

# 基于实时流量数据的交通网络脆弱性研究

郑州轻工业大学 王静怡、钟毅、朱姣姣

## 摘要

在全球化和信息化的背景下,中国政府发布《交通强国建设纲要》,旨在通过科技创新推动交通行业的质量、效率和动力变革。智慧交通作为关键领域,利用大数据等技术提高交通管理的智能化和精细化,对于促进经济社会发展具有重要作用。

本文提出了一个基于实时道路流量数据的交通网络脆弱性研究框架,并给出了交通网络静态和动态脆弱性分析方法。首先,本研究选用贵州省某城市交通网络道路信息数据集,对数百万条数据进行清洗与处理,并进行时间序列分析和聚类,为实现精准的交通流量预测和网络脆弱性分析奠定了基础。其次,构造了基于复杂网络理论的拓扑结构,从度、介数、接近度中心性和特征向量中心性等角度构建了静态脆弱性测度体系,提出了静态脆弱性分析的框架与方法。再次,基于交通流量预测数据,构建了交通网络级联失效模型,进行了动态脆弱性分析,模拟了早高峰时段可能发生的级联失效情况。最后,针对识别出的脆弱路段,给出了科学的交通管理对策与建议,不仅为交通管理部门提供了技术支持,还为城市交通政策的制定和优化提供了决策依据。

本研究提出的交通流量预测与网络脆弱性分析框架和方法不仅为城市交通流量的精准预测和交通网络鲁棒性评估提供了科学的方法,还给实时交通流量监测、应急政策制定和交通基础设施建设等提供了有力的政策支持。

**关键词:**交通流量预测;级联失效模型;网络脆弱性分析;智慧交通

## 一、引言

### (一)研究背景和意义

在全球化和信息化时代背景下,中共中央、国务院于2019年印发的《交通强国建设纲要》标志着我国对交通运输行业的未来发展方向进行了全面而深远的规划。该纲要不仅是一份宏观的战略决策,更是一项具体的行动计划,旨在通过科技创新驱动加快建设交通强国,实现交通运输行业的质量变革、效率变革、动力变革。这一决策深刻体现了我国政府对交通发展重要性的认识,并且明确了建设交通强国的具体目标和行动指南,为我国未来一段时间内的交通发展提供了清晰的路线图和行动框架。

《交通强国建设纲要》中提出的发展目标,具体包括到2020年完成全面建成小康社会交通建设任务,到2035年基本建成交通强国,以及到本世纪中叶全面建成人民满意、保障有力、世界前列的交通强国。这些宏伟目标涵盖了构建现代化综合交通体系、提升人民满意度、实现城乡区域交通协调发展等多个方面,为交通管理的智能化和精细化提供了明确的方向。实现这些目标,需要依托于新一代信息技术与交通运输的深度融合,即智慧交通的发展。智慧交通作为推动交通运输行业创新实践最为活跃的领域,不仅代表了新型基础设施建设的重要方向,也是数字经济的重要组成部分,对于提升交通运输全要素生产率、推动经济社会发展具有重大意义。

智慧交通与新能源车辆的协同发展还将带动相关产业的发展,如智能交通设备制造、交通信息服务、车联网技术服务等,进一步提高道路利用效率、优化道路通行成本。智慧交通将运用先进的信息技术,采集到海量实时数据,实现交通管理的智能化、精细化,从而提高交通系统的运行效率和安全性。因此,如何高效利用智慧交通的大规模交通信息数据,提高道路通行效率,已成为一个亟待解决的现实难题。

本研究立足智慧交通的创新发展需求,基于对实时交通流量数据的挖掘和处理,给出交通网络脆弱性

分析框架与方法,不仅丰富了智慧交通领域的理论基础,提供了新的研究视角,而且为智慧交通技术的进步和实际应用开辟了新的道路。通过准确预测交通流量,可以更有效地进行交通需求管理、优化交通流的分配,减少交通拥堵,从而提高道路服务效率。

本研究还可以直接指导交通管理部门制定和实施更为科学和精确的交通管理策略,在交通高峰时段提前识别潜在的交通瓶颈,及时采取预防措施,如调整交通信号灯的时序,实施交通管制或发布出行建议,从而有效缓解交通压力。此外,本研究还能为城市交通规划提供数据支持,帮助决策制定者优化交通基础设施布局,提高公共交通的服务质量和效率。

## (二)文献综述

本文基于实时交通流量的网络脆弱性分析为研究目标,需要基于实时流量数据来预测可能发生的道路流量超载引起的网络级联失效。因此,本部分将从以下三个方面的研究现状展开阐述。

### 1. 时间序列预测

对于时间序列预测问题,线性模型和 ARIMA<sup>[12]</sup>被广泛应用。随着机器学习和深度学习研究的深入,时间序列预测有了新的发展方向,SVM 算法一直被称为金融时间预测的经典算法<sup>[13-14]</sup>。Jain 等人<sup>[15]</sup>将传统的时间序列方法与神经网络相结合来预测时间序列,显示出两种方法的优势。Zhang 等人证明了神经网络在线性时间序列预测中的有效性<sup>[16]</sup>,并利用神经网络对带有趋势成分的季节性时间序列进行建模和分析<sup>[17]</sup>,随后提出了 ARIMA 和神经网络的混合模型来预测时间序列<sup>[18]</sup>。Liu<sup>[19]</sup>、Siami-Namini<sup>[20]</sup>、Sagheer<sup>[21]</sup>等人验证了使用 LSTM 算法预测时间序列的优异性能。

### 2. 级联失效

城市交通网络级联失效模型的研究主要集中于在负重一承载模型上,该模型假设承载能力与初始负荷成正比<sup>[10-11]</sup>,并模拟了不同攻击模式下网络级联失效的演化过程<sup>[10-12]</sup>。Zhengetal.<sup>[10]</sup>提出了基于城市交通网络特征的 Crucitti-Latora-Marchiori(CLM)优化模型,引入拥堵效应确定节点负担,并将实际容量引入成本函数,以此反映城市交通网络级联失效。Wang 等人<sup>[10]</sup>提出了灾害传播的动态模型,模拟了城市交通网络级联失效的演变,并<sup>[10]</sup>建立了基于城市交通网络双层模型的灾害传播动态模型。经过对修复因子等参数的仿真分析,模型可以模拟网络失效的演化过程。国内的王正武等<sup>[11-13]</sup>基于道路交通的级联失效现象改进了路网节点重要度的测算法,随后采用正交实验法完成了交通网络级联失效的影响源识别。

### 3. 网络脆弱性分析

过去 20 年来,复杂网络一直在广泛的研究热点。1999 年 Wang 在《自然》杂志上发表了“Small World”,1999 年 Barabasi<sup>[22]</sup>在《科学》杂志上发表了“ScaleFree”,复杂网络的结构和特性研究越来越受到关注,并已渗透到数学、计算机科学、社会学、交通网络<sup>[23]</sup>和电力网络<sup>[24]</sup>等领域。在交通领域,网络的鲁棒性已被广泛研究,如公路网络、海洋网络、地铁网络<sup>[25]</sup>和公路网络<sup>[26-27]</sup>。网络的性能可以用脆弱性的概念来衡量,在交通系统中脆弱性通常被定义为对可能导致网络可用降低的因素的易感性<sup>[28]</sup>。网络可靠性降低的可能性<sup>[29]</sup>。

一般来说,许多研究都讨论过城市道路网络(URN)的脆弱性和稳健性,这些研究可分为两类。一些研究人员关注 URN 的拓扑结构,在 URN 中,事故和拥堵通常发生在交叉路口或阻塞路段,故而后影响最严重的关键位置被认为是关键网络<sup>[30]</sup>。Bono 和 Gutierrez<sup>[31]</sup>通过复杂网络理论和空间分析,对 2010 年海地地震后的道路网络可达性进行了分析和评估。通常,研究人员将连通性或可达性作为复杂网络的性能指标,如度、密度、聚类系数<sup>[32-33]</sup>等。研究人员根据节点的位置和节点间不同的流量研究节点间的相互作用,从而检验 URN 的脆弱性。

另一些研究人员专注于对严重的交通问题建立交通流量模型。在这些模型中,有一种是宏观交通模型,主要通过研究交通密度、交通流量和交通速度等交通参数之间的关系来描述宏观交通特性<sup>[34]</sup>。由于整个城市路网的交通流量、研究范围大、数据获取复杂,因此研究难度很大。目前,一些研究人员考虑利用交通加权法构建网络,并将其纳入城市道路网络的脆弱性评估中。随着城市的交通发展和人口的增加,城市交通网络变得越来越复杂,对通行可靠性的要求也越来越高<sup>[35]</sup>。道路交通网络本质上是加权的动态网络,在受到干扰后会发生演变。因此,从拓扑角度审视 URN 的属性和中心性已逐渐不能满足发展的需要。He 等人<sup>[36]</sup>开发了一个用道路网络构建,发现交通量对脆弱性有怎样的影响。Wang 等人<sup>[37]</sup>提出了一种基于真实出租车轨迹的新路网脆弱性分析框架。Zeng 等人<sup>[38]</sup>通过道路在一天中不同时段的流量并利用时序网络确

定了拥堵传播过程。Ma 等人<sup>[39]</sup>研究了网络结构如何影响不同速度下的拥堵率。

### 4. 文献评价

基于以上研究分析发现,对于时间序列预测,虽然上述提出的模型已经具备较好的性能,但针对交通网络流量的时间序列预测的准确度仍有提升的空间;对于级联失效,城市交通网络的级联失效模拟具有较为完善的理论模拟方法,但基于实时数据的级联失效模拟仍有不足,对于网络脆弱性分析,尽管存在基于交通网络特征和交通流量的网络脆弱性分析,但缺乏系统的基于实时流量数据的交通网络脆弱性分析框架,因此本研究将展开时间序列模型预测方法的优化,基于实时流量数据的网络级联失效模拟以及网络脆弱性分析框架的构建三项工作。

### (三)研究内容和方法

#### 1. 研究内容

本文首先搜集道路交通数据,对其进行 3σ 异常值检测和剔除,再对产生的缺失值进行移动平均插值数据,然后基于道路网络的连接情况构建城市交通网络的拓扑图,再根据城市交通网络流量的时间序列数据利用 XGBoost 和 ARIMA 的加权融合模型预测网络中每一条道路的指定时间内的流量数据,最后基于该流量预测数据,通过负载容量模型模拟城市交通道路网络级联失效,最后同时利用静态的传统网络拓扑结构测度和动态的城市交通网络级联失效推演完成交通网络的脆弱性分析,以此得到该道路交通网络中的最重要边。

#### 2. 研究方法

本文的具体研究方法包括:(1)文献研究法。梳理城市交通网络分析相关文献,形成研究架构,为项目开展提供理论基础。(2)数学建模与模型分析法。基于时间序列预测理论,级联失效理论,网络脆弱性分析理论,构建基于交通流量时间序列的预测模型,城市综合交通网络级联失效模型及分析城市交通网络的脆弱性。

#### 3. 技术路线图

本研究的技术路线图如图 1:

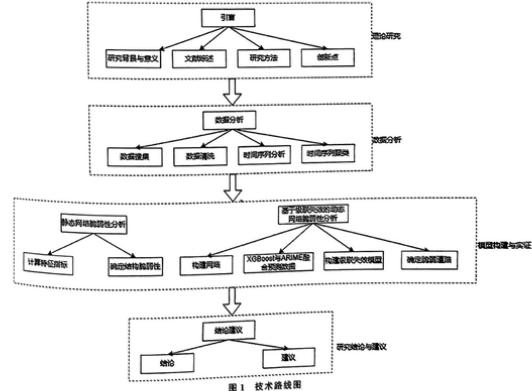


图 1 技术路线图

### (四)本文的创新点

本文创新性地构建了一个基于道路实时交通流量数据的交通网络脆弱性识别研究框架,提出了一种融合多源数据的交通流量预测模型,该模型能够准确预测早高峰时段的交通流量变化。在此基础上,本文给出了一种流量级联失效模型,用于模拟和分析交通网络在高峰时段可能发生的级联失效情况。通过该模型,本文对级联失效过程进行了详细的仿真分析,揭示了交通网络在面临各种干扰时的动态响应和失效传播机制。

进一步地,本文建立了一套静态和动态的交通网络脆弱性识别方法,通过定量分析网络结构的脆弱性,识别出早高峰时段最易失效的道路和静态网络拓扑结构上的脆弱路段,这些方法的应用,为交通管理部门提供了科学的决策支持,帮助其提前识别潜在的交通瓶颈和风险点,从而进行有针对性的交通组织和优化。

### 三、数据处理

#### (一)数据搜集

在数据收集阶段,首先分析了美国加州的 PEMS(Performance Measurement System)08 和 PEMS04 道路数据集,这些数据集因其较高的数据质量 and 较少的缺失值而受到关注。然而,美国与中国在道路设计、交通规则以及驾驶习惯等方面有显著差异,如美国与我国在汽车驾驶位置的不同,以及在人流量上的显著差异,若直接采用该道路数据平台分析中国交通网络存在局限性。然而,转向寻求国内的道路数据。深圳市有一个官方的数据服务平台,提供了包含实时交通流量信息的街道数据,该数据集包含超过五百万个道路记录,数据量充足。但是,该数据集缺少描述道路网连通关系的关键信息,且未提供道路编号与实际路段之间对应的关系,这限制了基于数据驱动的道路网构建和进一步分析。

经过深入调研,最终确定使用阿里云天池大数据众智平台提供的一份贵州省的道路数据集。该数据集不仅详细记录了道路的连接信息,而且数据量庞大,包含超过七百万条数据记录。此外,该数据集的道路平均旅行时间数据采集频率为每两分钟一次,这对于实现更精准时间尺度上的道路流量提供了可能。尽管如此,数据集本身存在的缺失值和异常值较多,这给数据处理工作带来了一定的挑战。

#### 1. 数据来源

来自于天池大数据众智平台<sup>①</sup>。

#### 2. 数据基本说明

贵州道路数据具有 132 条道路的静态信息,以及这些道路之间的上下拓扑结构,提供 2017 年 4 月 0 0:00 至 2017 年 6 月 30 日 23:58 每条路段的平均旅行时间,每两分钟记录一次。

#### ① 路段(link)属性表

每条道路的每个通行方向由多条路段(link)组成,数据集中会提供每条 link 的唯一标识,长度、宽度,以及道路类型,如表 1 所示;图 2 示例了地面道路 link1 和 link2 的属性信息。

表 1 路段属性表

属性	类型	说明
link_ID	string	每条路段(link)的唯一标识
length	double	link 长度(米)
width	double	link 宽度(米)
link_class	int	link 道路等级,例如 1 表示主干道

#### ② link 上下游关系表

link 之间按照行驶的方向存在上下游关系,数据集中提供每条 link 的直接上游 link 和直接下游 link,如表 2 所示;图 3 示例了 link2 的 in\_links 和 out\_links。

基于实时流量数据的交通网络脆弱性研究

图 2 路段属性信息示意图

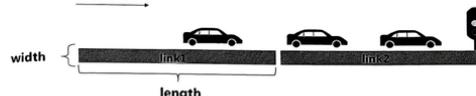


表 2 link 上下游关系表

属性	类型	说明
Link_ID	string	每条路段(link)的唯一标识
In_links	string	link 的直接上游 link.link_ID 之间以 # 分割
Out_links	string	link 的直接下游 link.linkID 之间以 # 分割

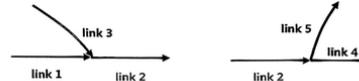


图 3 路段上下游关系图

③ link 历史通行时间表

数据集中记录了历史每天不同时间段内(2min 为一个时间段)每条 link 上的平均旅行时间,每个时间段的平均旅行时间是基于在该时间段内进入 link 的车辆在该 link 上的旅行时间产出。字段解释如表 3 所示:

表 3 link 历史通行时间表

属性	类型	说明
Link_ID	string	每条路段(link)的唯一标识
Date_time	date	日期,例如“2015-10-01”
Time_interval	string	时间段,例如“(2015-09-01,00:00,00,2015-09-01,00:00,00,10”
Travel_time	double	车辆在路上的平均旅行时间(秒)

#### ④ 数据清洗与处理

由于道路的各项交通数据可能受到交通管制、道路修理、极端天气的影响及对外信息披露透明度不足等问题,需对数据中异常值进行检测和处理。

#### 1. 数据正态性分析

##### ① 道路车辆平均旅行时间的分布

为探究道路车辆平均旅行时间的分布情况,如图 4 直方图所示,发现平均旅行时间是一个长尾分布,大多数的平均旅行时间比较小,只有很小一部分的平均旅行时间比较大。

因此对道路的平均旅行时间做对数变换。将道路平均旅行时间经过对数变换后,从图 5 可以看出数据的分布均匀,符合正态分布。

<sup>①</sup> 网址:<https://tianchi.aliyun.com/competition/entrance/231598/information>.

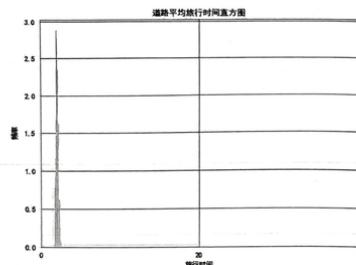


图 4 道路平均旅行时间直方图

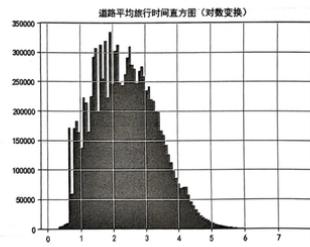


图 5 经过对数变换的道路平均旅行时间直方图

## ②正态性检验

为进一步确认道路的平均旅行时间数据是否符合正态分布,利用 Q-Q 图(Quantile-Quantile Plot)用来检验道路平均旅行时间的原始值与其实证分布的理论值是否同分布。如图 6 所示,大部分数据点近似地落在了一条直线上,表明数据近似服从正态分布,也有一些数据点(特别是在图的两端)偏离了这条直线,这表明数据的尾部行为与正态分布有所不同,因此考虑对这部分数据进行清洗。

## 2.3σ 准则检验异常值

根据上述正态性检验,得到道路平均旅行时间数据基本符合正态分布,因此基于  $3\sigma$  准则,剔除异常大或异常小的数据。

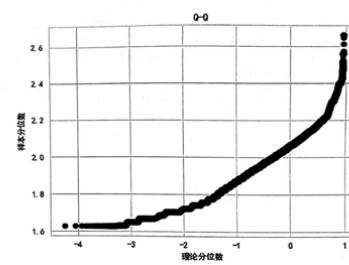


图 6 道路的平均旅行时间的 Q-Q 图

## 3. 缺失值补充

由于存在道路数据因为现实情况产生的缺失值以及前文经过  $3\sigma$  准则检测删除极端异常数据后出现的缺失值,为了保证数据的完整性与平稳性,需要对这些缺失值进行插值。

在时间序列分析中,移动平均插值(Moving Average Interpolation)用一个滑动窗口在时间序列上滑动处理缺失值,核心思想是利用时间序列中相邻数据点之间的相关性,通过取一段时间内的平均值来估计缺失数据点的值。因此,这种方法需要假设时间序列在短期内具有稳定性与连续性。

如图 7 为道路 ID 为 3377906280028510514 的路从 4 月到 6 月每个月的道路平均旅行时间的变化,离散点表示真实数据,直线表示对这个时间序列数据的趋势回归,图中可以看出这条路从 4~6 月的车辆平均旅行时间的趋势相对平稳,可以利用移动平均插值对缺失值进行补充。

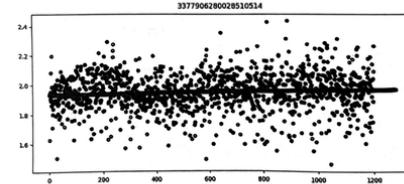


图 7 对道路平均旅行时间的回归示意图

对于道路平均旅行时间的移动平均插值,其公式如下:

$$\hat{Y}_t = \frac{1}{n} \sum_{i=t-k+1}^{t+k} Y_i \quad (1)$$

其中,  $\hat{Y}_t$  是时间点  $t$  的插值估计,  $n$  是时间窗口的大小(即  $2k+1$ , 其中  $k$  是窗口的一半长度),  $\hat{Y}_t$  是时间点

## z 的观测值。

插值步骤如下:首先,定义一个滑动窗口,该窗口包含 20 个连续的时间点的数据点,且随着时间序列的进行而移动。其次,当滑动窗口移动至包含缺失数据点的位置时,计算窗口内所有非缺失数据点的平均值。再次,将计算得到的平均值赋给窗口中心的缺失数据点,作为该点的估计值,之后继续移动滑动窗口,重复上述步骤,直到处理完所有的缺失数据点。

插值后的数据与原始数据的对比如图 8,粗线条是原始数据,细线条是插值后数据,可以看到经过移动平均插值后的数据保留了原始数据的趋势性,并且在数值上也符合  $3\sigma$  检验,这表明此插值合理且可使用。

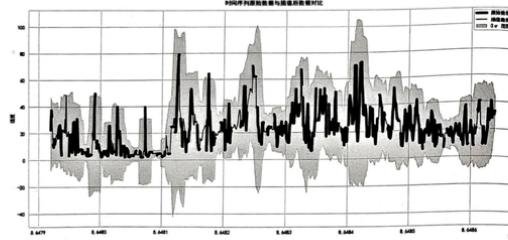


图 8 时间序列原始数据与插值后的数据对比图

## 4. 数据统计量描述

道路长度(Road Length):数据显示道路长度的分布非常广泛,从 5 米到 839 米不等,均值为 86.01515 米,但标准差高达 113.6054 米,表明道路长度之间存在很大的差异。这可能意味着道路的规模和类型非常多样化,从非常短的小巷到较长的主要干道都有。

道路宽度(Road Width):与长度相比,道路宽度的分布较为集中,均值为 9.931818 米,标准差为 3.014452 米,最小宽度为 3 米,而大多数道路(75%)的宽度在 9 到 12 米之间,这表明虽然存在一些较窄的道路,但大多数道路的宽度相差不大,显示出较好的一致性。

道路平均通行时间(Travel Time):道路的平均通行时间也显示出较小的变异性,均值为 8.367045 秒,标准差为 2.403522 秒,最短通行时间为 0.9 秒,而大多数道路(75%)的平均通行时间在 7.4075 到 9.75 秒之间。这可能反映出道路的通行效率普遍较好,尽管有一些道路的通行时间较短,而另一些较长,由表 4 所示:

表 4 原始道路车辆通行时间数据描述性统计量

columns	count	mean	std	min	25%	50%	75%	max
RoadLength	132	86.01515	113.6054	5	29	55.5	103.25	839
RoadWidth	132	9.931818	3.014452	3	9	9	12	15
TravelTime	132	8.367045	2.403522	0.9	7.4075	8.75	9.75	12.73

## (三)时间序列分析

## 1. ADF 检验

在交通流量预测工作中,需要基于平稳的交通流量数据进行后续的时间序列分析。本文采用增广广

基-富勒(Augmented Dickey-Fuller,简称 ADF)检验道路平均旅行时间的平稳性。

ADF 检验通过回归模型来描述时间序列数据的动态变化特征,并检验该回归模型的残差是否存在单位根。如果残差存在单位根,则说明时间序列数据是非平稳的。

其中,在一个自回归过程中:

$$y_t = b_{t-1} + a + \epsilon_t \quad (2)$$

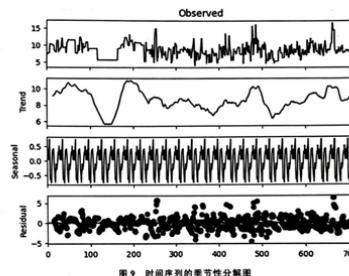
如果滞后项系数  $b$  为 1,就称为单位根,当单位根存在时,自变量和因变量之间的关系具有欺骗性,因为残差的任何误差都不会随着样本量(即时期数)增大而减小,也就是说模型中的残差的影响是永久的。

针对以上交通数据,通过 ADF 检验得到其 ADF 统计量的值是 -25.525488, p-value 值是 0.000000,因为 ADF 统计量在 1% 的显著性水平下,其他对值远大于 1,且 p-value 值远小于 5%,因此认为有足够的证据表明序列是平稳的。

## 2. 季节性分解

季节性分解将一个时间序列分解为三个主要的部分:趋势(Trend)、季节性(Seasonal)和残差(Residual)或称作不规则成分。这种分解有助于理解时间序列的结构和模式,从而可以进行更有效的预测和分析。

本文利用了每天 24 小时共 720 条的速度数据(使用道路长度数据除以道路平均旅行时间得到),经过季节性分解之后,生成了原始数据、趋势线、季节性成分、残差的示意图如图 9 所示。图解如下:



①趋势:趋势是指时间序列随时间推移而表现出的长期发展方向,从趋势线可以看出时间序列数据整体具有一定的周期性。

②季节性成分:季节性是指时间序列中重复出现的模式,反映出基于实时道路交通的时间序列数据具有明显的周期性,周期长度为一天。

③残差:残差是原始时间序列数据在去除趋势和季节性成分后的剩余部分。图中余留下的残差比较随机,表明趋势和季节性成分被准确估计,已经很好地从原始数据中分离出来。

以上分析可以推断得到经过处理的时间序列数据集具有很明确的周期性,故接下来的研究中可以通过原始的时间序列数据预测一天内的交通流量情况。

## 3. 自相关分析和偏自相关分析

后续的时序模型预测中,需要基于交通流量的时间序列数据来预测指定的时间范围内的交通流量,考虑到 ARIMA 模型这一传统时间序列回归模型,绘制了交通流量时间序列的自相关函数图和偏自相关函数图,如图 10。

自相关图：图中自相关系数在值为1后随着滞后的增加迅速衰减到0，故 ARIMA 模型中的  $q$  值可考虑取1。  
偏自相关图：图中偏自相关系数在值为2的滞后之后迅速衰减到0，故 ARIMA 模型中的  $p$  值可考虑取2。

再根据上述 ADF 检验，可知的时间序列数据是具有平稳性的，故 ARIMA 模型中的差分值可考虑取0。

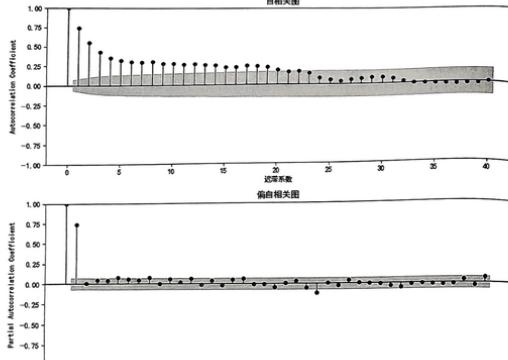
自相关图  


图 10 时间流量时间序列的自相关函数图和偏自相关函数图

#### (四) 时间序列聚类

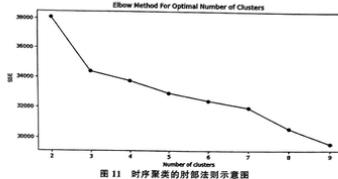
为将大量的时间序列数据分为具有相同模式或行为的时间序列类别，以方便进行某特定类的时间序列分析，本文对道路的平均车辆速度数据进行了时间序列聚类。

##### 1. 尖部法则

聚类度量：时序数据间欧几里得距离(Euclidean Distance)是一种非常常见的距离度量方式，尤其是在 K-Means 聚类算法中，欧几里得距离是最直观的距离度量方式，它测量的是多维空间中两点之间的直线距离。并且考虑到本文使用的时序数据集规模很大(包含七百万余条数据)，出于降低计算成本的考量，故选择使用欧几里得距离作为聚类度量，其数学描述如下：

肘部法则(Elbow Method)是一种用于确定聚类分析中最佳聚类数( $K$ )的启发式方法，尤其在 K-means 聚类中广泛使用。这种方法的核心思想是，随着聚类数  $K$  的增加，聚类的误差平方和(SSE，即簇内误差 Sum of Squared Errors)会逐渐减少，但是减少的速率会逐渐降低，直到达到一个点，之后每增加一个聚类，SSE 的减少量会骤减，形成一个“肘部”。

图 11 可以看到 SSE 值在聚类数为 3 处有一个肘部，但曲线斜率在此值处的变化并不很显著，故初步将聚类设置为三类进行可视化并观察实际聚类效果。



2. 时间序列聚类结果

观察所得的时序聚类效果图，如图 12，可以清晰地辨识出三种不同的类别特征，则该道路车辆速度数据的时间序列可以划分为三种不同情况。

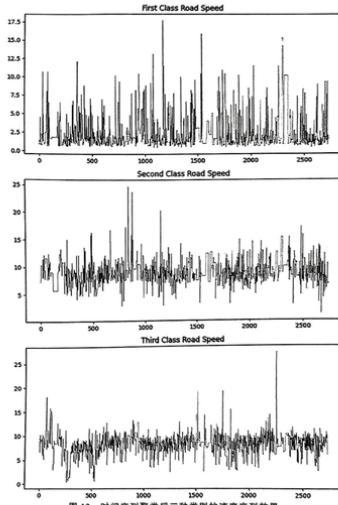


图 12 时间序列聚类后三种类别速度序列效果

### 三、基于道路实时数据的交通网络脆弱性分析

本文采用原始法来对道路网络进行抽象，即将现实中的道路网络中的交叉口抽象为拓扑网络结构中的节点，道路抽象为边。再根据交通网络的基本结构及特性，采用图论中的无向图来表示道路交通网络。其公式如下：

$$\begin{aligned} G &= (E, V) \\ V &= \{v \mid v \in \text{nodelement}\} \\ E &= \{e \mid e \in \text{edgeelement}\} \\ V &= \{v_i \mid v_i \in (e_{\text{out}}, e_{\text{in}})\} \end{aligned} \quad (3)$$

其中，曲线图 G 用于表示交通网络，V 是节点集合，E 是边集合。 $v_i \in (e_{\text{out}}, e_{\text{in}})$  表示两条边的上下游连接构成一个节点。

由于本文重点研究交通网络图中边的失效对于整体网路结构的影响，故节点仅具有“节点名称”属性——nodelement，而边具有“道路容量”“初始道路车辆数”“某时刻该道路平均速度”等属性——edgeelement。

#### (一) 交通拓扑网络构建

对于每一条道路，创建两个节点，分别代表道路的起始端和终止端。这两个节点通过一条边相连，边的权重和容量基于道路的长度和宽度计算得到。

为了简化网络结构，对网络中的节点进行了合并处理。即对于每一条道路，将其上游道路的终点端节点与该条道路的起始端节点合并，其下游道路的起始端与该条道路的终点端点合并。在这一过程中，特别注意避免产生自环，保留道路的原始空间特征。

然而，在道路网络构建过程中，发现有部分道路与主要道路离散，故选择剔除离散道路，只构建主要道路的网络拓扑图。

#### (二) 交通拓扑网络可视化

构建出的交通拓扑网络图如图 13 所示，图中较低的流量值用浅灰调表示，而较高的流量值则用黑色或深墨色表示，来直观地识别出道路交通流量的情况和潜在的交通问题。

从基于道路实时数据的交通网络模型图中可以观察到不同颜色的道路交会处的颜色变化，例如，灰色道路交会处呈现黑色，表明平均速度值较高的黑色道路的分支道路也会存在较高平均速度的路段。此外，浅灰色道路也会分支出黑色、灰色、浅灰色、深黑色道路，进一步证明了该网络模型能够反映现实世界的道路交通流量情况。

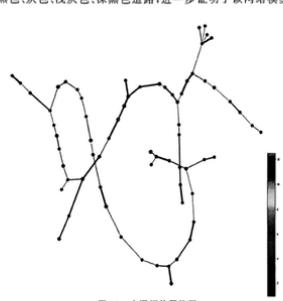


图 13 交通拓扑网络图

#### (三) 道路属性与车辆速度的相关性分析

由于调查道路本身属性与车辆速度的相关性，本文通过计算道路信息的三个关键属性—宽度、长度和速度的众数之间的相关系数矩阵对其进行相关性分析。为方便展示，对相关矩阵进行可视化，最终得到以下相关性系数图 14。

然而，图中并无明显黑色块，表明速度属性与道路的宽度和长度这些几何特征之间的相关性并不显著。由此可知本研究中道路的属性特征并不需要进行相关性处理。

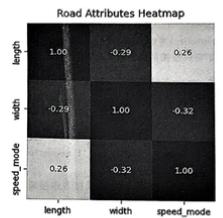


图 14 数据相关性矩阵图

然而，上图除了对角线中并无明显黑色块，表明速度属性与道路的宽度和长度这些几何特征之间的相关性并不显著。由此可知本研究中道路的属性特征并不需要进行相关性处理。

在建立的交通网络拓扑模型的基础上，本文通过搭建脆弱性度量体系从网络拓扑结构总体上评估该交通网络的静态脆弱性。

#### (四) 网络拓扑结构脆弱性度量

本文的交通网是由道路节点和道路边组合而成，当道路流量过高，超过道路本身的车辆容量后，该道路就会发生拥堵，若存在特性相同的道路来替代该失效的道路或该道路在该网络中连接的道路较少，则对网络脆弱性影响较小。而当道路发生拥堵后，不存在特性相同的道路来替代该失效的道路或该道路在该网络中连接的道路较少，则会影响网络的连通性。复杂网络理论中，拓扑结构度量是从连通性角度评估网络脆弱性的基本指标。本文选取常见的拓扑结构度量网络拓扑结构脆弱性度量。

##### 1. 度

度(Degree)是一个节点的相邻连接数量。对于每条边来说，它的度是连接该边的两个节点的度之和。边的度越高，表示连接该边的节点在网络中具有更多的相邻节点，因此该边在网络中的重要性更高。

计算公式如下：

$$\text{Degree}(e) = \sum_{n \in N(e)} A_n \quad (4)$$

其中， $n$  是边  $e$  的邻居节点集合， $A_n$  表示第  $n$  个与边  $e$  之间的连接权重

##### 2. 介数

介数(Betweenness)衡量了一条边在网络中作为最短路径的中介程度。如果一条边位于许多节点间的最短路径上，那么它的介数值就会很高，表示它在网络的信息传播中起到重要作用。

计算公式如下：

$$\text{Betweenness}(e) = \sum_{s \neq t \neq e} \frac{\sigma_{st}(e)}{\sigma_{st}} \quad (5)$$

其中， $\sigma_{st}$  是从起始边的上游节点  $s$  到目标下游节点  $t$  的最短路径数量， $\sigma_{st}(e)$  是从节点  $s$  到节点  $t$  的最短路

径中通过边 $e$ 的数量。

### 3. 接近度中心性

接近度中心性(Closeness Centrality)衡量了一条边到达其他节点的平均距离。较高的接近度中心性意味着一条边到达其他节点的平均距离较短,表示该边连接的节点更为集中。计算公式如下:

$$\text{Closeness}(e) = \frac{1}{\sum_{e' \in N(e)} d(e, e')} \quad (6)$$

其中,  $d(e, e')$  是边 $e$ 到边 $e'$ 的最短路径长度。

### 4. 特征向量中心性

特征向量中心性(Eigenvector Centrality)考虑了边连接到的节点的质量,即一个边的重要性不仅取决于它的两个节点连接的边数,还取决于它所连接的边的重要性。计算公式如下:

$$\text{Eigenvector}(e) = \frac{1}{\lambda} \sum_{e' \in N(e)} C_G(e') \quad (7)$$

$C_G(e')$  是边 $e'$ 的特征向量中心性,  $\lambda$  是最大的特征值,  $N(e)$  是边 $e$ 的节点相邻连接的边。

### (五)基于网络拓扑结构的网络脆弱性分析

计算得到该交通网络中每一条道路的拓扑特性指标值如表 5 所示:

通过对表中数据,以“Betweenness”“Closeness Centrality”“Eigenvector Centrality”这三个指标进行降序排序,得到重要度最高的边 road\_30,其具有如下特性:

发现 Degree 为 6,说明上下游各连接着两条道路,故这条路共连接着 5 条道路,表明该道路是一个连接节点较多的边。

介数(Betweenness)衡量的是边在网络中的桥梁作用,即边在连接网络中不同部分的能力。该道路的介数值为 0.507024,这个数值较高,表明这条边在网络中的许多最短路径上扮演了中介角色,对网络的连通性有较大影响。

表 5 交通网络中每一条道路的网络拓扑特性指标值

	Degree	Betweenness	Closeness Centrality	Eigenvector Centrality
road_0	4	0.112379	0.060514	0.045081
road_1	4	0.085601	0.057354	0.023777
road_2	4	0.454346	0.073075	1.74E-05
road_3	4	0.463565	0.074786	2.86E-05
road_4	5	0.138279	0.063919	0.084940
road_5	5	0.085601	0.041917	4.44E-05
road_6	5	0.112379	0.043580	5.5E-06
road_7	3	0.029412	0.041762	2.75E-06
road_8	4	0.491659	0.077550	0.001576
road_9	4	0.485953	0.076402	0.002842
road_10	5	0.294996	0.059631	0.350801
road_11	8	0.138279	0.057056	0.525314
road_12	5	0.163201	0.057102	0.346308
road_13	4	0.085601	0.049619	0.051588
road_14	4	0.057946	0.047408	0.026572
road_15	4	0.029412	0.051932	2.35E-06
road_16	6	0.029412	0.054202	0.424738
road_17	4	0.454346	0.071976	0.013662
road_18	4	0.444249	0.070242	0.024501
road_19	4	0.444249	0.071288	1.07E-05
road_20	4	0.112379	0.051966	0.097718

	Degree	Betweenness	Closeness Centrality	Eigenvector Centrality
road_21	4	0.313872	0.062003	0.199447
road_22	3	0.029412	0.045323	0.012543
road_23	5	0.479368	0.075118	0.005272
road_24	4	0.503512	0.079104	0.000123
road_25	4	0.505707	0.079573	0.000208
road_26	6	0.029412	0.054202	0.424738
road_27	4	0.433275	0.069443	6.69E-06
road_28	5	0.507024	0.079857	0.000368
road_29	4	0.029412	0.077034	0.000335
road_30	6	0.507024	0.079857	0.000563
road_31	6	0.029412	0.054202	0.424738
road_32	6	0.029412	0.054202	0.424738
road_33	4	0.506439	0.078456	7.99E-05
road_34	4	0.210711	0.046934	4.11E-06
road_35	4	0.233099	0.048673	3.35E-06
road_36	4	0.138279	0.054458	0.184120
road_37	5	0.313872	0.054179	2.78E-06
road_38	4	0.331870	0.056036	2.1E-06
road_39	4	0.365233	0.059837	1.58E-06
road_40	4	0.380597	0.061766	1.68E-06
road_41	4	0.057946	0.054202	0.012259
road_42	3	0.029412	0.051691	0.005787
road_43	4	0.433275	0.068449	0.045423
road_44	5	0.421422	0.066613	0.085040
road_45	5	0.275241	0.052320	3.08E-06
road_46	4	0.254609	0.050468	3.02E-06
road_47	5	0.187445	0.042555	5.4E-06
road_48	4	0.421422	0.067554	4.29E-06
road_49	5	0.471905	0.076405	4.78E-05
road_50	5	0.085601	0.074616	4.7E-05
road_51	5	0.496488	0.077839	6.29E-05
road_52	5	0.057946	0.076938	0.000495
road_53	3	0.029412	0.071594	0.000238
road_54	4	0.029412	0.071901	0.004856
road_55	5	0.057946	0.043525	4.98E-06
road_56	4	0.408692	0.065636	2.86E-06
road_57	4	0.395083	0.058703	2.06E-06
road_58	4	0.029412	0.040305	2.98E-06
road_59	4	0.348990	0.057924	1.71E-06
road_60	3	0.029412	0.065306	1.36E-05
road_61	5	0.506439	0.079386	0.000704
road_62	4	0.496488	0.078549	0.000950
road_63	5	0.331870	0.064447	0.130187
road_64	4	0.029412	0.040305	2.98E-06
road_65	4	0.057946	0.069724	2.66E-05
road_66	5	0.463565	0.073635	0.008443

### 五、基于 XGboost-ARIMA 加权融合模型的流量预测

基于上述时间序列分析结果,尽管道路的几何特征与行车速度的相关性不高,但时间序列分析揭示了交通流量数据在不同时间点之间存在显著的统计依赖性,故采用时间序列预测方法来对交通流量进行建模和预测,利用时间序列数据的自相关性和偏自相关性的分析以及 ADF 检验可知,该时序数据是一个平稳时间序列数据,适合用 ARIMA 模型来拟合。考虑到时间序列并不总是线性的,融合 XGBoost 模型,利用 ARIMA 处理数据的线性趋势,XGBoost 照顾到非线性因素,可以增强模型对非线性和非平稳时间序列的预测能力,并提高模型的鲁棒性、适应性。

#### (一)预测模型建立

##### 1. 预测时段的选择

在本研究中,专注于对贵州省道路车辆数据集中的早高峰时段(8:00 至 10:00)的车流量进行预测。该时段被选作研究对象,原因在于其交通流量的复杂性较高,且车流量集中,易于发生交通拥堵,这为研究道路级联失效现象提供了理想的条件。数据涵盖本身覆盖了 3 个月内每天 24 小时以 2 分钟为一个时间段的交通流量信息,提供了充足的数据来挖掘和分析交通流量的特征。

##### 2. ARIMA 模型的参数调整

在 ARIMA 模型中,有三个超参数需要给定:  $p$  (自回归部分的阶数),  $d$  (差分阶数),  $q$  (移动平均部分的阶数), 如图 15。



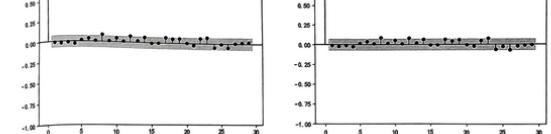
由于超参数的初始设置对于模型性能的影响较大,故用不同组超参数分别拟合的 ARIMA 模型并给出预测数据,然后使用皮尔逊相关系数计算预测值和原始的时间序列数据的实际情况来评估不同超参数设置的模型的性能,如表 6,拟合出的模型对应的相关系数可以看出在(2,0,1)时已经有了显著的增加,再增加阶数,模型预测表现也没有明显的增加,再结合前文偏自相关图——图中偏自相关系数在值为 2 的滞后之后迅速衰减到 0,而前文 ADF 平稳性检验——数据在不经过差分(即 0 阶差分)的情况下平稳性便已经足够高,同时为避免模型过拟合,选择(2,0,1)作为 ARIMA 模型的超参数。

表 6 不同超参数的性能评估表

	model_1	model_2	model_3	model_4	model_5	model_6
$p$	1	1	2	3	4	1
$d$	0	1	0	0	0	0
$q$	1	1	1	1	1	2
performance	0.740485	0.733301	0.740499	0.746341	0.746748	0.741253
	model_7	model_8	model_9	model_10	model_11	model_12
$p$	2	2	2	2	2	2
$d$	0	0	0	0	1	1
$q$	2	3	4	5	2	3
performance	0.746214	0.746476	0.747071	0.748253	0.736024	0.735995

图 16 是拟合了该模型之后的时间序列数据的自相关图和偏自相关图,因为拟合数据点展现了明显的“拖尾”效果,故可说明拟合后数据的具有明显的平稳序列特征。结合前文的各组参数组合对应的模型表现,则可充分证明 ARIMA 模型的( $p, d, q$ )参数选取(2,0,1)是合理的。

图 16 拟合了 ARIMA 模型之后的时间序列数据的自相关图和偏自相关图



#### 3. XGboost 模型的超参数调整

XGBoost 模型的超参数设置基于对交通数据周期性的理解,由于本文着眼于早高峰时段(8:00 至 10:00)的道路交通网的相关研究,故选择了 840 作为滞后的时间周期数(每隔 2 分钟采集一次数据,1 小时 30 分钟),则每天 8:00 至 10:00 采集  $2 \times 30 = 60$  条数据,2 星期采集  $2 \times 7 \times 60 = 840$  条数据。

#### 4. 加权融合模型的权重选择优化

为了更好的结合 XGBoost 和 ARIMA 模型的优点,将各模型的预测结果赋予权重并加以得到与真实值相比误差更小的预测结果。对于权重值的选择,则有以下步骤:

##### ①设置最优化目标函数:

$$\min \text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i \cdot \text{XGBoost}_i + x_2 \cdot \text{ARIMA}_i - \text{Truth}_i| \quad (8)$$

其中,  $i$  表示道路 id,  $\text{XGBoost}_i$  和  $\text{ARIMA}_i$  表示两模型在道路  $i$  的预测结果,  $x_1$  和  $x_2$  为模型结果赋予的权重值。

##### ②优化约束条件:

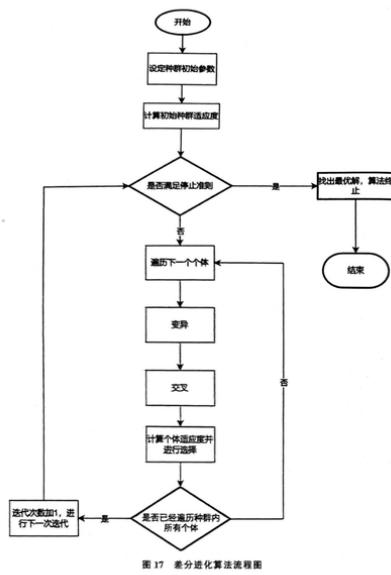
$$\begin{aligned} \text{不等式约束: } & 0 \leq x_1 \leq 1 \\ & 0 \leq x_2 \leq 1 \\ \text{等式约束: } & x_1 + x_2 = 1 \end{aligned} \quad (9)$$

##### ③优化方法:

差分进化算法的基本步骤如下:

首先初始化,生成一个初始种群,种群中的每个个体都是问题的潜在解,这里指的是一组权重  $x_1, x_2$ ;后变异,通过组合现有个体生成新的个体,即变异操作;再交叉,通过交叉变异后的个体与现有个体产生后代,后选择,在后代和现有个体之间选择适应度更好的个体,以形成下一代种群;最后迭代,重复变异、交叉和选择步骤,直到满足最大迭代次数或其他终止条件。此处设置种群大小 Pop 为 20,最大迭代次数 Malter 为 100,得到最优权值结果为:  $x_1 = 0.48050451, x_2 = 0.51947581$ 。

综上,关于使用差分进化算法来优化模型融合的权值的流程如图 17:



## (二) 预测模型评价

本研究使用平均绝对误差 (MAE)、均方误差 (MSE)、均方根误差 (RMSE)、平均绝对百分比误差 (MAPE) 来评估 XGBoost 模型、ARIMA 模型、XGBoost 与 ARIMA 加权融合模型的性能。公式如下：

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (12)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (13)$$

其中,  $n$  是预测的道路数量,  $y_i$  是第  $i$  个真实的道路平均速度,  $\hat{y}_i$  是第  $i$  个预测的道路平均速度。

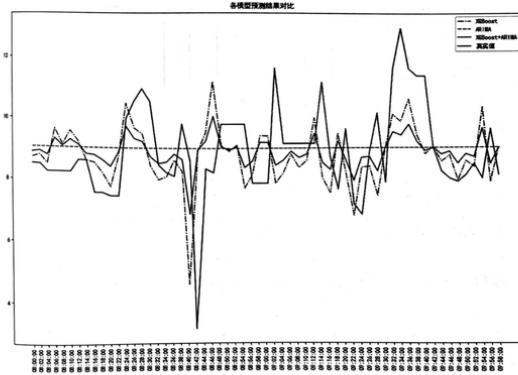
由表 7 中数据, 显然, XGBoost 与 ARIMA 加权融合模型的各项评估值都明显优于单独的模型。

表 7 XGBoost 模型、ARIMA 模型、XGBoost 与 ARIMA 融合模型的评价

	XGBoost	ARIMA	XGBoost+ARIMA
MAE	2.73823	2.411524	1.766125
MSE	11.83923	9.390131	4.955728
RMSE	3.440818	3.064332	2.226146
MAPE	1.104913	0.628383	0.319851

从图 18 中可以看出, 单一的 XGBoost 模型和 ARIMA 模型都不能很好的拟合真实数据, 但二者的加权融合则可以结合模型的优势——XGBoost 模型可以得到与真实数据较为一致的波动趋势但与真实数据在数值上相差较大, ARIMA 模型与真实数据相比在数值上较为贴合但无法拟合出足够优秀的波动趋势。

由此可以验证前文所言, “ARIMA 模型处理数据的线性趋势, XGBoost 照顾到非线性因素, 可以增强模型对非线性和非平稳时间序列的预测能力”, 故经加权融合的 XGBoost 与 ARIMA 模型可以得到更优的预测效果。



## 五、基于流量预测的交通网络动态脆弱性分析

在本研究的上文中, 构建了一个基于无向图的道路网络模型, 以模拟和分析交通系统中的级联失效现象。级联失效是指由于某些道路的失效导致邻近道路相继失效, 进而可能引发整个交通网络性能的下降。本文建立负载容量失效模型来模拟道路交通网络的级联失效。其建立流程如图 19, 级联失效示意图如

图 20。

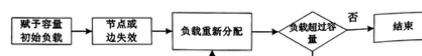
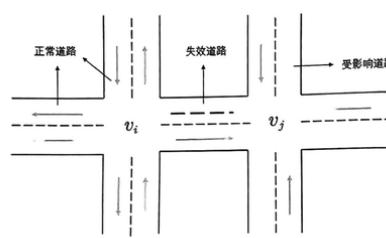


图 20 级联失效示意图



## (一) 道路级联失效的假设

假设 1: 道路失效的条件为当前道路边  $e(i)$  的车辆数量超过其承载容量, 边  $e(i)$  发生拥堵。

假设 2: 车辆在道路上的行驶涉及安全距离的考量, 即车辆与前车之间应保持足够的间隔以避免追尾事故, 故车辆占据的道路长度由车辆长度加上安全间距决定, 该间距与车辆的速度相关, 经查, GB7258—2017 机动车运行安全技术条件中对乘用车行车制动器制动力的要求中, 车辆的制动加速度  $a = 5.9m/s^2$ 。并考虑到人类的平均反应时间  $t$  约为 0.3 秒, 道路平均速度  $v$  根据给定时刻的实时预测数据得到, 则根据运动学公式, 前后两辆车之间的道路安全距离计算公式如下:

$$SafetyDistance = vt + \frac{v^2}{2a} \quad (14)$$

假设 3: 道路的车辆容量由道路的长度、车道数以及每辆车占据的道路长度共同决定,

$$RC_{i,0} = \frac{RL_{i,0} \times NL_{i,0}}{CL + SD}, i \in E \quad (15)$$

其中, 表示第  $i$  条道路的道路容量  $RoadCapacity$ ,  $RL_{i,0}$  是道路  $i$  的长度  $RoadLength$ ,  $NL_{i,0}$  是道路的车道数  $NumberLanes$ ,  $CL$  是车辆长度  $CarLength$ , 在此处设置  $CL = 4$  米,  $SD$  是基于假设 2 得到车辆行驶安全距离  $SafetyDistance$ 。

假设 4: 由于本文重点研究早高峰时段(6:30 至 10:00)交通网络结构, 在该时段的道路上车辆较多, 故设置每条道路的初始车辆数量为其实容量的 80%,

边  $e_{i,0}$  的初始车辆数  $OriRC_{i,0}$  与该道路边的容量是正比关系, 计算公式为:

$$OriRC_{i,0} = aRC_{i,0}, i \in E, (0 < a < 1) \quad (16)$$

其中,  $RC_{i,0}$  代表边  $e(i)$  的容量,  $a$  为控制道路初始车辆数的系数, 在本研究中为 80%, 则  $a = 0.8$ 。

假设 5: 道路的车辆流入/流出量在该道路的在时刻  $t$  的实时预测速度、每轮推演时间间隔、道路的车道数决定, 计算公式为

## (二) 道路网络失效的模拟推演

级联失效模拟过程中, 计算每个时间间隔  $Time$  车辆的流入和流出, 以此更新每条道路的车辆数量, 通过移除车辆数量超过其实容量的边来模拟道路失效, 该模式持续进行, 直到达到预设的模拟截止时间或图中没有更多的道路会发生失效为止。

应用上述假设和流量分配逻辑, 设定每轮推演的时间间隔  $Time = 2$  (分钟), 设定当前时刻为  $t = 8:30$ , 交通网中剩余可通行道路数随时间变化的模拟推演过程如下图所示。

由图 22 可观察到从 8:30 到 10:00, 交通网中剩余可通行道路数逐渐减少, 表明道路网络的失效在逐步扩大。并且, 从图中也可以直观看出, 从 9:08 开始到 9:38 这一段时间的道路失效数量急剧增加, 而在则相对平稳, 这也与现实生活中该时段早高峰多数道路会发生堵塞的实际情况相符合。

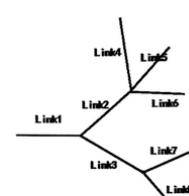


图 21 流量分配时道路连接示意图

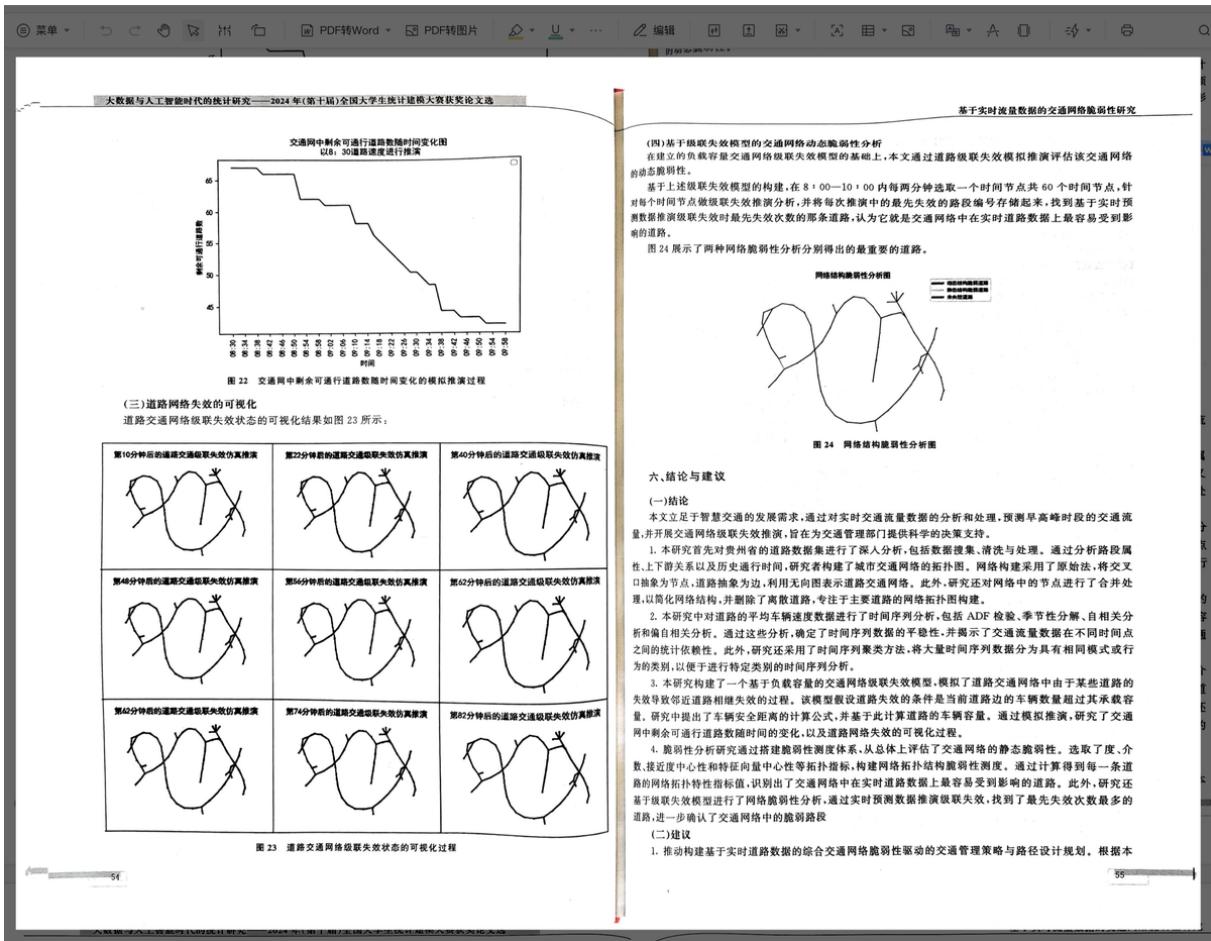


图 22 交通网中剩余可通行道路随时间变化的模拟推演过程

### (三) 道路网络失效的可视化

道路交通网络级联失效状态的可视化结果如图 23 所示：

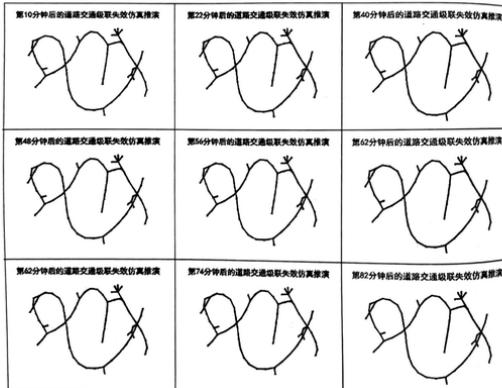


图 23 道路交通网络级联失效状态的可视化过程

## 基于实时流量数据的交通网络脆弱性研究

(四) 基于级联失效模型的交通网络动态脆弱性分析

在建立的负载容量交通网络级联失效模型的基础上,本文通过道路级联失效模拟推演评估该交通网络的动态脆弱性。

基于上述级联失效模型的构建,在 8:00—10:00 内每两分钟选取一个时间节点共 60 个时间节点,

针对每个时间节点做级联失效推演分析,并将其每次推演中的最先失效的路段编号存储起来,找到基于实时预测数据推演级联失效时最先失效次数的那条道路,认为它是交通网络中在实时道路数据上最容易受到影响的路。

图 24 展示了两种网络脆弱性分析分别得出最重要的道路。



图 24 网络结构脆弱性分析图

## 六、结论与建议

### (一) 结论

本文立足于智慧交通的发展需求,通过对实时交通流量数据的分析和处理,预测早高峰时段的交通流量,并开展交通网络级联失效推演,旨在为交通管理部门提供科学的决策支持。

1. 本研究首先对贵州省的道路数据集进行了深入分析,包括数据采集、清洗与处理。通过分析路段属性,上下游关系以及历史通行时间,研究者构建了城市交通网络的拓扑图。网络构建采用了原始法,将交叉口抽象为节点,道路抽象为边,利用无向图表示道路交通网络。此外,研究还对网络中的节点进行了合并处理,以简化网络结构,并剔除了离散道路,专注于主要道路的网络拓扑图构建。

2. 本研究中对道路的平均车辆速度数据进行了时间序列分析,包括 ADF 检验、季节性分解、自相关分析和偏相关分析。通过这些分析,确定了时间序列数据的平稳性,并揭示了交通流量数据在不同时间点之间的统计依赖性。此外,研究还采用了时间序列聚类方法,将大量时间序列数据分为具有相同模式或行为的类别,以便于进行特定类别的时间序列分析。

3. 本研究构建了一个基于承载容量的交通网络级联失效模型,模拟了道路交通网络中由于某些道路的失效导致邻近道路互相影响的过程。该模型假设道路失效的条件是当前道路边的车辆数量超过其承载容量。研究中提出了车辆安全距离的计算公式,并基于此计算道路的车辆容量,通过模拟推演,研究了交通网中剩余可通行道路随时间的变化,以及道路失效的可视化过程。

4. 脆弱性分析研究通过搭建脆弱性测度体系,从总体上评估了交通网络的静态脆弱性。选取了度、介数、接近度中心性和特征向量中心性等拓扑指标,构建网络拓扑结构脆弱性测度,通过计算得到每一条道路的网络拓扑特征指标值,识别出了交通网络中最容易受到影响的道路。此外,研究还基于级联失效模型进行了网络脆弱性分析,通过实时预测数据推演级联失效,找到了最先失效次数最多的道路,进一步确认了交通网络中的脆弱道路。

### (二) 建议

1. 推动构建基于实时道路数据的综合交通网络脆弱性驱动的交通管理策略与路径设计规划。根据本

## 基于实时流量数据的交通网络脆弱性研究

大数据与人工智能时代的统计研究——2024 年(第十四届)全国大学生统计建模大赛获奖论文选

文提出的研究方法和理论,在拥堵发生时鼓励使用网络中的冗余或备用路径,从而在关键路段发生失效时,能够快速重新分配流量,减少对网络的影响,而不会由于流量的转移造成进一步的拥堵。

2. 强化交通网络结构的鲁棒性,加强交通基础设施的弹性建设,改进交通的基础建设与配套设施,以减少因道路发生意外事故导致的交通中断时间。

3. 制定相应的应急处置预案分级制度,建立预案管理平台,实现预案的数字化存储,快速检索和实时更新,提高预案管理的效率。构建一个更加科学、高效、透明的应急管理体系,有效提升交通管理部门应对各种突发事件的能力。

### 【参考文献】

- [1] Barabási, A. L., Albert, R. Emergence of scaling in random networks[J]. Science, 1999, 286 (5439), 509–512.
- [2] Bono, F., Gutierrez, E. A network-based analysis of the impact of structural damage on urban accessibility following a disaster: the case of the seismically damaged Port Au Prince and Carrefour urban road networks[J]. Journal of Transport Geography, 2011, 19: 1445–1455.
- [3] Cao, L. J., Tay, F. E. H. Support vector machine with adaptive parameters in financial time series forecasting[J]. IEEE Transactions on Neural Networks, 2003, 14(6): 1506–1518.
- [4] Cats, O., Koppenol, G. J., Warner, M. Robustness assessment of link capacity reduction for complex networks: application for public transport systems[J]. Reliability Engineering and System Safety, 2017, 167: 544–553.
- [5] Duan, Y., Lu, F. Robustness of city road networks at different granularities[J]. Physica A, 2014, 411: 21–34.
- [6] He, Z., Navneet, K., Dam, W. V., et al. Robustness assessment of multimodal freight transport networks[J]. Reliability Engineering and System Safety, 2021, 207, 107315.
- [7] Jain, A., Kumar, A. M. Hybrid neural network models for hydrologic time series forecasting[J]. Applied Soft Computing, 2007, 7(2), 585–592.
- [8] Jakšić, T., Androćec, I., Sprčić, P. Electricity price forecasting-ARIMA model approach[C]. In: 2011 8th International Conference on the European Energy Market (EEM), 2011; 222–225.
- [9] Kim, K. J. Financial time series forecasting using support vector machines[J]. Neurocomputing, 2003, 55(1–2): 307–319.
- [10] Liu, H., Zhong, X., Ouyang, M., et al. Vulnerability analysis of public transit systems from the perspective of urban residential communities[J]. Reliability Engineering and System Safety, 2019, 189: 143–156.
- [11] Lu, Q. C., Zhang, L., Xu, P. C., et al. Modeling network vulnerability of urban rail transit under cascading failures: a Coupled Map Lattices approach[J]. Reliability Engineering and System Safety, 2022, 221.
- [12] Ma, J., Li, M., Li, H. J. Traffic dynamics on multilayer networks with different speeds[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2021, 69(3): 1697–1701.
- [13] Michael, A. P., Taylor, S. Remoteness and accessibility in the vulnerability analysis of regional road networks[J]. Transport Research Part A, 2012, 46: 761–771.
- [14] Peng, S. Disaster sprawling dynamics model for the cascading failure in urban road traffic network [D]. Changsha University of Science & Technology, Changsha, China, 2014.
- [15] Sagheer, A., Kotb, M. Time series forecasting of petroleum production using deep LSTM recurrent networks[J]. Neurocomputing, 2019, 323: 203–213.
- [16] Siami-Namini, S., Tavakoli, N., Namin, A. S. A comparison of ARIMA and LSTM in forecasting time series[C]. In: 2018 17th IEEE International Conference on Machine Learning and Applications
- [17] Spiliopoulou, A., Kontorinski, M., Papageorgiou, M., et al. Macroscopic traffic flow model validation at congested freeway off-ramp areas[J]. Transport Research Part C, 2014, 41: 18–29.
- [18] Tay, F. E., Cao, L. Application of support vector machines in financial time series forecasting[J]. Omega, 2001, 29(4): 309–317.
- [19] Tian, J., Fang, H., Wang, Y., Ren, C. On the degree correlation of urban road networks[J]. Transactions in GIS, 2018, 22: 119–148.
- [20] Wang, S., Chen, C., Zhang, J., et al. Vulnerability assessment of urban road traffic systems based on traffic flow[J]. International Journal of Critical Infrastructure Protection, 2022, 38: 100536.
- [21] Wang, S., Lv, W., Zhang, J., et al. Method of power network critical nodes identification and robustness enhancement based on a cooperative framework[J]. Reliability Engineering and System Safety, 2021, 207, 107313.
- [22] Wang, Z. W., Chen, X. L., Li, C. L. Identifying influence sources of cascading failure for road traffic network[J]. China Journal of Highway and Transport, 2015, 28(10): 98–104.
- [23] Wang, Z. W., Kuang, A. W., Wang, H. J. Calculation of traffic network node importance of cascading failure[J]. Highway Traffic Technology, 2012, 29(5): 96–101.
- [24] Wang, Z. W., Peng, S., Huang, Z. X. Disaster sprawling dynamics model for the cascading failure in urban road traffic network[J]. Journal of Safety and Environment, 2014, 14(3): 96–101.
- [25] Wu, J., Sun, H. J., Gao, Z. Y. Cascading failures on weighted urban traffic equilibrium networks [J]. Physica A: Statistical Mechanics & Its Applications, 2007, 386(1): 407–413.
- [26] Wu, J. J. Studies on the Complexity of Topology Structure in the Urban Traffic Network[D]. Beijing Jiaotong University, 2008.
- [27] Wu, J. J., Gao, Z. Y., Sun, H. J. Effects of the cascading failures on scale-free traffic networks[J]. Physica A: Statistical Mechanics & Its Applications, 2007, 378(2): 505–511.
- [28] Yunpeng, L., Di, H., Junpeng, B., Yong, Q. Multi-step ahead time series forecasting for different data patterns based on LSTM recurrent neural network[C]. In: 2017 14th Web Information Systems and Applications Conference (WISAC). 2017; 305–310.
- [29] Zeng, J., Xiong, Y., Liu, F., et al. Uncovering the spatiotemporal patterns of traffic congestion from large-scale trajectory data: a complex network approach[J]. Physica A, 2022, 604: 127871.
- [30] Zhang, G. P. An investigation of neural networks for linear time-series forecasting[J]. Computers & Operations Research, 2001, 28(12): 1183–1202.
- [31] Zhang, G. P. Time series forecasting using a hybrid ARIMA and neural network model[J]. Neurocomputing, 2003, 50, 159–175.
- [32] Zhang, G. P., Qi, M. Neural network forecasting for seasonal and trend time series[J]. European Journal of Operational Research, 2005, 160(2): 501–514.
- [33] Zhang, J., Wang, M. Transportation functionality vulnerability of urban rail transit networks based on moving block: the case of Nanjing metro[J]. Physica A, 2019, 535: 122367.
- [34] Zhang, L., Zeng, G., Li, D., Huang, H. J., Stanley, H. E., Havlin, S. Scale-free resilience of real traffic jams[J]. Proceedings of the National Academy of Sciences, 2019, 116: 8673–8678.
- [35] Zheng, J. F., Gao, Z. Y., Zhao, X. M. Modeling cascading failures in congested complex networks [J]. Physica A: Statistical Mechanics & Its Applications, 2007, 385(2): 700–706.
- [36] Zheng, Z., Huang, Z., Zhang, F., et al. Understanding coupling dynamics of public transportation networks[J]. EPJ Data Science, 2018, 7: 1–16.
- [37] Zhou, D., Hu, F., Wang, S., et al. Power network robustness analysis based on electrical engineering and complex network theory[J]. Physica A, 2021, 564: 125540.