决策；专题四中可以根据重要路段的分布信息，划定信息发布点，优化信息发布行为。

## 1.2 研究内容

本文从交通领域的诸多实际问题出发，研究高速公路关键路段挖掘问题。针对现有复杂网络关键路段挖掘技术的不足，提出以下两个研究思路：

1) 提出一种基于贪心算法的关键路段挖掘模型，并给出次模性证明。在高速路网上建设基础设施、部署警力，增加维护成本，都可以归类于对高速公路中路段的资源投放问题。本文提出关键路段挖掘模型，以每条路段都具有一定的损毁概率为基础，定义当路段获得投资后，损毁概率下降；结合路段损毁概率矩阵，计算用户的出行效率，最终获得高速公路的通行效率。文中分析了目标模型的性质，证明了模型的次模性，并给出了贪心求解步骤。

2) 提出一种基于复杂网络社群划分算法的关键路段挖掘模型。本文提出的关键路段挖掘模型，本质上可归类于概率规划问题，时间复杂度达指数级。在小范围路网中，模型计算时间较少，可以实现静态路网中关键路段的挖掘。但是在大范围路网中，算法的时间消耗会急剧增长，而且算法对于需要实时动态分析关键路段的需求无法满足。本文同时分析了以往复杂网络社群划分的局限性，给出一种结合高速公路的网络特性的社群划分方法。根据高速公路路段特性，结合模拟退火和变权算法，解决传统社群划分方法中的低分辨率问题 [2] 和极端退化问题 [2]，结合投资问题，在有限的误差下快速求解关键路段。

### 1.3 研究意义

研究意义主要包括两个方面：

1) 应用实践价值：从应用的角度来看，随着我国高速公路网络规模的逐渐扩大，路网结构日益复杂，人们的出行需求逐年增加，高速公路遇到的挑战不断增加，高速公路的稳定越来越重要。高速公路监测设施不断完善，可以实时监测高速公路中的车流信息。本文结合高速公路的特性，解决传统关键路段挖掘方法中的不足，优化时间复杂度，使得模型可以在可承受时间复杂度内求解，最终实现静态、动态挖掘路网的关键路段。

2) 理论研究价值：本文提出的“面向高速公路的关键路段挖掘模

型”具有一定的理论创新性。将一些高速公路建设上的具体问题抽象成一类基础问题，并给出一个可行解。传统高速公路关键路段研究，大多是研究高速公路的统计特性（如流量的大小，路段周围城市的重要程度，路段周围环境的变化等），本文利用数学模型描述各类高速公路现象，从宏观角度给出并求解一种具有普适性的高速公路关键路段挖掘模型。

## 1.4 论文结构

第一章为绪论，介绍了本文的研究背景，提出了本文的研究内容。第二章介绍了相关研究，主要介绍了高速公路关键路段研究、复杂网络关键路段研究的相关工作，结合交通问题的特点分析了现有方法的优势与不足；同时对复杂网络社群划分方法及其相关研究进行了介绍，通过对现有社群划分方法的分类对比，分析了它们的优势与不足；第三章论述本文的主要研究内容，提出了一种复杂网络关键路段挖掘模型，分析模型的贪心可解性，并给出了贪心解法与实验。第四章提出了一种基于复杂网络社群划分的关键路段挖掘方法，给出了高效的优化算法和详尽的理论分析，并在多个数据集下的进行了验证。第五章给出了本文实现的一个原型系统，并且已经在实际数据上运行。第六章给出了全文的总结与未来工作展望。

## 第二章 相关研究

### 2.1 高速公路关键路段挖掘相关研究

高速公路的关键路段挖掘问题研究较少，主要分为基于统计的研究方法和基于路网拓扑结构的研究方法。

2006 年，Jenelius 等人提出了一种基于链接重要指数和现场曝光指数的路段重要性排序方法 [3]，他们认为高速公路中的关键路段是结合路段发生危险概率和路段脆弱性得到的。路段脆弱性有很多定义，包括①对一般事故的敏感度，即受事故影响的大小；②对微小事故的敏感度，即当网络压力足够大时，哪些路段在发生小型事故后依然会对路网造成严重影响；③可靠性，研究当处于严峻条件下路网的可操作性。同时脆弱性的使用范围也不同，有①一条路段的脆弱性；②一小块网络的脆弱性；③整体网络的脆弱性。作者认为路网的脆弱性需要结合路段损毁概率和路段损毁后所造成的后果综合考虑，给出了判断路段重要性的目标函数：

$$L(k) = \frac{\sum_i \sum_{j \neq i} w_{ij} (c_{ij}^{(k)} - c_{ij}^{(0)})}{\sum_i \sum_{j \neq i} w_{ij}} + \frac{\sum_{k \in E^{nc}} \sum_{i \in V_m^d} \sum_{j \neq i} w_{ij} (c_{ij}^{(k)} - c_{ij}^{(0)})}{L^{nc} \sum_{i \in V_m^d} \sum_{j \neq i} w_{ij}}$$

式中  $c_{ij}^{(k)}$  表示当路段或子路网集合  $k$  断裂时，交通网络的通行代价， $c_{ij}^{(0)}$  是初始通行代价。 $w_{ij}$  表示  $k$  的重要性。该研究虽然可以度量高速公路中关键路段的重要性，但是函数局限于研究一条路段或者一个完整区域，没有考虑到多个路段、区域之间的相互影响。同时该文没有对算法效率进行优化，时间复杂度较高。

2002 年，Karlaftis 等人提出了一种结合道路几何特征与事故发生

概率的交通事故概率预测模型 [4]。该研究从微观角度出发，提出了一种定量评估公路几何特征对事故发生率的影响的方法，之后给出一种基于 HTBR 的道路事故率预测方法。该方法解决了 NB 回归的一些缺陷。2016 年 8 月，Kerner 提出了一种基于微观道路信息的关键路段挖掘模型 [5]，该模型结合道路的集合形状，考虑驾驶员的视线等因素，利用路段安全性度量函数来挖掘关键路段。这些方法都是从微观角度出发，没有体现整体的网络特性。

2016 年，Yip 等人基于高速公路统计学方法 [6] 研究高速公路路段的重要程度。该文章从高速公路拥堵情况出发，讲述了高速公路关键路段挖掘的意义，并且基于弗吉尼亚的交通管控系统，文章中使用 jackson network 来进行路段事故概率预测，流程如图??。通过 VDOT 发布的 SSP 数据，结合路段长度、平均行驶速度等数据，获取路段的事故概率，构建概率模型，构建 jackson network 预测模型来预测关键路段。

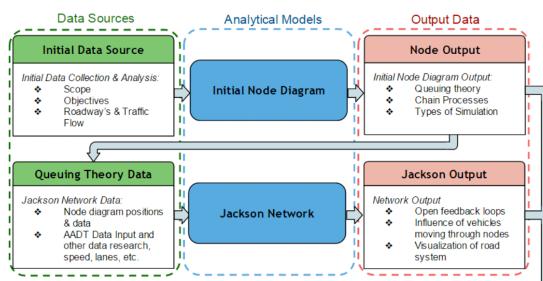


图 2.1 基于统计分析的关键路段识别模型

2017 年，Yacine 结合路段滑坡敏感性，给出了君士坦丁公路中路段的敏感性挖掘方法 [7]，该方法结合路段的历史信息，通过解读航拍信息，遥感图像和实地调查，定义并研究了高速公路路段的滑坡性质。同时使用地理信息系统（GIS）为每个地理环境因素生成主题层图，从地质数据库中提取岩性单位和断层图的距离，结合数字高程模型（DEM）计算的斜坡坡度，坡度和距离，在 GIS 环境中对滑坡相关因素与滑坡事件之间的关系进行分析，获取易发生滑坡的路段。2014

年, Ren 等人基于节点的重要程度进行链接位置的优化, 他们用节点的地理指标和集群特性来判定节点的重要程度, 研究如何优化路段的连接方式和连接位置。这些方法从微观角度出发, 通过研究节点和路段的特性来计算重要性, 忽略了宏观交通流对整体网络的影响。

2015 年, Niu 等人通过节点重要性评估方法研究关键节点 [8], 他们提出了一种基于图论和网络理论评估公路网节点重要性的新方法, 首先为每个具有 GDP, 里程, 容量等权重的节点提出了一个效益函数, 并计算出一个效益指标, 表示每个节点的本地重要性; 其次, 为了描述整个网络的全球重要性, 对节点之间的距离估计进行加权。通过地方和全球重要性的加权平均来计算每个节点的整体重要性。2011 年, Song 等人基于因子分析法, 结合 k-means 算法对每个节点的重要程度进行进一步的划分 [9]。2011 年, Wang 等人基于节点的度、介数和交通量进行关键路段研究, 同时结合 C-means 聚类方法进行进一步排序 [10]。2013 年, 他们又根据节点删除法来进行关键节点的识别 [11], 通过研究高速公路的级联失效性质, 提出了基于节点失效、考虑连锁故障的交通网络节点重要性评估方法, 它使用级联故障网络的拥塞状态来描述节点的重要性。2010 年, Chen 等人认为影响节点重要程度的因素有很多, 而且它们的权值固定不变, 基于这个思想, 提出了一种基于 Matlab 聚类分析的解点重要性划分方法。

2010 年, Peeta 等人提出了基于两层决策模型的关键路段挖掘方法 [12], 综合研究路段断裂后的连通性和路网通行效率, 提出了一个度量高速公路路段重要性的模型。忽略路段之间的一阶联系, 对离散函数进行求导, 最终转化为背包问题进行线性时间复杂度的求解。该方法从宏观角度综合考虑了路段对整个高速公路系统的影响, 但是该研究只针对桥梁, 并且没有分析只考虑一阶路段影响的误差。

现有的高速公路的关键路段研究较少, 且大都是基于微观角度、基于统计学特性、基于路网结构求解, 有各自的局限性。基于微观角

度的研究方法虽然可以模拟真实情况下的路网状况，但是这类研究单纯的研究细节层面的高速公路网络规律，忽视了宏观层面的整体情况。基于统计学的研究方法虽然可以简单直观的根据历史先验经验，快速总结出规律，挖掘出路网的关键路段，但是这类基于经验的方法效果无法得到保证。基于网络拓扑结构虽然是研究各种复杂网络特性的经典方法，但是高速公路网络属于稀疏网络，网络复杂程度低，路网主要受到车流流量的影响。

## 2.2 复杂网络关键路段挖掘相关研究

复杂网络中，关键节点的研究较多，关键路段的研究较少。因为传统的复杂网络（如社交网络）的边不具备实际的物理意义。近年来，复杂网络中节点重要性排序研究受到越来越广泛的关注。不同类型的网络中节点的重要性评价方法各有侧重，且应用领域极广，研究者们从不同的实际问题出发，设计出各种各样的方法。这些方法可以作为关键路段研究的参考。

在交通研究中，可以对路网进行对偶拓扑，路段变成节点，节点变为边，关键路段挖掘转化为对路网中的关键节点挖掘。复杂网络关键节点研究中，主要用节点的空间分布、平均距离、连通性、聚类系数、度相关性等参数来度量节点的重要程度。通过网络的抗毁性、传播性等数据来测试网络的稳定性和完备性。其中，基于空间分布的方法在此并不适用，路段的结构比之节点更加简单，无法体现复杂网络的空间复杂性。

### 2.2.1 基于特征向量的排序方法

基于特征向量的关键路段挖掘方法研究路段的性质对网络的影响。特征向量中心性和累计提名方法主要用于无向网络，对于高速公路并不适用。PageRank 算法和 LeaderRank 算法主要用于研究与识别

网页重要性。在此主要介绍 HITs 算法。

1999 年, Kleinberg 提出 HITs 算法 [13], 该算法认为网络中路段的重要性不能由单一指标确定。HITs 算法定义每个节点具有两个度量值:①权威值②枢纽值。权威值指路段对信息的原创性, 即由路段本身的价值; 枢纽值是指连接该路段的邻居的价值。在一个包含  $n$  个路段的网络中, 定义  $a_i^t$  和  $b_i^t$  分别路段  $e_i$  在  $t$  时刻的权威值和枢纽值, 目标函数:

$$L = \sum_{j=1}^n a_{ji} h_j^t - 1 + \sum_{j=1}^n a_{ij} a_j^t$$

式中  $\sum_{j=1}^n a_{ji} h_j^t - 1$  表示路段的权威值,  $\sum_{j=1}^n a_{ij} a_j^t$  表示路段的枢纽值, 可以看出这两个值之间相互影响。在高速公路领域。权威值是直接与路段相连的收费站的流量, 枢纽值则是相邻节点向路段输入的流量。HITs 方法可以结合路段特性与路段之间的相互影响, 整体度量路段重要性。

### 2.2.2 基于路段移除和收缩的排序方法

2004 年, 李鹏翔提出一种基于节点移除和收缩的关键节点挖掘方法 [14]。他认为路段的重要性往往体现在该路段被移除之后对网络的破坏性。从网络结构来说, 某些关键路段移除后, 整个网络有可能陷入瘫痪。李鹏翔等人用路段集被删除后, 新路网中所有不直接相连的节点对之间的最短距离来度量路段的重要程度。用  $C_i$  ( $k=1, 2, \dots, u$ ) 表示路段被删除后, 网络分化成的  $u$  个连通子图中第  $i$  个子图的节点数。易知此时不再连通的节点的数目为  $\sum_{i=1}^u \sum_{r=i+1}^u N_i N_r$ , 定义路段的重要性等价于所有不再联通的节点之间最短距离之和:

$$DSP(i) = \sum_{(j,k) \in E} d_{jk}$$

式中,  $d_{jk}$  为删除路段之前节点 j 与节点 k 之间的最短距离。该方法在衡量某些路段集的重要性方面优势比较突出。但是, 在大规模网络中, 仅删除一个或几个路段时, 网络的连通性没有遭到严重破坏, 该方法的效果不明显。

2004 年, 陈勇等人提出了节点删除的生成树法 [15]。他们认为, 一个节点或者路段删除后, 网络的生成树的数目越少, 路段越重要。2006 年, Dangalchev 对接近中心性进行了改进, 提出了残余接近中心性方法 [16], 该方法对接近中心性进行了改进, 提升短路径的影响力, 易于计算和扩展。作者认为, 路段的删除使得网络变得更加脆弱, 通过度量脆弱程度来挖掘网络中的关键路段。定义当删除路段 i 之后, 网络残余接近中心性为

$$RCC(i) = \sum_j \sum_{k \neq j} \frac{1}{2^{d_{jk}(-i)}}$$

其中  $d_{jk}^{(-i)}$  为删除路段 i 后节点 j 与 k 的最短距离。该方法在测度网络的脆弱性方面比其他方法优秀。

### 2.2.3 路段重要性排序方法的评价标准

主要分为①用网络的鲁棒性和脆弱性评价排序算法②用传播动力学模型评价排序算法。

网络科学的研究的早期, 网络结构简单, 规模小。对于这些小规模网络, 可以通过调查问卷等方式对每个节点的重要性进行打分, 用调查问卷的结果作为评判指标, 与其他算法结果进行比较, 分析各种算法的优劣。但是现在社会中, 网络规模迅速增长, 很难得到一个客观有效的节点重要性评价指标。目前主流的评价思路是将排序算法得出的目标作为研究对象, 考察这些对象对网络某种结构和功能的影响程度, 断排序结果是否正确。

## 网络的鲁棒性和脆弱性评价

本类方法主要考察网络中路段移除后，对网络结构和功能的影响。影响越大路段越重要。用  $k(i/n)$  表示移除  $i/n$  比例的路段后，网络中属于巨片 [17] 节点数目的比例。网络鲁棒性指标如下 [18]:

$$R = \frac{1}{n} \sum_{i=1}^n k(i/n)$$

文献 [19] 研究了在网络中使用几种排序方法，选出关键路段，并且移除路段后对网络最大连通集的影响。

## 用传播动力学模型评价排序算法

复杂网络传播研究范围非常广泛 [20]，如①通信网络中的病毒传播 [21]②现实世界的病菌传播③电力网络中的相继故障④经济网络中的危机扩散 [22]⑤社会网络中的信息传播 [23] 等。所以，传染病模型也是在评价各种节点重要性挖掘方法的重要模型。传染病模型主要包括 SIS 模型 [24] 和 SIR 模型 [25]。在第一个模型中，节点的传播能力被定义为当节点处于稳态下时，该节点被感染的概率；在 SIR 模型中，节点的传播能力与该节点的传播范围有关。高速公路是一个物理路网，路段事故的传播具有时延性，事故的传播范围和传染概率对研究路段的重要性意义不大。

## 2.3 复杂网络社群划分相关研究

现实中，许多情景都可以归类于复杂网络，如：人际关系网络、流行病传播网络、电子社交网络、电话网络、因特网和万维网等等。2002 年 Girven 和 Newman 引提出社群挖掘的概念。在社会网中，社群代表了相似的人群；在生物网络中，社群表示具有相似功能的生物组织；万维网络中，社群代表具有很多相似性的文档类。近几年来，复杂网

络一直属于热门研究领域，研究者分别采用自物理学、数学和计算机科学领域的理论和技术，进行复杂网络社群划分研究。

### 2.3.1 基于划分的社群挖掘

2002 年，Girvan 和 Newman 提出了 GN 社群挖掘方法 [26]，该方法采用的启发式规则，认为社群间链接的边介数应远大于社群内链接的边介数。边介数是指网络中，经过这条边的任意两点间最短路径的条数。算法通过边介数识别社群间链接，进而删除社群间链接，以自顶向下的思路建立一棵层次聚类树。该算法简单实用，适合计算小规模网络，但是同样的，该算法的缺点是计算速度过慢。2003 年，Tyler 等人结合统计方法，对 GN 算法进行了一定的改进 [27]。他们采用蒙特卡洛方法，近似估算出部分链接的边介数，通过牺牲精度来提高 GN 算法的运行效率。2010 年，刘大有等人基于结构相似度来度量路段是否属于社群间链接 [28]，他们认为社群间链接的结构相似度应小于社群内链接的结构相似度，每次删除结构相似度最小的边，直到完全社群划分。该方法主要针对社交网络，对于交通网络并不适用。

### 2.3.2 基于模块性优化的社群挖掘

2004 年，Newman 和 Girvan 提出了一个模块性函数  $Q$ ，用来刻画网络社群结构化的优劣 [29]。该算法通过不断合并两个现有的社群来进行社群划分。初始，定义社群数量，每个社群仅包含一个结点，作为候选解。每轮迭代中，选择合并后使函数  $Q$  增量最大的社群对进行合并（要求增量必须大于 0）。自底向上的层次聚类，最终输出一棵层次聚类树，将使得  $Q$  最大的社群划分作为最终结果。2005 年，Guimera 和 Amaral 提出了基于模拟退火的模块性优化算法 [30]，该算法随机生成一个初始解，每轮迭代中，在当前解的基础上，随机产生一个新的候选解，由模块化函数  $Q$  判断其优劣，结合模拟退火策略中的

Metropolis 准则，决定何时停止迭代。该算法产生新候选解的随机策略有以下几种：①将结点移动到其他社群②交换不同社群的结点③分解社群④合并社群。该算法聚类质量优秀，但是运行效率低。2006 年，Newman 朝将谱图理论引入模块性优化中 [31]。通过将模块化函数转化成一个图拉普拉斯矩阵，优化社群划分效果。2008 年，Blondel 等人提出了一种新的模块性优化方法 [32]。该方法结合了局部优化与多层次聚类技术，首先局部计算社群，之后将每个社群看作一个超级节点，再次进行合并。

### 2.3.3 基于标签传播的社群挖掘

2007 年，Raghavan 等人提出了标签传播算法 [33]，该在初始化时为每个结点赋一个唯一标签，每次迭代中，每个节点更新自身标签，更新策略是参照大多数邻居的标签。当所有结点的标签都不再变化时，社群挖掘结束。2009 年，Leung 等人将标签传播算法作为分析大规模在线社会网的工具 [34]，结合实际情况，研究算法的优势和限制，进而对算法进行了修正。同年，Barber 等人将算法 LPA 等价为一个优化问题，并给出对应的目标函数 [35]，设计了带约束的标签传播算法，该算法恰好就是社群划分的模块化函数  $Q$ 。2010 年，Di 等人发现，标签传播算法得到的社区划分具有陷入局部最优解的倾向 [36]，提出一种多步层次贪婪算法，修正算法，每次可合并多个社群对。但是，在高速公路网络中，和节点的标签相比，流量更加重要，所以该类方法并不适用。

### 2.3.4 基于动力学的社群挖掘

2000 年，Dongen 等人提出了 Markov 聚类算法 [37]，该算法基于 Markov 动力学理论，通过改变和调节马尔可夫链，呈现出网络社群结构。2007 年，在符号网络社群挖掘领域，杨博等人提出了基于

Markov 随机游走模型的启发式社群挖掘算法 [38]，该算法认为，从任意社群出发，网络中的随机游走过程到达起始社群内结点的期望概率将大于到达起始社群外结点的期望概率。2008 年，Rosvall 等人提出了映射平衡算法 [39]。该方法结合最小描述长度原理，使用信息传播扩散技术来探测网络社群结构。2011 年，Morfirescu 等人将复杂网络视为一个 agent 网络 [40]，每个 agent 拥有一个信念值，这些 agent 和邻居不停的进行会话，同时依据可信邻居的状态来更新自己的信念值，信念值的更新阈值不断衰减，最终达到收敛。该方法虽然已经广泛用于社群挖掘领域，但是算法普遍时间复杂度高，不符合高速公路实时性需求。

### 2.3.5 基于仿生计算的社群挖掘

2007 年，Liu 等人基于自然中蚂蚁的个体行为，提出了一个用于探测邮件网络社群结构的蚁群聚类算法 [41]，并且在之后结合聚类目标与群体相似度之距离的计算，作出改进。2009 年，Sadi 等人采用蚁群优化技术发现网络中的团，并将这些团视为新结点而构建一个简化网络，然后通过传统社群挖掘算法来探测社群结构 [42]。2010 年，刘大有等人从仿生角度出发提出一个基于 Markov 随机游走的蚁群算法 [36]。2007 年，Tasgin 等人提出了用于社群挖掘的遗传算法 [43]。仿生网络结合了自然生物学知识，在具有客观规律的网络中十分适合。

## 2.4 本章小结

本章介绍了现有的高速公路关键路段挖掘方法、主流的复杂网络关键路段挖掘方法与复杂网络社群划分方法，并且分析先有方法的不足，无法很好的应用于高速公路关键路段挖掘。

## 第三章 基于贪心的高速公路关键路段识别模型

本章主要分四个部分。第一节讲述了关键路段挖掘的意义，分析了现有研究的不足；第二节定义了高速公路关键路段挖掘模型，并给出次模性证明；第三节针对不同的挖掘方法进行实验，从多个角度分析本文方法的效果；第四节对本章内容做了一个小结。

### 3.1 引言

对于交通运输 [44]、水利传输 [45]、能源和通信等基础设施系统，在遭遇自然灾害或者人为灾害时，会对整个系统的性能造成显著的影响，带来重大的经济损失。所以在发生事故或者自然灾害的时候，维护这些网络的完整性至关重要。

灾难管理是一个多阶段的过程，从防灾减灾和准备，着眼于长期消除或降低风险的措施，延伸到灾后响应、恢复与重构。投资基础设施系统在缓解中起着至关重要的作用活动，它可以增强链接的稳定性。但是，将所有的路段稳定性都增强到坚不可摧，在管理人员看来是十分浪费的。本章节主要研究如何在有限的资源下，找到可以最大化网络通行效率的关键路段集合。将资源布置到这个关键路段集合中，最大化提升关键路段通行效率，增加路网稳定性，实现事故前的预防，事故后的快速恢复。

本章主要研究如何对高速公路关键路段挖掘问题进行建模，并且围绕安徽、山西、北京的收费站车辆数据，进行真实数据集上的实验。

## 3.2 问题定义

### 3.2.1 问题定义

高速公路具有成网性，给定一个有向图  $G = \{V, E\}$ ，其中  $V$  代表收费站（节点）的集合； $E$  表示高速公路中路段的集合。对于经过高速公路的车辆，定义  $O$  为车辆的出发节点， $D$  作为车辆的目标节点。定义  $P_e$  ( $0 \leq P_e \leq 1$ ) 为路段的损毁率，这个概率通过历史上的高速公路路段损毁事件得到，同时可以随着交通管理者对路段进行管理而减小。定义管理者的决策向量  $y=\{y_1, y_2, \dots, y_n\}$ ， $y$  是一个  $n$  维向量，每一维  $y_i$  的数值取 0 或 1，1 表示这条路段属于关键路段集合，管理者会对其进行维护和投资管理，改善路段状况；0 表示非关键路段。基于每一条路段都有一定概率损毁，定义  $C_{e_i}$  表示第  $i$  个路段的状态。当  $C_{e_i}$  等于 1 时，路段保持完好，当  $C_{e_i}$  等于 0 时，路段因为事故损毁。定义  $c=\{C_{e_1}, C_{e_2}, \dots, C_{e_n}\}$ ， $c$  表示路网的某一种拓扑结构， $C$  表示路网的所有拓扑结构的集合。对于行驶在高速公路上的车辆，定义车辆的出行时间为  $X_i$ ，这个出行时间由车辆的路径选择、出行时途径路径的车流密度决定。定义当高速公路路段断裂严重，车辆无法抵达目的地时，车辆的出行时间定为常量  $M$ 。 $M$  的大小在一定程度上代表了路网连通性的权重， $M$  越大，高速网络的连通性就越重要。为了更好的求解目标函数，在此提出两个假设：

1) 路段之间的损毁概率相互独立：假设处于静态模型下，所有路段都有一定的概率损毁，这些损毁率之间没有相互影响。传统研究网络可靠性的相关文献 [3] 都是基于这个假设所做的研究。

2)  $M > \text{Max}(X_i)$ :  $M$  必须要大于连通路网中的最大出行代价。当车辆在路段中找不到一条可以抵达终点的路径时，定义车辆此次出行的代价要绝对大于路段仍然连通情况下的任意时间。即默认断裂对路网造成的影响大于路段仍然连通的情况。

根据高速公路的历史事故数据，通过结构分析和统计调查 [23]，确定路段损毁的概率。使用该概率作为本研究的先验概率。高速公路建设管理者可以通过在高速路段上建立基础设施，投放人力资源等方式管理路段，增强路段的稳定性。假设路段  $e$  以概率  $P_e$  损毁，以概率  $(1 - P_e)$  保持完好。基于路段的损毁率，我们可以计算路网拓扑结构概率矩阵  $Z$ ：

$$\begin{matrix} C_{e_1}^1 & C_{e_2}^1 & \cdots & C_{e_n}^1 & P^1 \\ C_{e_1}^2 & C_{e_2}^2 & \cdots & C_{e_n}^2 & P^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{e_1}^m & C_{e_2}^m & \cdots & C_{e_n}^m & P^m \end{matrix}$$

矩阵中， $C_{e_i}$  表示第  $i$  条路段的状态，0 表示遭遇事故，已经损毁，1 表示完好无损；对于每一行来说，前  $n$  项  $C^j = \{C_{e_1}^j, C_{e_2}^j, \dots, C_{e_n}^j\}$  表示路网的拓扑结构，全 0 表示全部路段断裂，路网瘫痪；全 1 表示路网完全连通。第  $n+1$  项  $P^j$  表示高速公路网络拓扑变成这个拓扑结构的概率。其中  $P^j = \prod_{i=1}^n (P_{e_i} C_{e_i}^j + (1 - P_{e_i})(1 - C_{e_i}^j))$ 。在交通管理者选取关键路段，并进行一定的决策处理后，路段的损毁概率发生变化，进而概率矩阵  $Z$  也同时会发生变化。

### 3.2.2 模型定义

在此提出关键路段挖掘模型：

$$L(\mathbf{y}) = -E(T(\mathbf{c}|\mathbf{y})) \quad (3.1)$$

$$T(\mathbf{c}|\mathbf{y}) = P(K|\mathbf{c}) \sum_{k \in K} X_k \quad (3.2)$$

$\mathbf{y}$  是一个  $n$  维向量，表示关键路段集合，即管理者想要投资维护的路段。 $T(\mathbf{c}|\mathbf{y})$  表示当关键路段集合为  $\mathbf{y}$ ，路网拓扑结构为  $\mathbf{c}$  的时候，高速公路的整体通行时间。式3.1中对路网通行时间的期望取负，转化为通行效率。模型的目标是选取出关键路段，对这些关键路段增加投入，使得在同样的投入情况下，整个路网的通行效率得到最大的提升。结合式3.1，式3.2，得到展开式：

$$\text{Max}(L(\mathbf{y})) = -\text{Min}_{\mathbf{y}} \sum_{\mathbf{c} \in C} P(\mathbf{c}|\mathbf{y})P(K|\mathbf{c}) \sum_{k \in K} X_k \quad (3.3)$$

式中  $\mathbf{y}$  表示关键路段集合，假设高速公路网络的路段数量为  $n$ ，则  $\mathbf{y}$  为  $n$  维向量，对于  $\mathbf{y}$  的第  $i$  个维度，如果值为 0，则表示第  $i$  个路段不是关键路段，反之表示第  $i$  个路段是关键路段； $\mathbf{c}$  表示路网的拓扑结构， $C$  是高速公路网络所有拓扑结构的集合； $P(\mathbf{c}|\mathbf{y})$  表示当关键路段集合为  $\mathbf{y}$  时，高速路网的拓扑结构为  $\mathbf{c}$  的概率； $k$  表示第  $k$  个车辆的出行路径， $K$  表示所有车辆的出行路径集合； $P(K|\mathbf{c})$  表示当路网拓扑结构为  $\mathbf{c}$  时，高速公路车辆出行路径集合为  $K$  的概率； $X_k$  表示当车辆的行驶路径为  $k$  时，车辆的行驶时间。这个时间可以用交通动力学理论求解 [46]。

### 3.2.3 次模性分析

#### 单调性证明

定义  $\mathbf{y}$  是关键路段集合， $e_i$  代表路段  $i$ ，定义  $\mathbf{y}^1 = \mathbf{y} + e_i$ ， $\mathbf{y}^1$  所代表的关键路段集合比  $\mathbf{y}$  多出关键路段  $e_i$ 。定义  $\Delta L$ ：

$$\Delta L = L(\mathbf{y}^1) - L(\mathbf{y})$$

结合式3.2，得到：

$$\Delta L = -\left(\sum_{c \in C} P(c|\mathbf{y}^1)P(K|c)X_k - \sum_{c \in C} P(c|\mathbf{y})P(K|c)X_k\right)$$

化简，得到：

$$\Delta L = -\left(\sum_{c \in C} \Delta P(c|e_i = 1)P(K|c)X_k - \sum_{c \in C} \Delta P(c|e_i = 0)P(K|c)X_k\right)$$

式中， $\Delta P(c|e_i)$  表示在  $\mathbf{y}^1$  和  $\mathbf{y}$  两种投资方式中，路网拓扑结构概率差值的绝对值。易知当路网结构只有  $e_i$  不同时， $\Delta P(c|e_i = 0) = \Delta P(c|e_i = 1)$ 。公式化简为：

$$\Delta L = -\left(\sum_{c \in C} \Delta P(c|e_i = 1)P(K|c)(X_k^{e_i=1} - X_k^{e_i=0})\right)$$

式中， $X_k^{e_i=1} - X_k^{e_i=0} \leq 0$ 。所以  $\Delta L \geq 0$ ，函数单调性得证。

## 次模性证明

次模函数 (submodular function) 是一种具有“边际效应递减”效应的函数，即对于一个集合函数，如果  $S \subseteq V$ ，那么在  $V$  中增加一个元素所增加的收益要小于等于在  $S$  的子集中增加一个元素所增加的收益。形式化表述就是：对于函数  $f$  而言，若  $A \subseteq B \subseteq V$ ，且  $\varepsilon \in V - B$ ，则  $f(A \cup \{\varepsilon\}) - f(A) \geq f(B \cup \{\varepsilon\}) - f(B)$ ；或者若  $A \subseteq \Omega$   $B \subseteq \Omega$ ，则  $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$ ；或者对于任意  $X \subseteq \Omega$   $x_1, x_2 \in \Omega$ ，下面的式子一定成立： $f(X \cup x_1) + f(X \cup x_2) \geq f(X \cup x_1, x_2) + f(X)$ 。满足这三个条件中的任意一个，函数  $f$  即满足次模性。

假设  $\varepsilon$  是某一条路段， $\mathbf{y}$  和  $\mathbf{Y}$  都是关键路段的集合， $\Omega$  是关键路段集合的全集空间， $\mathbf{y} \subseteq \mathbf{Y} \subseteq \Omega$ 。 $\varepsilon \in \Omega - Y$ 。 $\{y + \varepsilon\}$  表示对于关键路段集合  $\mathbf{y}$ ，将  $\varepsilon$  作为新的关键路段加入，形成新的关键路段集合。

定义：

$$I = L(\mathbf{y} + \boldsymbol{\varepsilon}) - L(\mathbf{y}) - (L(\mathbf{Y} + \boldsymbol{\varepsilon}) - L(\mathbf{Y})) \quad (3.4)$$

不妨假设  $\mathbf{Y} = \mathbf{y} + \boldsymbol{\varepsilon}_2$ , 公式 3.4 转化为:  $I = L(\mathbf{y} + \boldsymbol{\varepsilon}_1) - L(\mathbf{y}) - (L(\mathbf{y} + \boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2) - L(\mathbf{y} + \boldsymbol{\varepsilon}_2))$

令  $J = L(\mathbf{y} + \boldsymbol{\varepsilon}_1) - L(\mathbf{y})$ , 要证明  $I \geq 0$ , 即证  $J$  单调非增。

$J$  属于有限离散函数, 对  $J$  进行求导化简 [3], 得到:  $\frac{dy}{dx} = \sum_{c_1|\mathbf{y}+\boldsymbol{\varepsilon}} (\sum_{c_1|\mathbf{y}+\boldsymbol{\varepsilon}} P(c_1) - \sum_{c_2|\mathbf{y}} P(c_2)) X_k$ 。显然  $\sum_{c_1|\mathbf{y}+\boldsymbol{\varepsilon}} P(c_1) * X_k$  具有单调非减性, 导数恒大于等于 0。模型的次模性得到证明

由于对于具有次模性的模型, 贪心求解的精度误差不会超过  $\frac{1}{e} * OPT$ , 所以模型可以用贪心方法近似求解。

### 3.2.4 基于贪心法的关键路段求解

#### 贪心法求解

贪心算法主要用于优化算法的复杂度, 采用逐步获取局部最优解的方式, 不断循环, 直到得到最终解集。局部最优解求解思路是: 遍历  $n$  条路段, 计算每一条路段被选为关键路段之后对高速公路通行效率的影响, 选出影响最大的那条边作为关键路段。

定义路段损毁概率矩阵  $P = \{p_1, p_2, \dots, p_n\}$ , 其中  $p_i$  代表路段  $i$  出现事故的概率。定义  $U = \{u_1, u_2, \dots, u_n\}$  为路段概率变化矩阵,  $u_i$  表示管理者对路段  $i$  采取措施后, 路段  $i$  事故率的变化量。 $Y = \{y_1, y_2, \dots, y_n\}$  表示关键路段集合, 其中  $y_i$  取值 0 或 1, 表示路段  $i$  是否属于关键路段, 1 表示属于, 0 表示不属于。定义管理者处理关键路段后, 路段的损毁概率矩阵  $P^1 = P + U^T * Y$ 。易知在贪心求解过程中,  $|Y| = 1$ , 定义贪心方法的目标函数:

$$\underset{|Y|=1}{\operatorname{Max}}(L(Y))$$

使用上式，求得第一条关键路段  $k$ ，更新  $P^1 = P + U^T * Y_{y_{k=1}}$ ， $P = P^1$ ，将  $k$  记录为关键路段，并将  $y_k$  恒置为 0，循环搜索下一条关键路段，直到关键路段数量达到预算值。伪代码如下：

---

**Algorithm 1** 贪心算法求解模型

**Require:** 高速车辆 O-D 数据，高速公路网络拓扑结构，关键路段数量，路段损毁率

**Ensure:** 高速公路关键路段集合

```

1: function GREEDY(ODMatrix G = V, E B Pe)
2:   res  $\leftarrow 0$ 
3:   Array  $\leftarrow \text{Null}$ 
4:   k  $\leftarrow 0$ 
5:   l  $\leftarrow 0$ 
6:   while len(Array)  $\leq B do
7:     for i  $\in E - \text{Array} do
8:       if L(Array.Append(i)) > k then
9:         k  $\leftarrow L(\text{Array.Append}(i))$ 
10:        l  $\leftarrow i$ 
11:      end if
12:    end for
13:    res  $\leftarrow k$ 
14:    Array  $\leftarrow \text{Array.Append}(l)$ 
15:  end while
16:  return Array
17: end function$$ 
```

---

为验证贪心算法的效果，在此引入对比方法：

算法2使用枚举方法，获取最优解。从  $n$  条路段中，利用枚举方法，枚举计算所有  $C_n^B$  种关键路段情况，计算每一种情况下的高速公路通行效率的变化，从中选取最优解。

算法3利用高速公路网络拓扑结构，抽取关键路段。算法中的  $Z(i)$  是计算路段  $i$  的中心性函数 [47]，该方法引用经典路段中心性来度量关键路段。

算法4基于统计学方法，计算路段重要程度，获取关键路段。结合式中  $f_i$  表示路段  $e$  的流量。：

**Algorithm 2** 枚举

**Require:** 高速车辆 O-D 数据, 高速公路网络拓扑结构, 关键路段数量

**Ensure:** 高速公路关键路段集合

```

1: function ENUMERATION(ODMatrix G = V, E B Pe)
2:   res  $\leftarrow 0$ 
3:   Array  $\leftarrow \text{Null}$ 
4:   k  $\leftarrow 0$ 
5:   for l  $\in \Omega$  and len(l)  $\leq B$  do
6:     if L(l)  $> k$  then
7:       k  $= L(l)$ 
8:       Array  $= l$ 
9:     end if
10:   end for
11:   return Array
12: end function
```

**Algorithm 3** 拓扑中心性

**Require:** 高速公路网络拓扑结构, 关键路段数量

**Ensure:** 高速公路关键路段集合

```

1: function ENUMERATION(ODMatrix G = V, E B)
2:   res  $\leftarrow 0$ 
3:   Array  $\leftarrow \text{Null}$ 
4:   k  $\leftarrow \text{Null}$ 
5:   for i  $\in E$  do
6:     k.Append({i, Z(i)})
7:   end for
8:   SortByValue(k)
9:   Array  $\leftarrow k[0 : B]$ 
10:  return Array
11: end function
```

## 复杂度分析

最优解法为直接枚举, 即穷举所有可能的路段组合, 一一计算当这些路段被选为关键路段时, 路网通行效率的提升量。选出对路段通行效率提升最大的关键路段集合。对关键路段的选取时间复杂度为  $O(C_n^B) = O(n^B)$ , 复杂度属指数级别, 随着路网规模  $n$  与关键路段规模  $B$  的增大而急剧增加。

贪心法选取关键路段的时间复杂度是  $O(n * B)$ 。通过  $B$  轮循环, 选取关键路段, 每次循环中, 选取节点的时间复杂度为  $n$ 。

**Algorithm 4** 统计**Require:** 高速公路网络拓扑结构, 关键路段数量, 高速公路路段损毁概率**Ensure:** 高速公路关键路段集合

```
1: function ENUMERATION( $G = V, E, B, P_e$ )
2:    $res \leftarrow 0$ 
3:    $Array \leftarrow Null$ 
4:    $k \leftarrow Null$ 
5:   for  $i \in E$  do
6:      $k.Append(\{i, f_i * P_i\})$ 
7:   end for
8:    $SortbyValue(k)$ 
9:    $Array \leftarrow k[0 : B]$ 
10:  return  $Array$ 
11: end function
```

---

### 3.3 实验及结果

本节针对各种方法在真实的交通数据集中进行实验, 通过对比已有的关键路段挖掘方法, 评估模型的效果。实验环境为: Windows Server 2008, 64GB RAM, Inter(R)Xeon(R) CPU E7-4830 2.13GHz 2.13GHz (2 处理器), 后续章节的实验均在相同的实验环境下进行。特别地, 实验中采用了两个国内高速公路网的数据: 安徽省和山西省高速公路网数据。

#### 3.3.1 实验数据

本节的实验数据来自于安徽省和山西省的高速公路路网, 其中的主要数据为高速路网中车辆的行驶 O-D 数据。路网中包含 142 个出口位置和 142 个入口位置。为了方便研究, 将车辆的 O-D 数据整合为出行 O-D 矩阵 ODMatrix:

$$\begin{matrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{matrix}$$

其中,  $a_{ij}$  表示以收费站 i 为起点 O, 以收费站 j 为终点 D 的车辆数量。

高速公路路段损毁概率通过统计历史的路段损毁次数获得。这些数据一部分数据库已有, 一部分通过新闻抓取获得。路段的损毁包括交通事故损毁, 重大自然灾害损毁, 严重堵车等。

### 3.3.2 实验结果

图3.1, 3.2给出了在不同时间区段下, 几种方法的最终结果比较。图3.1是基于 2010 年 10 月 30 日一天的实验结果, 纵坐标代表路网整体通行效率 (路网整体通行时间取负) 的绝对值, 横坐标代表一天内的不同时间段, 本实验中以 1 小时为一个时间段, 采样八个时间点  $[0 - 1, 3 - 4, 6 - 7, 9 - 10, 12 - 13, 15 - 16, 18 - 19, 21 - 22]$ 。由图3.1可以发现, 在整体上贪心算法明显优于统计算法, 同时统计算法又比直接基于高速公路拓扑结构获取关键路段有效, 原因是高速公路整体网络结构比较简单, 路网拓扑结构的某些性质体现的不明显。在不同的时间段, 高速公路的流量在不断变化, 不同方法的效果之间的差异也在变化。在高速公路车流最少的午夜, 几种方法差异达到最小, 从早上六点开始, 到流量最高的中午, 三种方法之间的差异逐渐增大, 这体现了高速公路流量对关键路段选取结果的影响, 流量越大, 关键路段维护后造成的效益越大; 流量越大, 选取关键节点的误差造成的影响就越大。图3.2是基于从 2010 年 10 月 10 日开始, 到 2010 年 10 月 16 日为止一周数据的实验, 纵坐标和图3.1一样, 表示网络整体的通行时间。纵坐标以一天为一个时间段, 从当天 0 点采样到当天 24 点, 采样七天 (从周日到下一个周六)。可以发现, 在以一整天的 O-D 矩阵为数据集进行研究时, 不同天之间的路网通行效率变化较小, 不同方法之间的差异也趋于平稳。这证明了高速公路具有一定的稳定性和规律性, 即可以通过研究历史数据, 获取静态关键路段。这些关键路

段具有普适性。

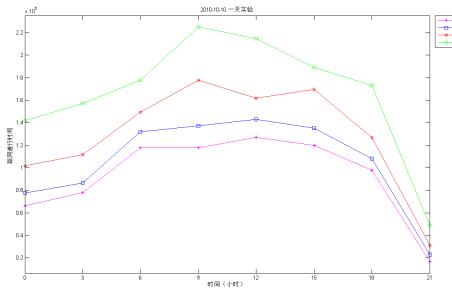


图 3.1 关键路段挖掘：以 1h 为区间

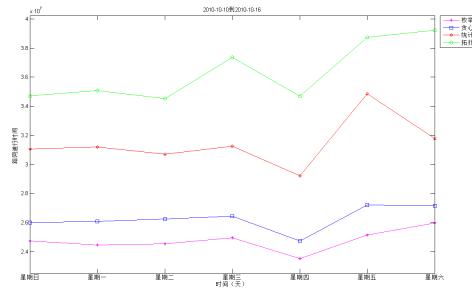


图 3.2 关键路段挖掘：以 1d 为区间

表3.1给出了不同方法求得的路网通行效率。表格中的数值是  $L(\mathbf{y})$  的绝对值（路网整体通行时间），可以看出贪心算法最接近最优解，而且误差在可接受范围内。

	枚举	贪心	统计	拓扑
一天	926030.06	1053575.26	1287439.55	1660243.55
一周	21674024.80	22989458.02	27510044.42	31790488.20

表 3.1 算法结果集

图3.3给出了关键路段在路网中的分布图，图3.3(a) 是基于贪心算法求解的关键路段集合，图3.3(b) 是基于高速公路统计方法获得的路段集合。图3.3(c) 是基于枚举所得的最优解集，图3.3(d) 是基于路网拓扑结构选取的关键路段集合。对比图3.3(a) 和图3.3(b) 可以发现，直观上重要的点（承载流量较大的路段，事故多发路段等）并不一定在路网中属于关键路段集合，关键路段集合需要经过计算才能求出；直接枚举的路段集合与贪心算法求得的路段集合十分接近，平均有 80% 以上的相似度，而统计和度中心性方法求得的关键路段和枚举方法差距较大。

在实验过程中，我们还发现当时间区间缩短到 1h 时，高速公路网络中的关键路段具有随着时间和流量变化而变化的特性。图3.4是凌晨 3 点时的高速公路关键路段集合，图3.5是早上九点时的高速公

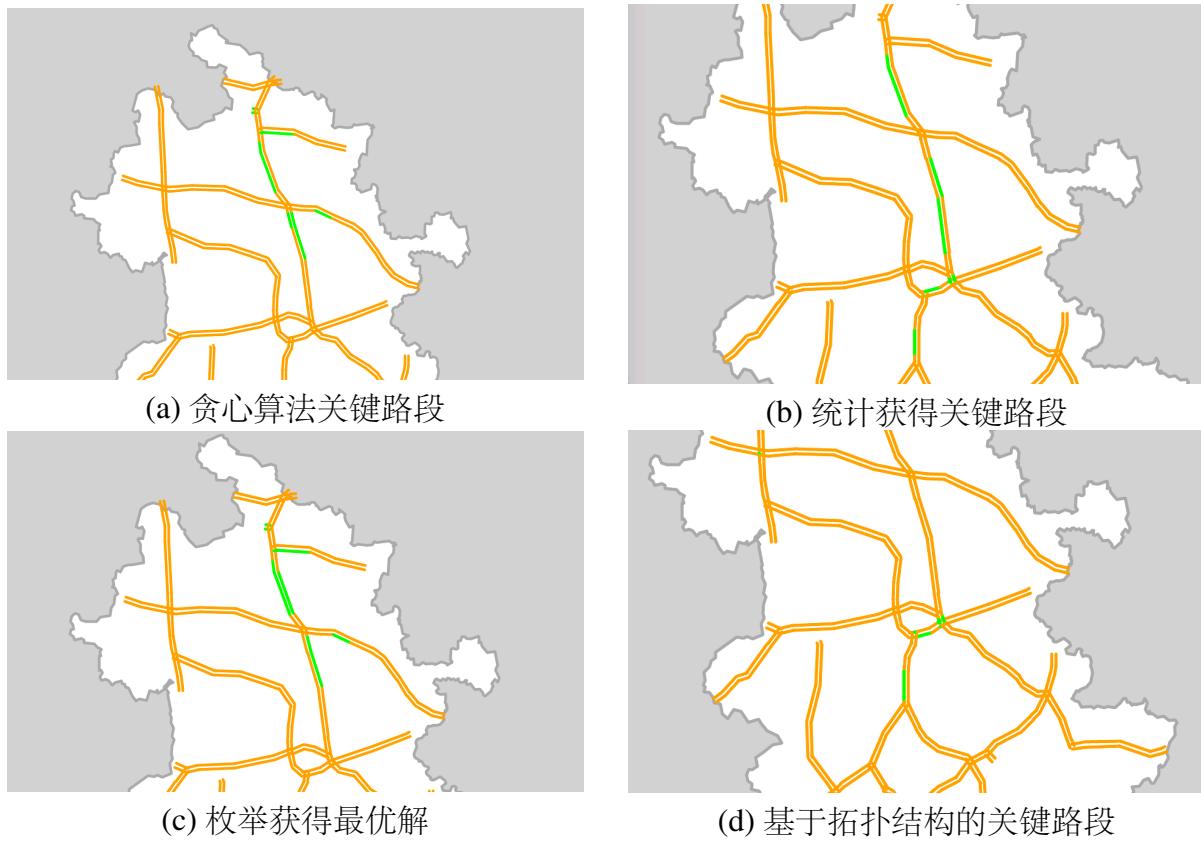


图 3.3 不同方法求得的关键路段结果图

路关键路段集合。这一发现证明了高速公路的关键路段具有动态变化特性，即高速公路的关键路段不是一成不变的，而是会随着整个路网的流量的变化而变化。

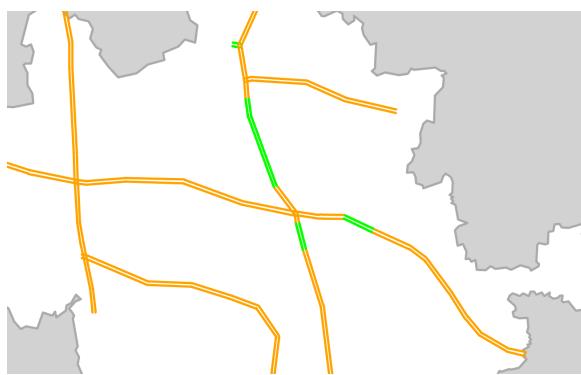


图 3.4 关键路段挖掘: 03: 00

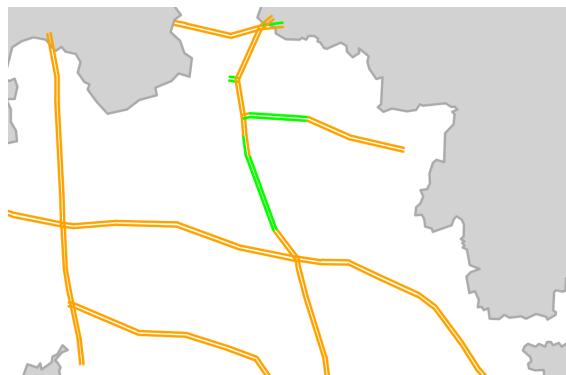


图 3.5 关键路段挖掘: 09:00

### 3.3.3 时间分析

基于暴力枚举方法的时间复杂度:  $O(n^B * 2^n)$

基于贪心算法的时间复杂度： $O(n * B * 2^n)$

基于统计路段重要性方法的时间复杂度： $O(n * \log(n))$

基于路网拓扑结构方法的时间复杂度： $O(n * \log(n))$

实验时间如表4.2所示，第一行代表实验的方法，第一列代表实验数据的范围，表格内的数值是实验的平均运行时间。

	枚举	贪心	统计	拓扑
一小时	1day	30min	1min	1min
一天	6day	2h	2min	1min
一周	7day	3h	5min	1min
一月	7day	3h	8min	1min

表 3.2 不同算法的运行时间

由表格可以看出，在以一个省为数据集的基础上，枚举方法已经处于一种较大的时间复杂度；贪心算法在一定程度上解决了算法过慢的情况，并且在精度上有一定的保证，可以应用于静态路网关键路段识别问题，但是对于动态实时应用仍旧不够；基于统计领域的路段重要性排序方法、基于路网拓扑结构的关键路段挖掘方法虽然在时间上效率较高，但是在精度上达不到要求。一周和一月的时间差距较小的原因是：对数据进行预处理，O-D 相同的用户被归为一类。在时间区段增大到一定程度时，O-D 类数量不再增加，算法运行时间增长较小。

### 3.4 本章小结

本章提出了一种面向高速公路网络的关键路段挖掘模型，同目前高速公路关键路段已有的挖掘方法相比，该方法的优势是结合高速公路的特性，考虑高速公路上的车流流量、路段事故率，从宏观角度提出一个整体的优化模型。针对上述模型，本章分析证明了模型的次模性，并给出基于贪心方法的关键路段挖掘方法。特别的，本文通过枚举方法，在较低的时间效率下计算高速公路中的最优解。结果表明该模型的贪心算法解可以很好地逼近真实解，并且在时间复杂度上有了

规模性的优化，证明了贪心算法的可行性。然而，即使贪心算法可以在一定规模上优化整体的时间复杂度，并且可以在实际应用中运行良好，但是这是基于目前的研究目标是静态关键路段挖掘，同时高速公路也只有部分路段产生过断流等重大事故的情况下达成的。当任务环境更为复杂时(扩大到全国高速公路网络)，当管理者需要更加迅速得到实时反馈的时候，上述方法受到算法计算规模的约束，无法达到预期的效果。下一章将针对高速公路的网络特性，给出相应的解决手段。

## 第四章 基于社群划分的关键路段识别方法

本章主要分四小节。第一节讲述了模型的性质，分析了贪心算法的不足；第二节给出了基于社群划分的关键路段挖掘方法模型；第三节是实验结果；第四节总结了本章内容。

### 4.1 引言

上一章介绍了面向高速公路的关键路段挖掘模型，并给出了贪心算法。然而根据模型的定义，就算进行简化，认为关键路段已经选出，计算对关键路段进行维护之后的整体网络通行效率也需要  $2^n$  的时间复杂度。贪心算法的实际时间复杂度是

$$O(B * n * 2^t)$$

式中  $B$  表示关键路段数量， $n$  表示网络中路段的数量， $t$  表示网络中可能出现损毁的路段数量。虽然在实际应用中，高速公路中路段规模不大 ( $O(10^3)$ )，大部分路段的损毁概率是 0，贪心算法对于静态高速公路中的关键路段挖掘可适用性高，但是贪心算法的时间复杂度仍旧属于指数级别，当高速公路网络规模变大后，复杂度指数上升，对于有实时性要求的动态关键路段挖掘方法并不适用。本节给出基于复杂网络社群划分的关键路段挖掘方法。

模型需要从输入的代表关键路段的离散 0-1 向量  $y$ ，求得高速公路网络通行效率的期望。这种输入为整数或整数向量，并且内部具有概率事件的问题，属于随机整数规划问题。在数学优化领域，随机规划是一个涉及不确定性优化问题的框架。比如说两阶段线性规划。决策者在第一阶段采取一些行动，之后发生随机事件影响第一阶段决策

的结果。不断调整第一阶段的决策，使得整体期望收益达到最大。

现有的随机整数规划问题大都是基于班德斯分解方法（Benders Decomposition）进行研究，然而班德斯优化方法要求有两层模型，且两层模型之间互不影响。本研究中，第一层的决策变量  $\mathbf{y}$  会直接影响到第二层里面的路网拓扑结构概率，对于这种相互依赖的随机规划问题，现有的研究没有找到适用的优化方法。

## 4.2 问题定义

### 4.2.1 问题定义

高速公路网络除具有绝大多数复杂网络的特征外，作为空间网络还具有不同于抽象网络的特性，这些特性决定了高速公路网络的拓扑性质。具体可以归纳为：高速公路交通网络的节点存在于真实物理空间，高速公路网络中的边是一种实体联接，具有明确的空间意义。高速路网中的边与抽象网络中的边不同，高速公路交通网络中节点的边权与交通距离直接相关，这一特性直接影响着高速公路网络出现小世界行为的可能性 [1]；高速公路交通网络中单一节点所能联接的边的数目受到物理空间的限制，这种限制会影响到网络的复杂程度。

在高速公路项目研究中，我们发现低跳数的用户占大多数。如图 4.1，可以发现在高速公路中，低跳数的车辆占了大多数，10 跳以下的车辆占所有车辆总数的 90% 以上。再结合高速公路的异质性，复杂网络的社群性，我们认为高速公路网络应该也具备社群性质，即存在一个个社群，这些社群各自包含一些收费站和高速公路路段，高速公路中的车辆大都从社群内部的节点出发，在同一个社群的另一个节点驶离。社群之间的车辆交流尽量小。

为此，抽取某一天的高速公路 O-D 数据，将有 O-D 交流的收费站之间连线，流量越多，线的颜色越深，流量越少，线的颜色越浅。

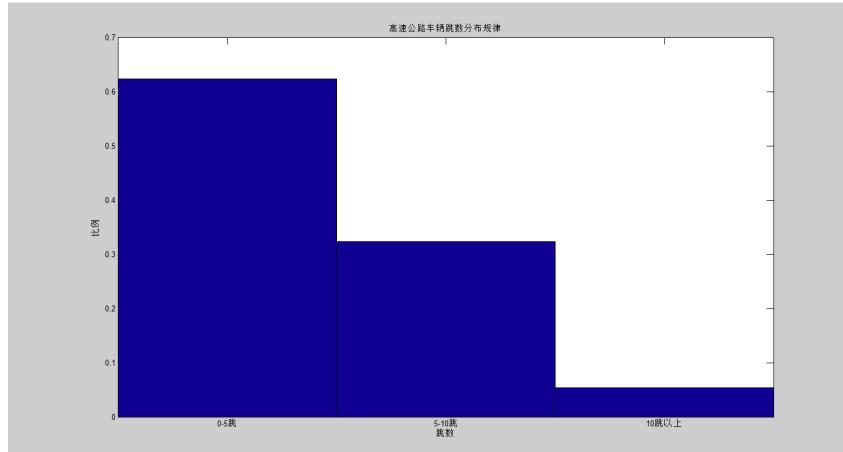


图 4.1 高速公路车辆跳数分布图

如图4.2，可以较直观的看出高速公路的社群特性。

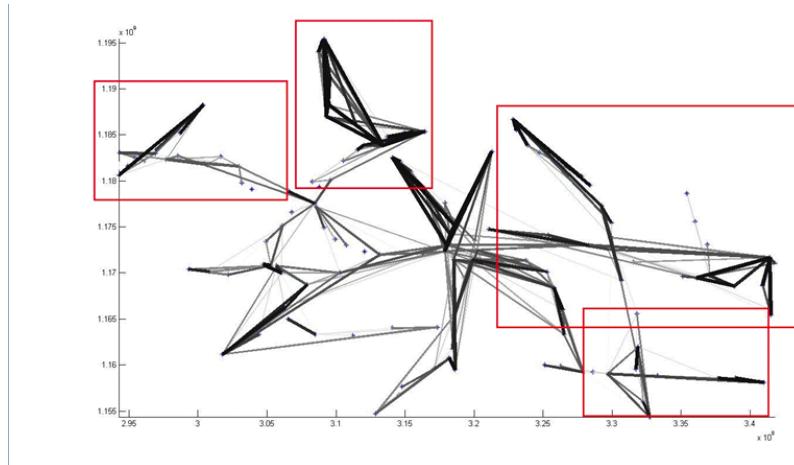


图 4.2 高速公路社群特性

高速公路社群划分的目的是将整个高速公路拓扑结构分成一个个社群，使得社群内部交流尽量多，社群之间的交流尽量少，最终在各自社群分别计算关键路段，分治计算，最后进行合并，达到优化算法效率的目的。在此引入基于模块性优化的社群划分方法。

复杂网络具有社群特性，高速公路属于复杂网络的一种。给定高速公路有向图  $G = \{V, E\}$ ，其中  $V$  代表收费站（节点）的集合； $E$  表示边的集合。定义社群  $c = \{v_1, v_2, \dots, v_m\}$ ，其中  $v_i$  是网络中的节点，即收费站或者交叉路口；社群集合  $C = \{c_1, c_2, \dots, c_u\}$ ；其中  $v_i \in V$ ，

$$c_i \cap c_j = \emptyset, \sum_{i=1}^u \sum_{v \in c_i} v = V.$$

基于高速公路社群划分的关键路段挖掘算法主要采用分治思想，将一个难以直接解决的复杂问题，分割成一些规模较小的问题，逐个计算，分而治之。本文主要将路网分成一个个子路网，在子路网中分别计算关键路段，之后再用一定方法进行合并。在此需要解决两个问题：

- 1) 如何分群：定制分群策略，使得不同社群之间的相互影响尽量小。
- 2) 如何合并：定制合并策略，将各个社群内部的计算结果合并到整个路网。

传统的复杂网络社群划分方法中，大都是针对虚拟网络（如社交网络）进行研究。高速公路网络和虚拟网络有很大的不同。在虚拟网络中，两个点之间只要有交流，那就代表有边相连；在高速公路中，我们认为只要两个收费站有流量交流，即 O-D 不为 0，那么这两个收费站之间就有边连接（不同于传统的路网定义）。但是这个边和其他的复杂网络如社交网络不同，社交网络中两个节点之间的空间距离定义为 1 跳，但是对于物理网络来说，两个节点之间的边具有实体距离。高速公路中路段之间的影响也会根据物理距离的变化而变化，这些都是传统方法中没有考虑到的。

2004 年，Newman 和 Girvan[44] 提出了一个用于刻画网络社群结构优劣的量化标准，被称作模块化函数。简单的带权模块化函数定义如下：

$$Q = \frac{1}{2m} \sum_{ij} [A_{ij} - \frac{k_i k_j}{2m}] \delta(c_i, c_j) \quad (4.1)$$

式 4.1 中， $A_{ij}$  表示节点  $i$  和节点  $j$  之间的边权； $k_i = \sum_j A_{ij}$  表示

所有与节点  $i$  相连的边的边权和;  $c_i$  是指  $i$  所属的社群编号; 如果  $c_i = c_j$ , 那么  $\delta(u, v) = 1$ , 否则等于 0;  $m = \frac{1}{2} \sum_{ij} A_{ij}$ 。

模块化函数主要用于度量社群划分结构的优劣, 现有的基于模块化函数的分群算法不适用于高速公路这种物理网络, 并且在高速公路网络上出现了低分辨率特性和极端退化特性 [48]。

#### 4.2.2 模型定义

针对现有研究的不足, 结合高速公路的路网特性, 提出一种新的面向高速公路的社群划分模型。首先定义模块化函数  $Q$ :

$$\Delta Q = \left[ \frac{\sum_{in} C + 2k_{i,in}}{2m} - \left( \frac{\sum_{tot} C + k_i}{2m} \right)^2 \right] - \left[ \frac{\sum_{in} C}{2m} - \left( \frac{\sum_{tot} C}{2m} \right)^2 - \left( \frac{k_i}{2m} \right)^2 \right] - L(i) \quad (4.2)$$

公式4.2用于判断: 当节点  $i$  从一个社群转移到另一个社群的时候, 整体路网社群模块性的变化。节点的社群归属由  $Q$  变化的大小决定。式中,  $\sum_{in} C$  表示社群  $C$  内部的所有边的权重和;  $\sum_{tot} C$  表示所有与社群  $C$  中的节点相连的边的权重和;  $k_{i,in}$  表示  $i$  到  $C$  中所有节点之间的连线的权重和;  $k_i$  表示所有和节点  $i$  直接相连的边的权重和;  $m$  是路网中所有边的权重之和;  $L(i)$  是模型罚项, 代表  $i$  转移社群后, 不同社群之间交通流的变化。定义  $L(i)$ :

$$L(i) = \frac{k_{i,c_1} - k_{i,c_2}}{k_{c_1,c_2}} \quad (4.3)$$

式4.3中,  $k_{i,c_1}$  表示路段  $i$  流向社群  $c_1$  的流量,  $k_{i,c_2}$  代表路段  $i$  流向社群  $c_2$  的流量,  $k_{c_1,c_2}$  表示社群  $c_1, c_2$  中所有节点之间的流量和。

本文提出的模型中, 边的权重不止与两个节点之间的流量有关, 还与两个节点之间的物理距离有关。和传统复杂网络不同, 节点之间的距离不再由节点之间的最短跳数决定, 而是由节点之间的最短物理

距离  $L$  决定：

$$L_{ij} = \sum_{e \in E_{ij}} e \quad (4.4)$$

式4.4中， $E_{ij}$  是节点 i 和节点 j 之间的最短路径中路段的集合。定义边权重：

$$W_{ij} = \frac{f_{ij}}{L_{ij} * T} \quad (4.5)$$

为了解决传统社群划分中的低分辨率问题，本文中的社群划分方法也采用自底向上的聚类思想，首先定义每一个节点都是一个社群，社群集合  $C = \{c_1, c_2, \dots, c_u\}$ ，路段集合  $E = \{e_1, e_2, \dots, e_n\}$ 。初始情况下， $u = n$ 。在每次迭代过程中，遍历路段  $E$ ，利用模块化函数  $\Delta Q$ ，判断节点  $e_i$  所属的社群：

$$\underset{c_j \in C}{Max} (\Delta Q)$$

式中， $\Delta Q$  表示当路段  $e_i$  由原来的社群划分到社群  $c_j$  时，模块化函数的改变量。路段  $|E|$  经过一遍遍历后，形成一个新的社群集合  $C^1 = \{c_1^1, c_2^1, \dots, c_u^1\}$ ，结合社群集合和路段集合，进行下一轮遍历，直到收敛。

传统方法中，社群划分具有极端退化特性，即最终结果无法收敛到某一个确定的最优解，而是会收敛到一个解集合中，在一定规模的解集中循环。比如说，假设社群划分已经收敛，在第  $i$  次迭代过程中，得到了社群集合  $C^i$ 。 $C^{i+1} \neq C^i$ ，但是在经过  $k$  次迭代后， $C^{i+k} = C^i$ 。实验表明，在高速公路上进行这样的社群划分，最终结果具有空间交叉特性，即不同的社群之间存在物理空间上的交叉（如图4.3），这种情况下不同社群之间的相互影响比较大。为了解决社群划分的空间交叉与极端退化问题，我们采用多变权值的思想，初始权值为  $W_{ij} = \frac{f_{ij}}{L_{ij} * 2}$ ，目标模型收敛后，记录本次迭代解集的规模 k，变

化权值  $W_{ij} = \frac{f_{ij}}{L_{ij} * (2 - 0.1 * l)}$  ( $l$  为迭代次数)，通过加大路段距离的权重，不断减少解集规模，改善社群之间空间交叉情况；为了加速收敛，结合模拟退火思想，定义退火温度为  $T$ ，当轮迭代的解集数量为  $k_i$ ，上轮迭代的解集数量为  $k_{i-1}$ 。当  $k_i < k_{i-1}$  时，以概率  $\frac{(k_{i-1} - k_i)}{T(k_{i-1})}$  结束迭代。分群算法伪代码见 Algorithm 5。

---

**Algorithm 5** 高速公路社群划分方法
 

---

**Require:** 高速车辆 O-D 数据，高速公路网络拓扑结构，最大社群节点数量

**Ensure:** 高速公路社群划分结果

```

1: function COMMUNITY(ODMatrix G = V, E B)
2:   res  $\leftarrow [[\{0, 0\}, \{1, 1\}, \dots, \{n, n\}]]$ 
3:   tmp  $\leftarrow [\{\}]$ 
4:   pre  $\leftarrow [[\{0, 0\}, \{1, 1\}, \dots, \{n, n\}]]$ 
5:   k  $\leftarrow 0$ 
6:   l  $\leftarrow 0$ 
7:   T  $\leftarrow 100$ 
8:   while  $|len(res) - len(pre)| \leq T$  do
9:     res  $\leftarrow res[-1]$ 
10:    pre  $\leftarrow res$ 
11:    while res $[-1] \notin res[0 : -1]$  do
12:      tmp  $\leftarrow res[-1]$ 
13:      for i  $\in E$  do
14:        for ( do C  $\in tmp \& |C| \leq B)
15:          if  $\Delta Q > k$  then
16:            l  $\leftarrow C$ 
17:            k  $\leftarrow \Delta Q$ 
18:          end if
19:        end for
20:        tmp $[l] \leftarrow i$ 
21:      end for
22:      res add tmp
23:    end while
24:    T $\leftarrow$ 
25:  end while
26:  return res
27: end function$ 
```

---

### 4.2.3 基于社群划分的关键路段求解

基于已经分群的高速公路网络，提出关键路段挖掘方法。

本节算法的核心思想是分治法。首先分割社群，将每个社群看作独立的路网。在每个社群内部，用贪心算法选出关键路段，记录每个关键路段被处理后，路段通行效率的增量  $\Delta L$ 。忽略不同社团的关键路段之间的相互影响。这样，分治法的合并问题被归类于投资问题。

投资问题的定义如下：定义资产总额为  $B$ ，总共有  $\{X_1, X_2, \dots, X_u\}$   $u$  种货物，每种货物可以投资  $[0 - B]$  份资源。 $g(y * X_i)$  表示当货物  $X_i$  投资量为  $y$  时，它所带来的收益。本节中，定义关键路段的数目为  $B$ ，社团  $i$  由  $X_i$  表示，每个社团中，用  $g(t * X_i)$  计算改造关键路段造成的影响。在此提出目标模型：

$$Q = \text{Max} \left( \sum_{i=1}^u g(y * X_i) \right) \quad (4.6)$$

$$\text{Subject to. } \begin{cases} \sum_{i=1}^u y \leq B \\ x_i \geq 0 \end{cases} \quad (4.7)$$

**4.6** 属于投资问题，可以利用动态规划，在多项式时间内求解。求解步骤如下：

定义  $f_k(x)$  为前  $k$  个社团投入  $x$  份资源时，高速公路通行效率的最大提升量，首先赋初始值： $f_0(x) = 0$ ； $f_k(0) = 0$ ； $f_1(x) = g(x * X_1)$ 。递推公式：

$$f_i(j) = \text{Max}_{0 \leq y \leq j} (CMatrix[i][y] + f_{i-1}(j - y))$$

公式中， $CMatrix$  矩阵表示当社团  $i$  投资  $y$  条路段时，路网通行效率的增加量； $y$  表示本社团中投资路段的数量，根据  $y$  和贪心算法结果，可以推算出具体路段。当  $k = 2$  时， $CMatrix$  矩阵已知， $f_1(x)$  全部已知，据此可以推算出所有  $f_2(x)$  的值。递推，依次得到

$f_3(x), \dots, f_u(x)$  的值, 结合  $f_u(x)$  中 y 的值, 根据贪心算法, 获取关键路段, 反向递推, 得到最终关键路段的集合。动态规划伪代码:

---

**Algorithm 6** 关键路段挖掘方法求解
 

---

**Require:** 每个社团中选取不同路段的收益, 高速公路网络社团结构, 最大社群节点数量

**Ensure:** 高速公路关键路段集合

```

1: function COMMUNITY(CMatrix C = c1, c2, ..., cu B)
2:   定义  $f_k(x)$ : 当前 k 个社团投入 x 份资源时, 最大的通行效率提升量
3:    $f_0(x) \leftarrow 0$ 
4:    $f_k(0) \leftarrow 0$ 
5:    $f_1(x) \leftarrow CMatrix[1][x]$ 
6:    $i = j = 1$ 
7:   while  $i \leq u$  do
8:     while  $j \leq B$  do
9:        $f_i(j) = \text{Max}_{0 \leq y \leq j} (CMatrix[i][y] + f_{i-1}(j-y))$ 
10:       $j = j + 1$ 
11:    end while
12:     $i = i + 1$ 
13:  end while
14:  return res
15: end function
```

---

其中, 路段收益矩阵 **CMatrix**:

$$\begin{matrix} & c_{11} & c_{12} & \cdots & c_{1B} \\ & c_{21} & c_{22} & \cdots & c_{2B} \\ & \vdots & \vdots & \ddots & \vdots \\ & c_{u1} & c_{u2} & \cdots & c_{uB} \end{matrix}$$

矩阵中,  $c_{ij}$  表示第  $i$  个社群中, 选取  $j$  条关键路段进行资源投放后, 高速公路网络通行效率的提升量。这一数据由贪心算法在每个社群分别求得。

### 4.3 实验及结果

本章节出了针对每一种方法的有效性做出实验, 并将基于高速公路社群划分方法的实际效果与通过枚举得到的最优解进行对比。

基本的社群划分存在分辨率限制和极端退化特性。分辨率限制是指社群划分方法无法发现小于一定规模的社群，极端退化特性是指最终的社群划分结果会收敛于指数数量级的高分解决方案，而不是指向一个或少量最优解。Newman[29] 采用一种方法解决低分辨率问题：初始化时，将每一个节点看作一个独立的社群，之后根据模块化函数不断循环修正节点的所属社群。这个方法用在高速公路上时，虽然解决了低分辨率社群无法发现的问题，但是最终会产生一系列孤立点（如图4.3），这不符合社群划分的初衷。而且最终结果也没有避开极端退化特性，最终的社群划分结果在一个非常大的解空间中循环。

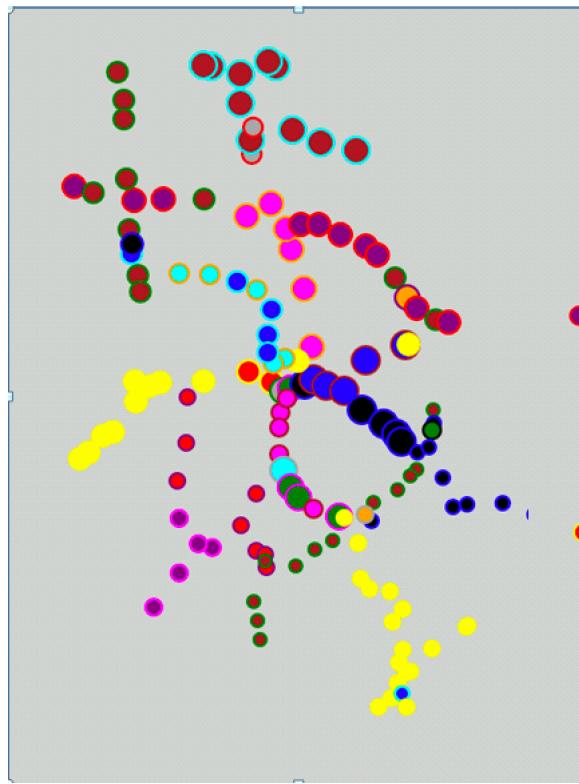


图 4.3 基于模块化函数的社群划分方法

图4.3给出了基本的基于模块化函数的分群结果，首先需要指出：使用基础方法的分群结果收敛于一个具有一千多个解的解集合。最终分群结果会在这些解集合内循环。由图我们可以看出两个问题：

- 1) 存在很多未被分群的孤立点。

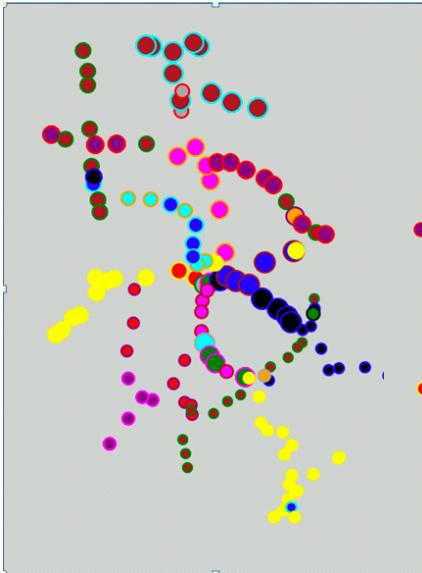


图 4.4 基于模块化函数的社群划分方法

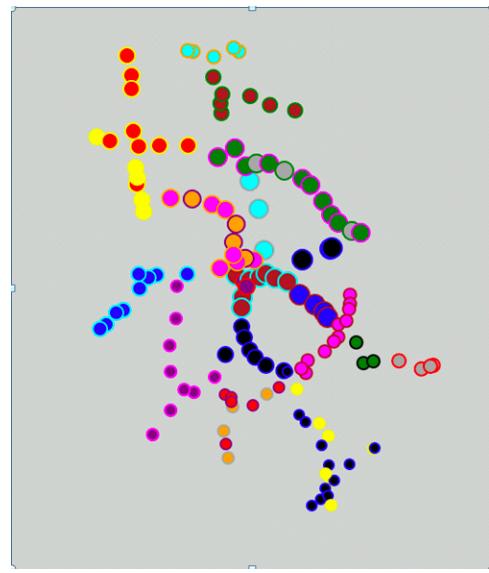


图 4.5 结合物理路网特性的社群挖掘

2) 很多社群存在物理意义上的交叉收费站。

孤立点的产生原因有两个，一是这个收费站本身流量较小，与其他站点交流不多；二是这个站点与其他站点之间的交流较为平均，站点不断流动于不同的社团中。图4.5是基于公式4.2的社群划分结果，该图由几百个社群划分组成的解集中选出，由图可以看出加入物理路经长度的情况下，可以在一定程度上消除孤立点，并且将高速公路划分成较为清晰的几类。但是我们发现仍旧有少量社群，存在物理层面的相互交叉情况，而这种情况不符合高速公路这种物理网络的社群划分特点。

经过数据分析，出现图4.5中不同社群内部的节点之间存在物理上的交叉情况的原因是——不同社群节点之间的流量差远大于节点之间的距离差。直接将具有交叉节点的社群进行合并虽然简单有效，但是不具有更大规模的适应性，这种方法得到的社群划分效果得不到保证，而且有可能出现过大的社团，不符合社群划分的目的。图4.6给出了基于模拟退火方法的迭代分群方法结果，该结果最终收敛于由 5 个结果组成的结果集，基本消除所有孤立点与社群交叉节点。

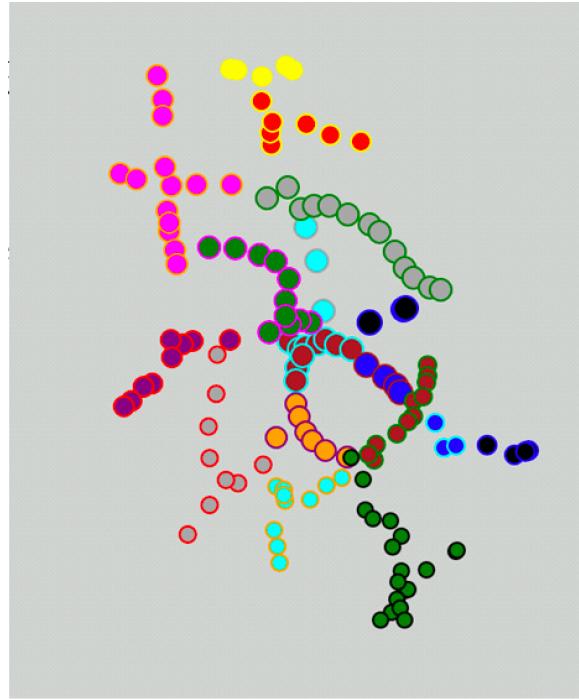


图 4.6 结合路段距离的变权社群挖掘结果

根据公式4.1给出的模块化函数  $Q$ ，表4.1给出了不同社群方法模块化的效果。可以看出将边权与物理距离结合考虑后，模块化效果得到了显著提高；虽然模拟退火方法的时间消耗较大，但是它提供了符合物理网络的分群结果，减少不同社群之间的交叉节点，将不同社群之间节点的相互影响降到最低。

	基于流量划分	基于流量/距离划分	基于变化距离的模拟退火
模块化效果	-1321.21	-1025.50	-1182.84
算法效率	1min	30s	1.5min
收敛度	$10^3$	$10^2$	0 – 10

表 4.1 不同社群划分方法效果对比

图4.7给出了一天时间内，基于分群算法和简单贪心方法的对比试验；图4.8给出了在一周时间内两种方法的对比试验。和上一章节一样，横坐标表示时间，纵坐标表示路网通行效率的绝对值（路网通行时间）。由图可以看出，简单贪心算法和基于分群算法的关键路段挖掘算法之间的误差较为平稳，并且一直维持在一个较低的水平线上。

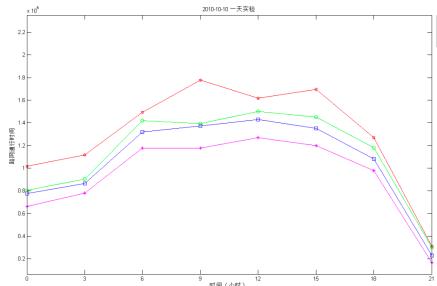


图 4.7 对比实验：以 1h 为区间

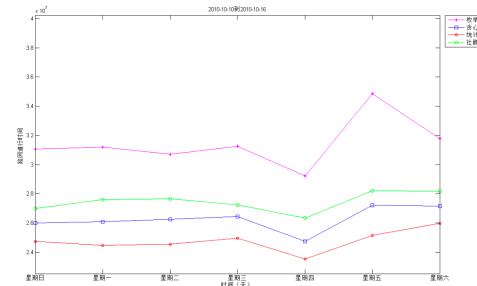


图 4.8 对比实验：以 1d 为区间

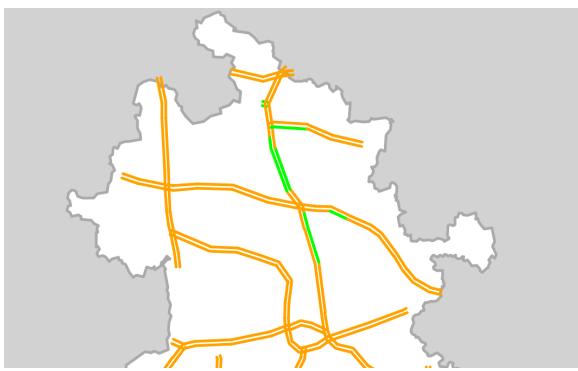


图 4.9 枚举求得关键路段

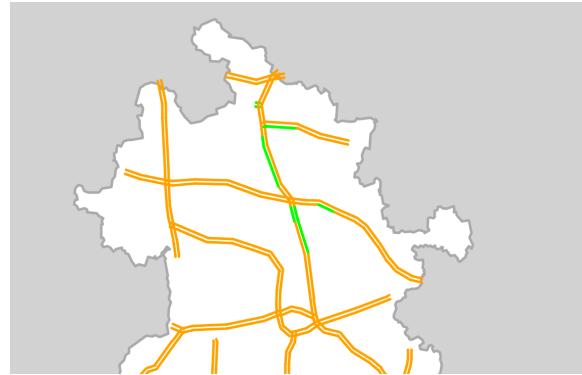


图 4.10 贪心求得关键路段

由图可以看出，分群算法可以在和统计算法相似的时间复杂度上，得到比统计算法优秀的解集。

下图给出不同方法选出的关键路段集合，图4.9给出了枚举方法选出的关键路段集合，图4.10给出了简单贪心算法给出的关键路段集合，图4.11给出了结合社群划分的关键路段识别算法的结果，图4.12给出了基于统计学的关键路段集合。观察图4.9和图4.10，发现两者选取的关键路段具有很强的相似性。

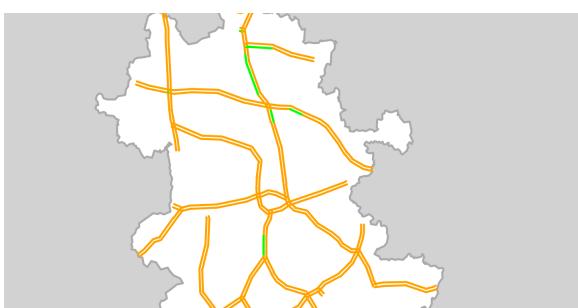


图 4.11 基于社群划分的关键路段

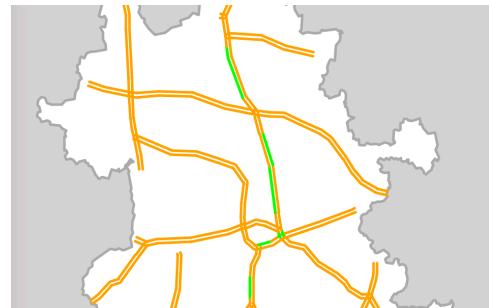


图 4.12 基于统计学的关键路段

误差分析：

	枚举-直接贪心	直接贪心-基于社群划分
一小时	14.63%	12.89%
一天	13.25%	13.26%
一周	13.10%	15.61%
一月	12.99%	11.59%

表 4.2 路网通行效率提升量误差分析

表4.2描述了枚举方法和直接贪心方法之间的误差，直接贪心和基于社群划分方法之间的误差。误差由高速路网的通行效率计算，可以看出误差在允许范围内。

关键路段选取平均误差：

	枚举-直接贪心	枚举-基于社群划分
一小时	0.18	0.25%
一天	0.14%	0.20%
一周	0.15%	0.19%
一月	0.14%	0.17%

表 4.3 关键路段选取误差分析

表4.3分析了关键路段选取情况的误差，采用欧式距离来刻画区别。可以看出，随着数据集的扩大，基于社群划分方法的关键路段准确率逐步上升。

运行效率分析：

	枚举	直接贪心	基于社群划分	基于统计
一小时	1day	30min	2min	1min
一天	6day	2h	5min	2min
一周	7day	3h	6min	5min
一月	7day	3h	7min	8min

表 4.4 不同方法运行效率分析

由表4.4可以看出，基于社群划分方法可以将整个算法的时间复杂度再降一个数量级，而结合表4.2来看，精度误差处于可接受范围( $1/e$ )。

#### 4.4 本章小结

本章提出了面向高速公路的社群划分方法，首先分析了传统方法的局限性，然后结合高速公路的独有特性，采用多变权值-模拟退火结合的方法，实现符合高速公路网络特点的社群划分方法。最后结合动态规划方法，实现多项式时间求解。



## 第五章 原型系统的设计与实现

### 5.1 系统功能

系统主要有如下功能：

- 1) 社群划分。根据输入的时间信息，结合数据库中的高速公路收费信息，进行社群划分。
- 2) 关键路段挖掘。根据输入的时间信息，结合数据库中的高速公路收费信息，进行关键路段挖掘。
- 3) 静态挖掘。根据历史信息，稳定挖掘关键路段，给管理者提供参考。
- 4) 动态挖掘。根据实时数据，实现动态挖掘关键路段。

### 5.2 系统架构

高速公路关键路段识别算法是为了解决高速公路管理者解决高速公路资源配比问题而开发的，整个系统基于 B/S 架构，从高层划分为前端和后端两部分，具体系统架构见图 5.1，后端为 Windows 服务，主要包括实时车辆数据的处理以及数据仓库的存储。本文使用了该系统的架构，通过编写“业务逻辑层”和“展现层”，实现原型系统。

系统逻辑如图 5.2 所示，首先对交通数据进行处理，剔除噪音数据，如：①车辆在半小时时间内横跨安徽省；②车辆丢失入口数据；③车辆丢失出口数据等。用筛选后的数据组成车辆 O-D 矩阵；从数据库内抽取重大交通事故信息，结合新闻中的高速公路事故信息，计算并输入路段的损毁概率。当选取贪心算法时，系统直接根据当前输入数据，进行贪心计算，注意本系统中将关键节点维护后的损毁率下降设为 0.1，具体数值可以根据实际应用调整；当选取基于社群划分的

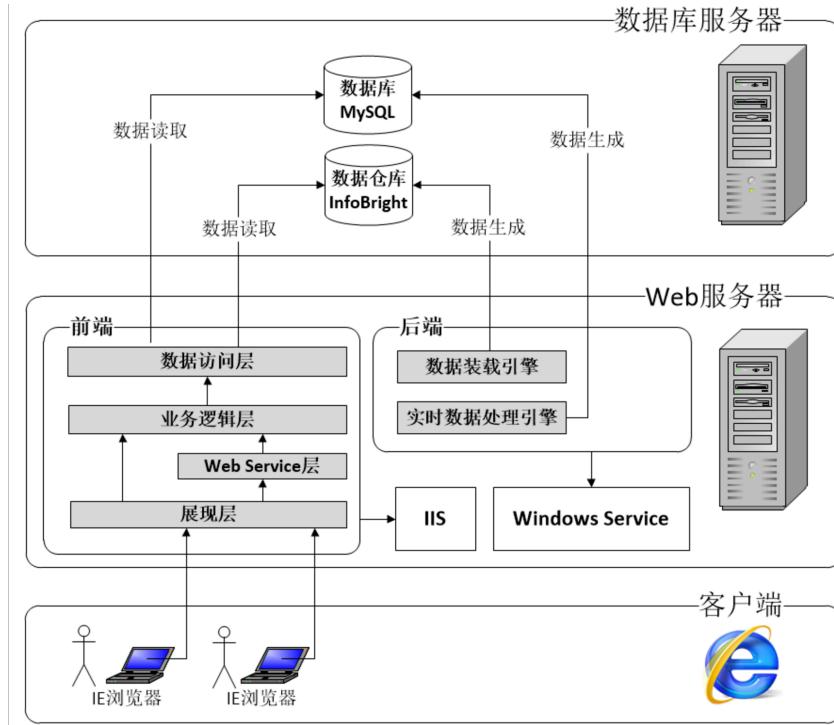


图 5.1 智能高速系统架构

关键路段挖掘策略时，首先根据 O-D 流量信息计算社群分布，并给出划分结果（图5.3）、网络模块化程度，之后对每一个社群内部进行贪心求解，这个求解过程并行执行，最后结合动态规划，求得最终解集合（图5.4）。

### 5.3 界面功能展示

系统的输入有时间，时间区间，路段损毁概率，管理者对关键路段的维护效果。

对于静态关键路段挖掘，直接输出关键路段结果（如图5.4）；对于动态关键路段挖掘，系统输出两层信息，第一层是关键路段分群效果（图5.3）；第二层是关键路段选取结果（图5.4）。

目前系统只适用于安徽高速系统，但是可以扩展到任何具有收费站数据的中国高速公路上。

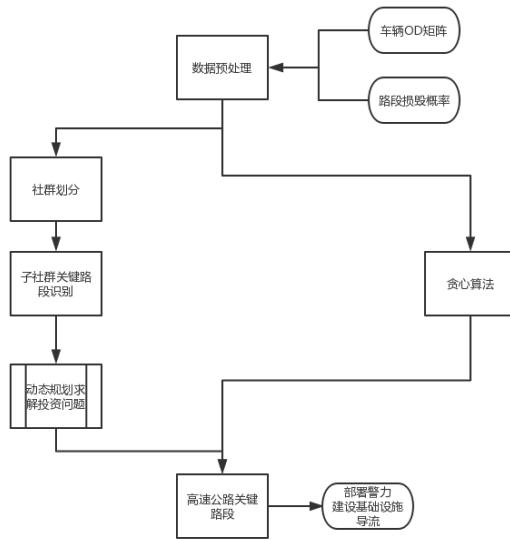


图 5.2 逻辑流程图

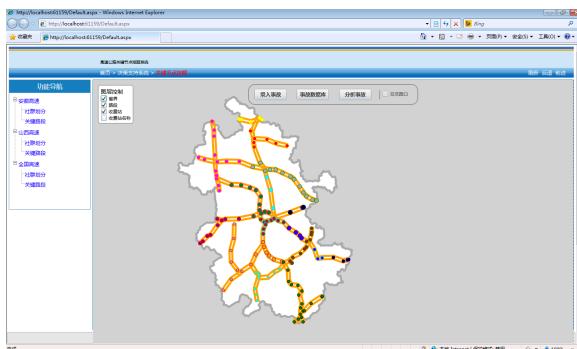


图 5.3 系统分群结果图

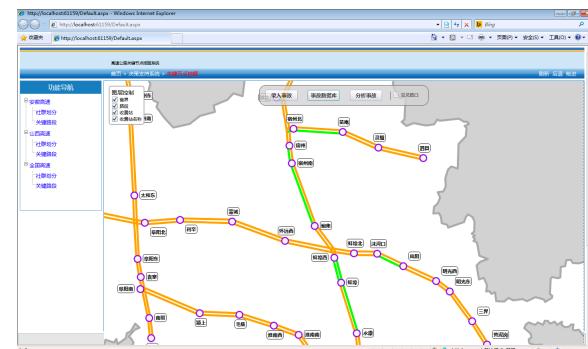


图 5.4 系统路段选取结果图

## 5.4 本章小结

本章介绍了一个面向高速公路的关键路段挖掘原型系统，给出了系统的流程图和样例图。该系统现在已经在安徽省高速公路网络上完全实现。该系统可扩展性强，所以可以很快的复用到其他省份乃至全国高速公路网络中。



## 总结与展望

交通系统中的关键节点识别非常重要。在高速公路系统中，关键节点的损毁会对整个系统的性能造成显著的影响，带来重大的经济损失。所以识别关键节点，在发生事故或者自然灾害之前进行维护巩固，在发生事故后进行快速修复，维护网络完整性非常重要。

### 5.4.1 主要工作

本文的主要工作和创新点如下：

1) 结合高速公路特性，提出高速公路关键路段挖掘模型，证明模型的贪心可解性，给出贪心解法。

重要节点一般数量非常少，但其影响却可以快速地波及到网络中大部分节点。例如，在对一个无标度网络的蓄意攻击中，少量最重要节点被攻击就会导致整个网络瓦解；网络的“小世界特性”和“无标度特性”的发现掀起了网络科学持续 10 多年至今丝毫没有降温的研究热潮。网络科学研究的热点逐渐从早期发现跨越不同网络的宏观上的普适规律转变为着眼于从中观（社团结构、群组结构）和微观层面（节点、链路）去解释不同网络所具有的不同特征。重要节点的挖掘研究也逐渐转为以微观研究为主。然而，从管理者的角度来说，过于微观的研究又无法体现高速公路系统的宏观特性。所以在此本文提出一种结合宏观（目标函数基于宏观理念）微观（基于路段损毁的随机规划问题）的高速公路关键路段挖掘模型，有效的解决了微观研究无法很好的顾及整体的问题。

2) 结合高速公路的复杂网络，提出一种基于复杂网络社群划分的关键路段挖掘方法。

传统的复杂网络社群划分与高速公路不同，首先高速公路是一种

相当稀疏的复杂网络，网络中的拓扑结构特性不是很复杂；其次高速公路和普通的复杂网络不同，他的不同的节点之间具有物理空间距离，和其他复杂网络如社交网络中的距离概念不同。所以传统的复杂网络社群划分方法已经不再适用。在此引入可变权值方法，有效的解决了传统分群算法中的低分辨率特性以及极端退化特性。同时采用模拟退火思想加速模型收敛过程。

#### 5.4.2 未来工作展望

本文工作虽然具有一定的创新性和实用性，但仍然存在一些局限和不足，需要在今后的研究中进一步探讨和完善，主要包括以下几个方面：

- 1) 由于时间限制，目前只在安徽路网上做完了原型系统。下一步的工作是将系统复用到全国的高速公路网络。
- 2) 目前分群算法结合高速公路的物理空间特性，在社群划分效果上可能有一定的损失。随着中国交通建设的不断完善，高速公路监测数据也越来越丰富。希望之后可以结合新的高速数据集，进一步完善高速公路关键路段挖掘方法。

## 参考文献

- [1] 姚寿康, 曹小军 and 刘杰. “浅析高速公路养护技术”. 工程技术: 全文版, **2017**, (2): 00155–00155.
- [2] A Khadivi, Rad A Ajdari and M Hasler. “*Network community-detection enhancement by proper weighting.*” *Physical Review E Statistical Nonlinear & Soft Matter Physics*, **2011**, 83(2): 894–901.
- [3] Erik Jenelius, Tom Petersen and Lars Göran Mattsson. “*Importance and exposure in road network vulnerability analysis*”. *Transportation Research Part A Policy & Practice*, **2006**, 40(7): 537–560.
- [4] Matthew G Karlaftis and Ioannis Golias. “*Effects of road geometry and traffic volumes on rural roadway accident rates*”. *Accident Analysis & Prevention*, **2002**, 34(3): 357–365.
- [5] Boris S. Kerner, Micha Koller, Sergey L. Klenov et al. “*The physics of empirical nuclei for spontaneous traffic breakdown in free flow at highway bottlenecks*”. *Physica A Statistical Mechanics & Its Applications*, **2015**, 438: 365–397.
- [6] Christian Yip, Phillip Fiorenzo, Kil Do Jung et al. “*A network-based congestion management model for Safety Service Patrol vehicle deployment*”. **2016**: 26–31.
- [7] Yacine Achour, Abderrahmane Boumezbeur, Riheb Hadji et al. “*Landslide susceptibility mapping using analytic hierarchy process and information value methods along a highway road section in Constantine, Algeria*”. **2017**, 10.
- [8] Shuyun Niu, Bin Li, Wujun Niu et al. “*Evaluation of Highway Network Node Importance via Node Benefit Function and Weighted Node Betweenness*”. **2015**.
- [9] Xinsheng Song, Xiaoxiao Wang, L. I. Aizeng et al. “*Node Importance Evaluation Method for Highway Network of Urban Agglomeration*”. *Journal of Transportation Systems Engineering & Information Technology*, **2011**, 11(2): 84–90.
- [10] Li Wang, Pei Zhou Yan, Ying Hong Li et al. “*Signal sub-control-area division of traffic complex network based on nodes importance assessment*”. **2011**: 5606–5609.
- [11] Zhengwu Wang, Aiwu Kuang and Hejie Wang. “*Calculating Node Importance Considering Cascading Failure in Traffic Networks*”. *Research Journal of Applied Sciences Engineering & Technology*, **2013**, 5(1): 264–269.
- [12] Srinivas Peeta, F Sibel Salman, Dilek Gunnecc et al. “*Pre-disaster investment decisions for strengthening a highway network*”. **2010**, 37(10): 1708–1719.
- [13] Jon M Kleinberg. “*Authoritative Sources in a Hyperlinked Environment*”. *Journal of the ACM (JACM)*, **1999**, 46(5): 604–632.
- [14] 李鹏翔, 任玉晴 and 席酉民. “网络节点(集)重要性的一种度量指标”. 系统工程, **2004**, 22(4): 13–20.
- [15] 陈勇, 胡爱群 and 胡啸. “通信网中节点重要性的评价方法”. 通信学报, **2004**, 25(8): 129–134.
- [16] Chavdar Dangalchev. “*Residual closeness in networks*”. *Physica A Statistical Mechanics & Its Applications*, **2006**, 365(2): 556–564.

- [17] Steffen Dereich and Peter Mörters. “*Random networks with sublinear preferential attachment: The giant component*”. *Annals of Probability*, **2013**, 41(1): 329–384.
- [18] C. M. Schneider, A. A. Moreira, Andrade Js Jr et al. “*Mitigation of malicious attacks on networks*.” *Proceedings of the National Academy of Sciences of the United States of America*, **2011**, 108(10): 3838.
- [19] Swami Iyer, Timothy Killingback, Bala Sundaram et al. “*Attack Robustness and Centrality of Complex Networks*”. *Plos One*, **2013**, 8(4): e59613.
- [20] 周涛, 傅忠谦, 牛永伟 et al. “复杂网络传播动力学研究综述”. *自然科学进展*, **2005**, 15(5): 513–518.
- [21] Linyuan Lü, Duan Bing Chen and Tao Zhou. “*Small world yields the most effective information spreading*”. *New Journal of Physics*, **2011**, abs/1107.0429(12): 825–834.
- [22] X. L. Peng, X. J. Xu, X. Fu et al. “*Vaccination intervention on epidemic dynamics in networks*”. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, **2013**, 87(2): 022813.
- [23] C. D Brummitt, R. M D’Souza and E. A Leicht. “*PNAS Plus: Suppressing cascades of load in interdependent networks*”.
- [24] Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin et al. “*Identification of influential spreaders in complex networks*”. *Nature Physics*, **2010**, 6(11): 888–893.
- [25] Phillip Bonacich. “*Factoring and Weighting Approaches to Clique Identification*”. *Journal of Mathematical Sociology*, **1972**, 2(1): 113–120.
- [26] M. Girvan and M. E. J. Newman. “*Community structure in social and biological networks*”. *Proceedings of the National Academy of Sciences of the United States of America*, **2001**, 99(12): 7821.
- [27] Joshua R Tyler, Dennis M Wilkinson and Bernardo A Huberman. *Email as spectroscopy: automated discovery of community structure within organizations*. Springer Netherlands, **2003**: 143–153.
- [28] Jin Di, Liu Jie, Zheng Xue Jia et al. “*k-Nearest-Neighbor Network Based Data Clustering Algorithm*”. *Pattern Recognition & Artificial Intelligence*, **2010**, 23(4): 546–551.
- [29] M. E. J. Newman. “*Fast algorithm for detecting community structure in networks*.” *Physical Review E Statistical Nonlinear & Soft Matter Physics*, **2004**, 69(6 Pt 2): 066133.
- [30] R Guimerà and L. A. Nunes Amaral. “*Functional cartography of complex metabolic networks*.” *Nature*, **2005**, 433(7028): 895.
- [31] M. E. Newman. “*Modularity and community structure in networks*”. In: *APS March Meeting*, **2006**: 8577–8582.
- [32] Soma Sanyal and James Glazier. “*Viscous instabilities in flowing foams: A Cellular Potts Model approach*.” *Journal of Statistical Mechanics Theory & Experiment*, **2006**, 2006(10): p10008.
- [33] U. N. Raghavan, R Albert and S Kumara. “*Near linear time algorithm to detect community structures in large-scale networks*”. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, **2007**, 76(3 Pt 2): 036106.
- [34] I. X. Leung, P. Hui, P Liò et al. “*Towards real-time community detection in large networks*.” *Physical Review E Statistical Nonlinear & Soft Matter Physics*, **2009**, 79(6 Pt 2): 066107.

- [35] M. J. Barber and J. W. Clark. “*Detecting network communities by propagating labels under constraints*”. *Physical Review E*, **2009**, 80(2 Pt 2): 026129.
- [36] Jin Di, Yang Bo, Liu Jie *et al*. “*Ant Colony Optimization Based on Random Walk for Community Detection in Complex Networks*”. *Journal of Software*, **2012**, 23(3).
- [37] Stijn Marinus Van Dongen. “*Graph Clustering by Flow Simulation*”. *Open Access*, **2000**.
- [38] 金弟, 杨博, 刘杰 *et al*. “复杂网络簇结构探测——基于随机游走的蚁群算法”. *软件学报*, **2012**, 23(3): 451–464.
- [39] Martin Rosvall and Carl T. Bergstrom. “*Maps of Random Walks on Complex Networks Reveal Community Structure*”. *Proceedings of the National Academy of Sciences*, **2008**, 105(4): 1118–1123.
- [40] Irinel Constantin Moroșescu and Antoine Girard. “*Opinion Dynamics With Decaying Confidence: Application to Community Detection in Graphs*”. *IEEE Transactions on Automatic Control*, **2009**, 56(8): 1862–1873.
- [41] Yan Liu, Qing Xian Wang, Qiang Wang *et al*. “*Email Community Detection Using Artificial Ant Colony Clustering*”. In: *Advances in Web and Network Technologies, and Information Management, APWeb/WAIM 2007 International Workshops: DBMAN 2007, WebETrends 2007, PAIS 2007 and ASWAN 2007, Huang Shan, China, June 16-18, 2007, Proceedings*, **2007**: 287–298.
- [42] Yan Liu, Junyong Luo, Huijie Yang *et al*. “*Finding Closely Communicating Community Based on Ant Colony Clustering Model*”. In: *International Conference on Artificial Intelligence and Computational Intelligence*, **2010**: 127–131.
- [43] Mursel Tasgin, Amac Herdagdelen and Haluk Bingol. “*Community Detection in Complex Networks Using Genetic Algorithms*”. *Corr*, **2007**, 2005(3120): 1067–1068.
- [44] M. E. Newman and M Girvan. “*Finding and evaluating community structure in networks.*” *Physical Review E Statistical Nonlinear & Soft Matter Physics*, **2004**, 69(2 Pt 2): 026113.
- [45] Fortunato and Santo. “*Community detection in graphs*”. *Physics Reports*, **2010**, 486(3–5): 75–174.
- [46] Duanbing Chen, Linyuan Lü, Ming Sheng Shang *et al*. “*Identifying influential nodes in complex networks*”. *Physica A Statistical Mechanics & Its Applications*, **2012**, 391(4): 1777–1787.
- [47] Phillip Bonacich. “*Factoring and weighting approaches to status scores and clique identification*”. *The Journal of Mathematical Sociology*, **1972**, 2(1): 113–120.
- [48] Chuanxiang Ren, Fasheng Liu and Li Zhou. “*The Position of Connected-way for Highway and Urban Road Based on Node Importance Analysis Method*”. *Journal of Shandong University of Science & Technology*, **2014**.
- [49] Zhi Wang and Jianzhi Zhang. “*In Search of the Biological Significance of Modular Structures in Protein Networks*”. *Plos Computational Biology*, **2007**, 3(3): e107.



## 附录 A 附件



## 致谢

时光荏苒，我的硕士生涯已接进尾声。这几年的时光既漫长又短暂，其中充满了酸甜苦辣，更有收获和成长。几年来，感谢陪我一起度过美好时光的每位尊敬的老师和亲爱的同学，正是你们的帮助，我才能克服困难，正是你们的指导，我才能解决疑惑，直到学业的顺利完成。

本人的学位论文是在我的恩师宋国杰副教授的殷切关怀和耐心指导下进行并完成的，衷心感谢我的恩师对我的淳淳教诲和悉心关怀。从课题的选择、项目的实施，直至论文的最终完成，宋老师都始终给予我耐心的指导和支持，我取得的每一点成绩都凝聚着恩师的汗水和心血。宋老师开阔的视野、严谨的治学态度、精益求精的工作作风，深深地感染和激励着我，在此谨向宋老师致以衷心的感谢和崇高的敬意。

感谢实验室的师弟师妹们，让我在最困难的时候看到你们的笑脸，被你们开朗的情绪感染。感谢你们带给我的帮助和鼓励！感谢一起打篮球的小伙伴，感谢谢老师组织的篮球活动，让我重新拾回篮球这一爱好。

最后，我要向百忙之中参与审阅、评议本论文各位老师、向参与本人论文答辩的各位老师表示由衷的感谢！人生的每个阶段都值得好好珍惜，这段美好岁月，因为有你们的关心和帮助，我很幸福。我会更加勤奋学习、认真研究，我会努力做得更好，我想这也是我能给你们的最好的回报吧。把最美好的祝福献给你们，愿永远健康、快乐！