Cartography M.Sc.

An Exploratory Study of Solving the Traffic Sensor Location Problem from the Perspective of Traffic Safety

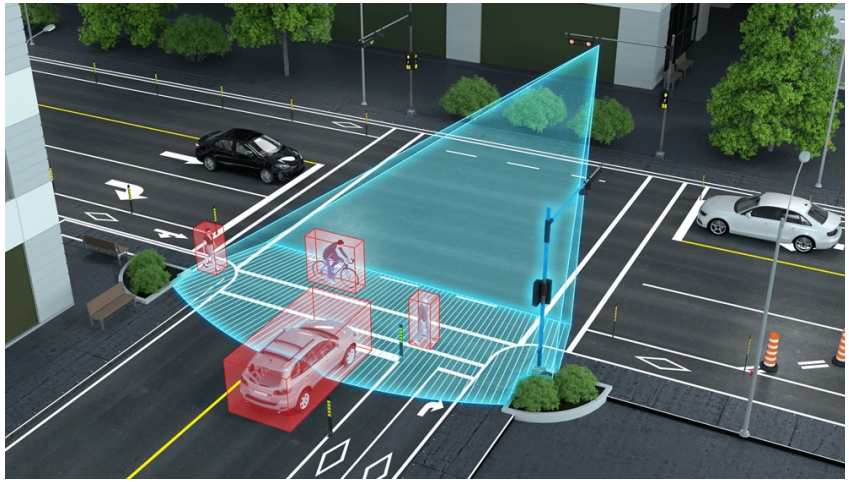
Zihan Liu

25.09.2023

Outline

1 Introduction and Motivation

- Nowadays, sensors with strong perception capabilities, like LiDAR and HD cameras, are being piloted as traffic sensors to make transportation smarter and more sustainable[1-3].
- These advanced traffic sensors can achieve accurate detection and tracking of pedestrians and vehicles and obtain trajectory-level data, which provides the possibility for their application in detecting traffic accidents and improving traffic safety [1-4].



An Example of Traffic Sensor with Strong Perception Capabilities (LeddarTech Inc., n.a.)

1 Introduction and Motivation

- Before traffic sensors can function in the field, their location strategy, namely "Where to Locate Traffic Sensors?" is a critical decision to be made. The related discussion is generalized as the Traffic Sensor Location Problem (TSLP)[5].
- The location strategy of traffic sensors is always determined by their functionality. Over time, TSLP has often been discussed from the perspective of traffic flow, while the discussion from the perspective of traffic safety has been rather rare[5-6].

Sensor Type	Current Mainstream Sensor	Strong Perception Sensor
Current Stage	mass deployment	start pilot
Sensor Function	traffic flow monitoring	traffic accident monitoring
Location Strategy	traffic flow-oriented TSLP	traffic safety-oriented TSLP
Relevant Study	persistent and extensive[5]	rare[13]

Comparison Table between Current Mainstream Traffic Sensors and Advanced Traffic Sensors and their Location Strategy (Liu et al., 2023)

1 Introduction and Motivation

- In this context, this thesis aims to conduct an exploratory study on dealing with the location strategy of traffic sensors in the application scenario of traffic accident monitoring.

"An Exploratory Study of Solving the Traffic Sensor Location Problem from the Perspective of Traffic Safety"

- Where to put traffic sensors? Put traffic sensors where traffic accidents or near accidents tend to happen.

In this scenario, TSLP is reduced to a problem of

traffic accident/near-accident hotspot detection

2 Research Objectives

- This thesis aims to extend the discussion on TSLP to the application scenario of traffic safety.
- Specifically, this thesis aims to explore the methods, data and workflow to detect historical traffic accident hotspots and predict traffic accident and near-accident hotspots.
- A case study is to be made to verify the effectiveness of the proposed workflow, methods, and data.

2 Research Objectives

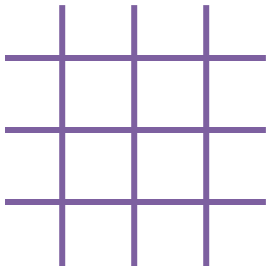
- Three methods are proposed in this thesis:

  Network Analysis: uses network analysis methods to detect the historical traffic accident hotspots on the road network.

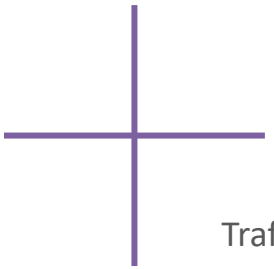
  Risk Analysis: uses machine learning methods to predict the traffic accident risk of traffic intersections.

  Rule Analysis: uses data mining methods to detect the association rules between geographic features and accident locations.

- Respectively, three research objects are as follows:



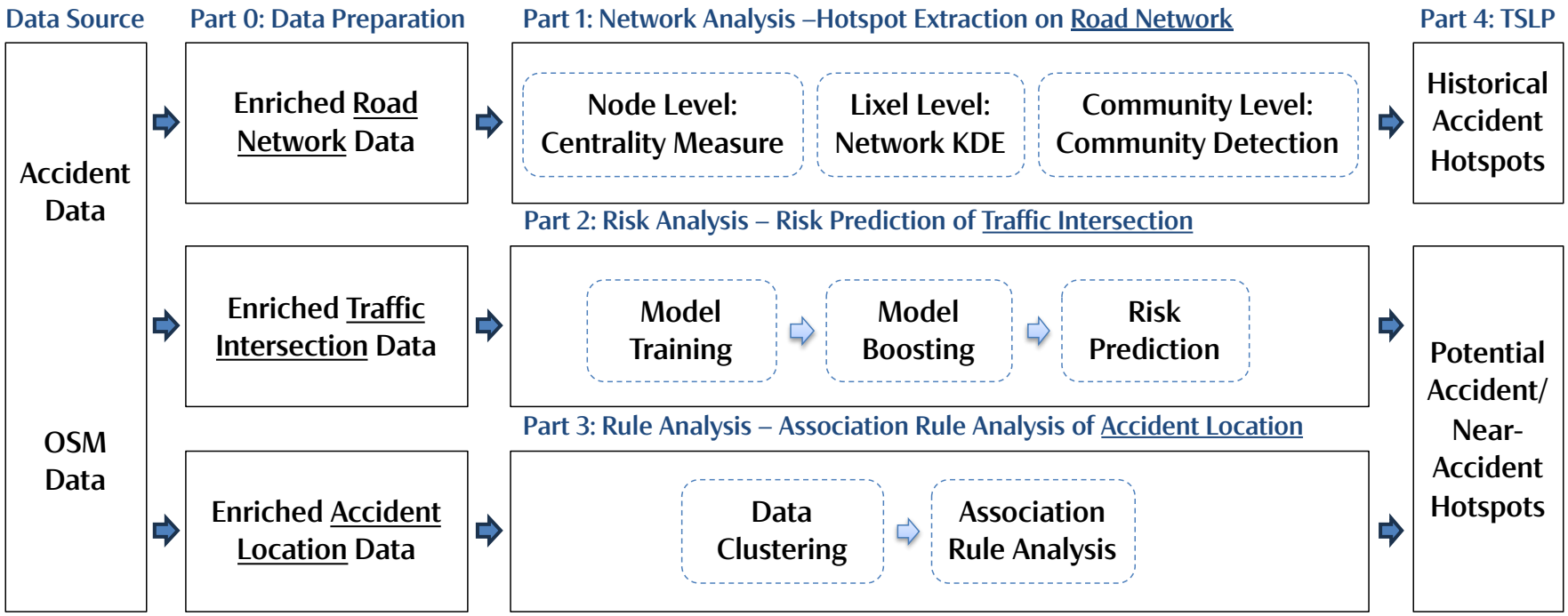
Road Network



Traffic Intersection

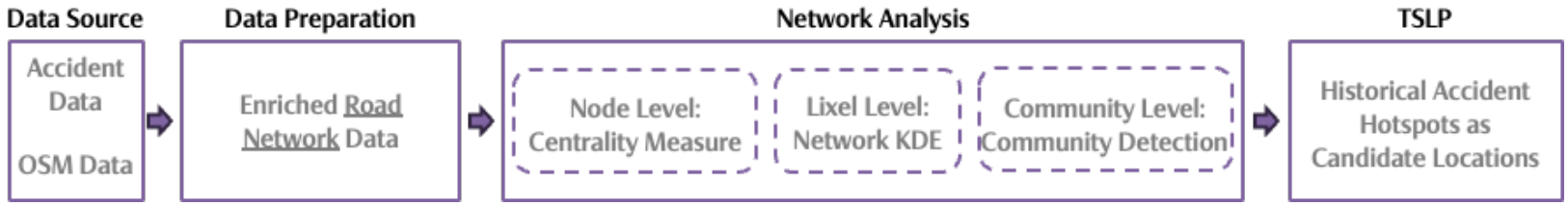
Accident Location

3 Methodology

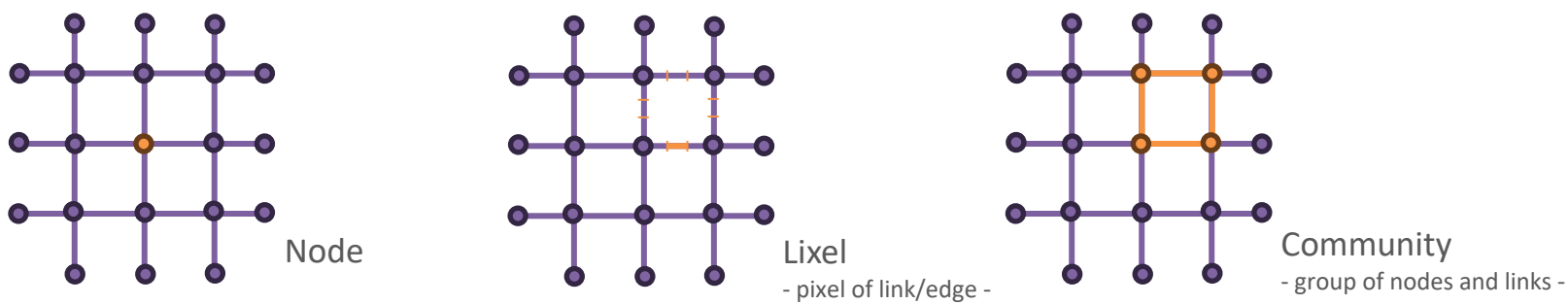


- Data Sources -> Data Preparation -> 3 Analysis Methods -> TSLP

3 Methodology



- Historical Hotspot Extraction on Road Network by Multiple Levels
  1) Node Level: centrality measures -> high accident centrality nodes
  2) Lixel Level: Network Kernel Density Estimation[7] (Network KDE) -> high kernel density lixels
  3) Community Level: community detection -> high accident risk communities



3 Methodology

- Risk Prediction of Traffic Intersections Based on Historical Data
  1) Data Preparation: intersection identification + data enrichment (accident data, nearby geographic feature data)
  2) Model Training: build the relationship between the accident risk and nearby geographic features of traffic intersections
  3) Model Boosting: improve the performance of the model with model boost methods
  4) Risk Prediction: use the model to predict the accident risk of traffic intersections

-> Predicted high-risk traffic intersections as candidate locations, including actual low-risk ones, which could be potential near-accident hotspots

3 Methodology

- Association Rule Analysis between Accident Locations and Nearby Geographic Features

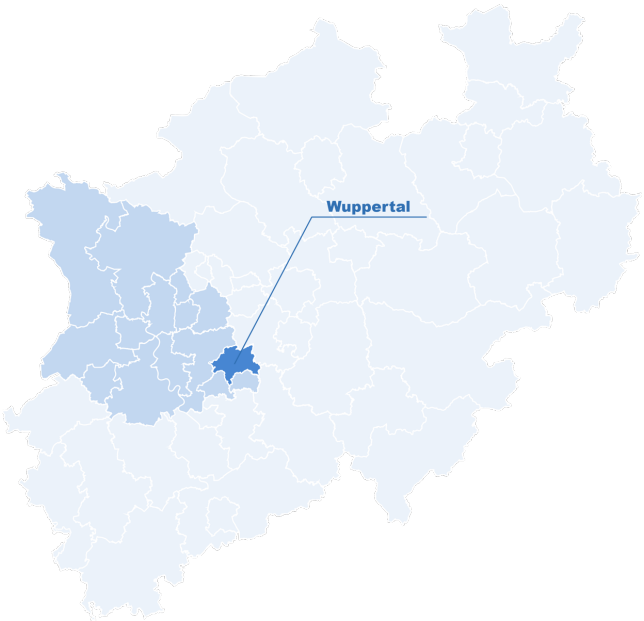
  1) Data Clustering: use data clustering methods or directly use their attributes to effectively cluster accident location data so that similar accidents can be grouped together.

  2) Association Rule Analysis: detect the association rules between one group of accidents and their nearby geographic features.

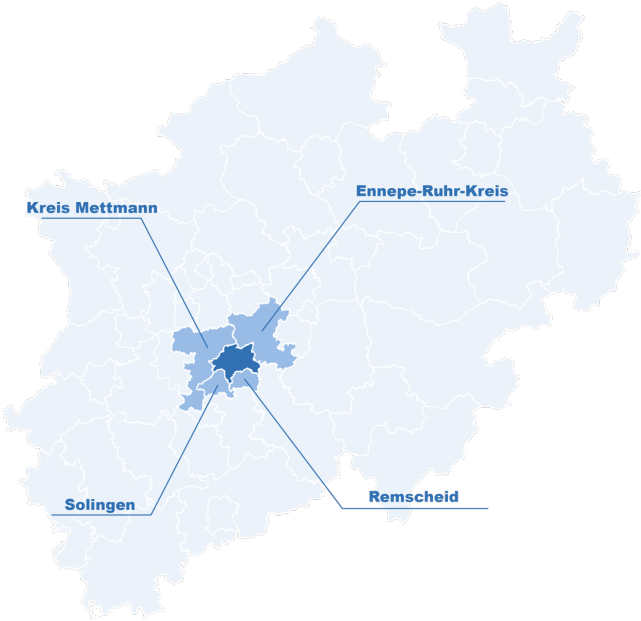
  -> The locations near accident-associated geographic features as candidate locations

4 Case Study

Study Area: Wuppertal, North Rhine-Westphalia, Germany



Location of Wuppertal in North Rhine-Westphalia (Basemap: OSM)



Location of Nearby Four Cities/Counties for Model Training ((Basemap: OSM)

4 Case Study

Data Source 1: German Accident Atlas

- An official, open and up-to-date data source published by the German Federal Statistical Office and updated annually.
- It records all the road traffic accidents that happened in Germany due to vehicular traffic on public roads or places.
- Accident locations, time, types, participants and other attributes (24 in total) are fully and well recorded.

Column Name	Content
ID	Serial Number
ULAND	Land
UREGBEZ	Administrative Region
UKREIS	Administrative District
UGEMEINDE	Municipality
UJAHR	Year of Accident
UMONAT	Month of Accident
USTUNDE	Hour of Accident
UWOCHENTAG	Day of the Week
UKATEGORIE	Road Accidents Involving Personal Injury
UART	Kinds of Accidents
UTYP1	Type of Accidents
ULICHTVERH	Light Conditions
IstRad	Accident with Bicycle
IstPKW	Accident with Passenger Car
IstFuss	Accident with Passenger
IstKrad	Accident with Motorcycle
IstGkfz	Accident with Goods Road Vehicle
IstSonstige	Accident with Other
USTRZUSTAND	Road Surface Conditions
LINREFX	Coordinates of the Place of Accident
LINREFY	(ETRS 89 / UTM Zone 32N)
XGCSWGS84	Coordinates of the Place of Accident
YGCSWGS84	(WGS84)

The Attribute List of German Accident Atlas (German Accident Atlas, 2021)

Data Source 2: OpenStreetMap (OSM)

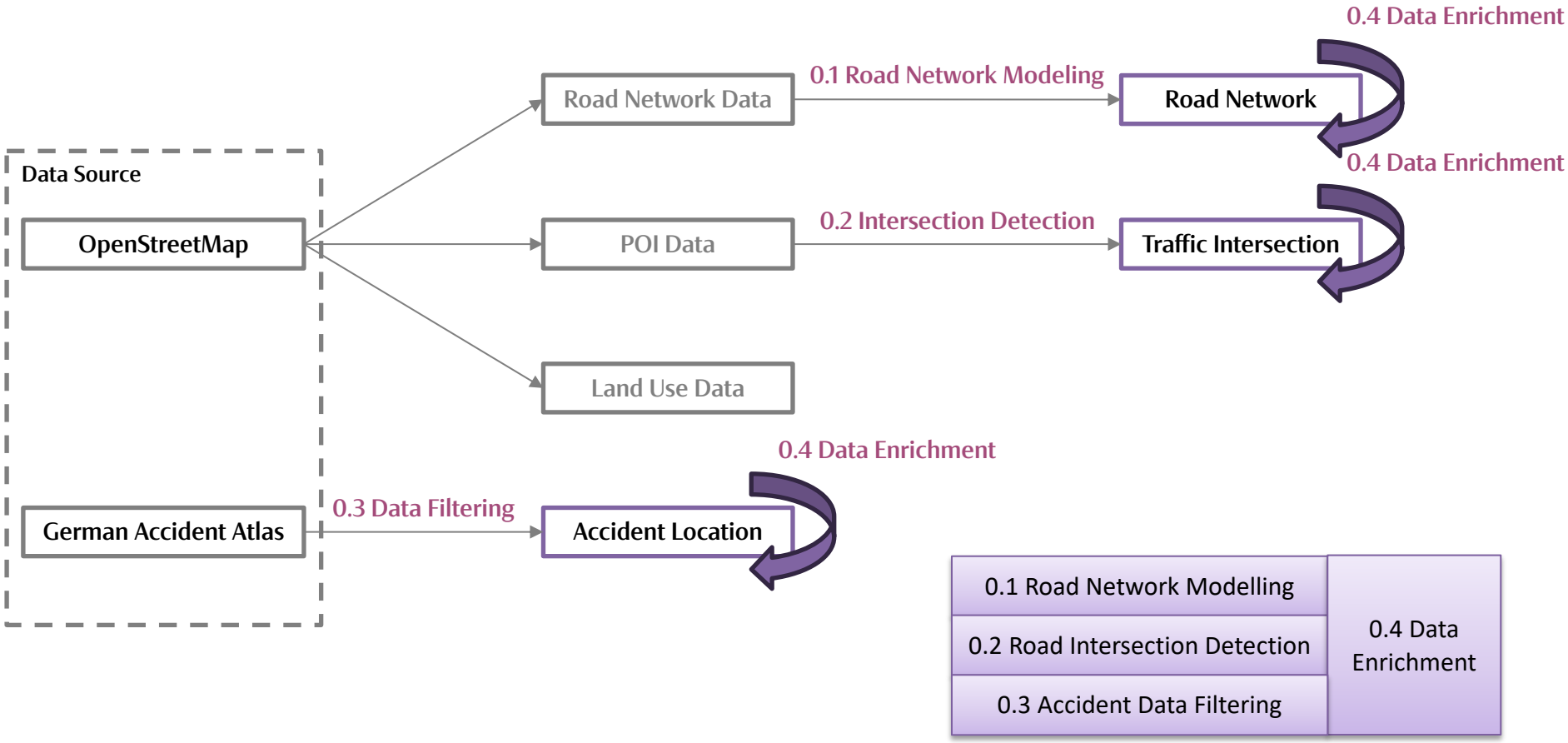
- A free, open, and up-to-date geographic database updated and maintained by a community of volunteers via open collaboration.
- From the OSM database, road network data, POI data, land use data and other data of the study area are extracted for further analysis.
- In total, 5 categories and 61 items are extracted.

Category	OSM Tag	Number
road network	*highway*	13
traffic facilities	*highway*	8
facilities	*amenity*	8
POI	*amenity*	6
land use	*landuse*	26
In Total: 61		

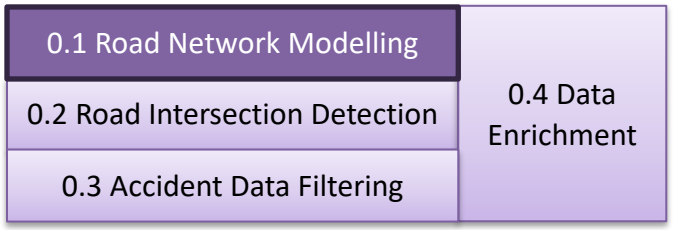
Extracted Data from OSM

4 Case Study

Part 0: Data Preparation

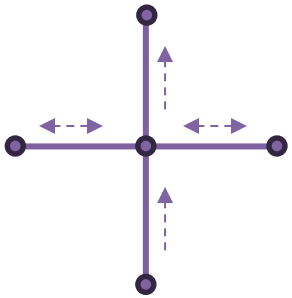


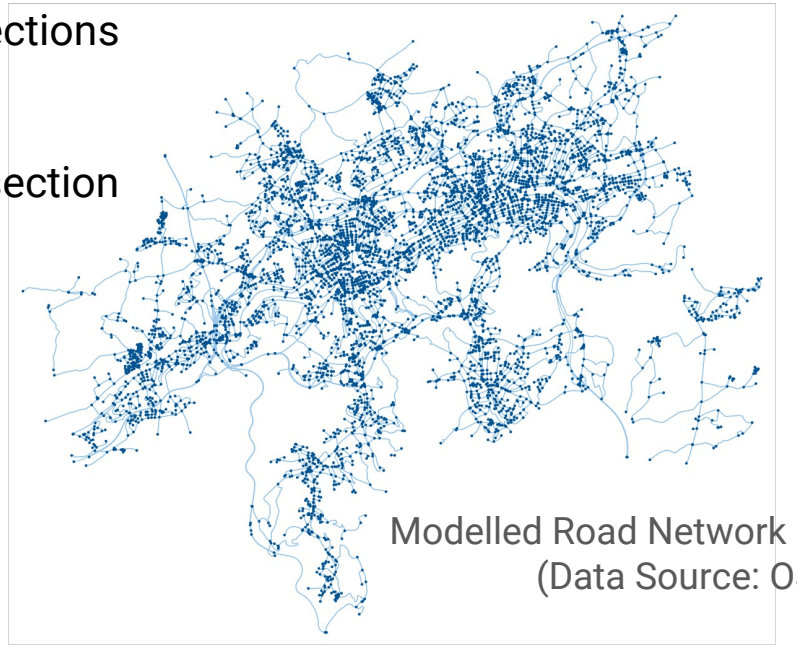
4 Case Study



Part 0: Data Preparation

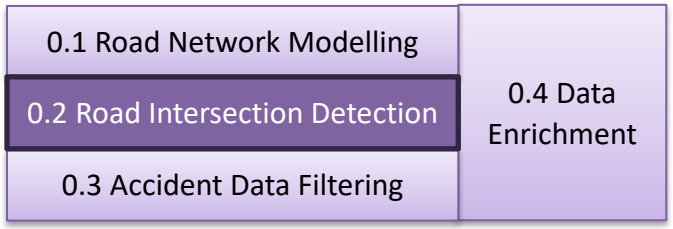
- All drivable public roads, except service roads, are modelled based on OSM.
- Road segments as edges, their directions are kept, road intersections and ends as nodes.
- Interstitial nodes (non-intersection nodes) are removed.





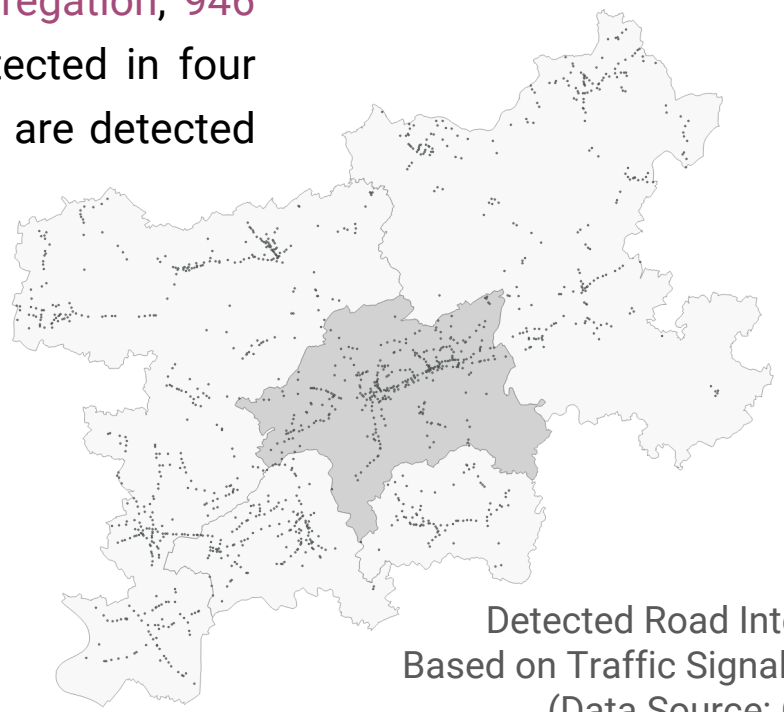
Modelled Road Network of Wuppertal
(Data Source: OSM)

4 Case Study



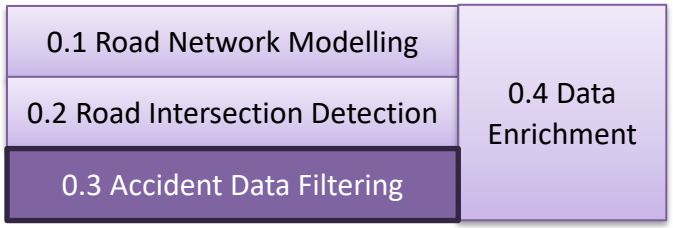
Part 0: Data Preparation

- There is no direct traffic intersection data.
- Based on traffic signal aggregation, 946 traffic intersections are detected in four nearby cities, and 345 ones are detected in Wuppertal.



Detected Road Intersection
Based on Traffic Signals Aggregation
(Data Source: OSM)

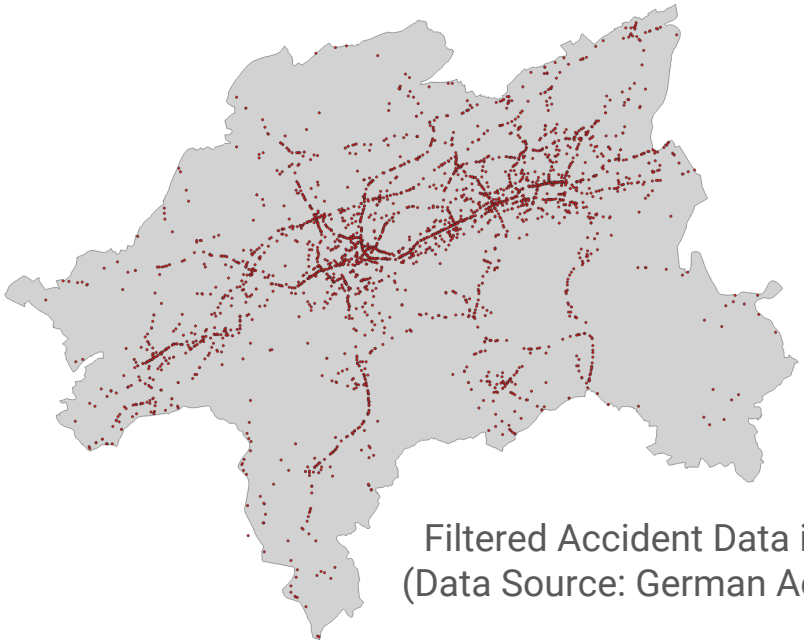
4 Case Study



Part 0: Data Preparation

- In total, 3396 traffic accidents happened in Wuppertal between the Years 2019 and 2022 are filtered from the German Accident Atlas.

Year	Number
2019	910
2020	770
2021	814
2022	902
In Total	3396



Filtered Accident Data in Wuppertal
(Data Source: German Accident Atlas)

4 Case Study

Part 0: Data Preparation

0.1 Road Network Modelling	0.4 Data Enrichment
0.2 Road Intersection Detection	
0.3 Accident Data Filtering	

- Data Enrichment: using the two data sources to enrich each study object for further analysis

1) Road Network: enriched with <u>accident linear density</u>
2) Road Intersection: enriched with <u>accident number</u> and nearby <u>geographic features number</u>
3) Accident Data: enriched with nearby <u>geographic features</u>

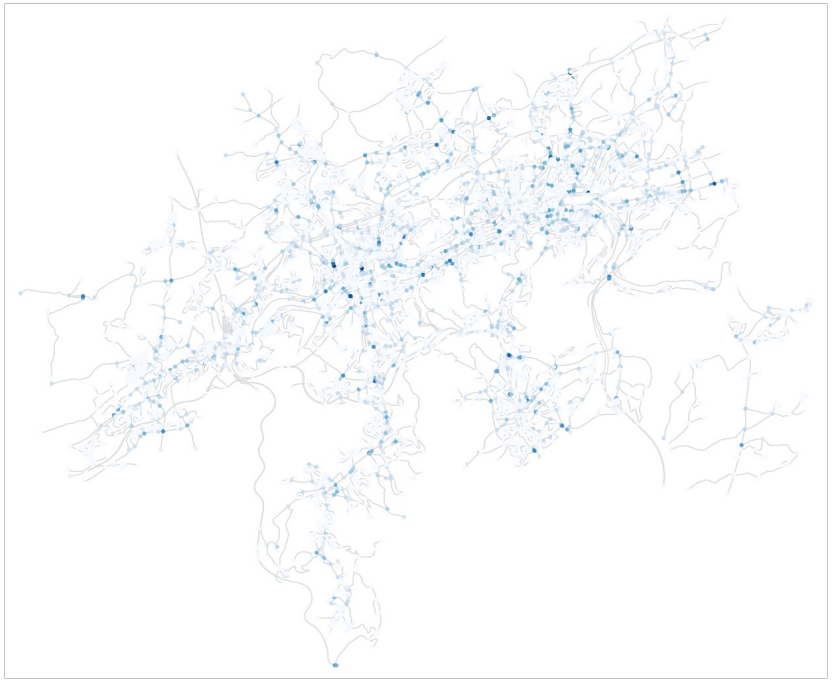
Attribute Name	Content	Category
nearbyNodes	number of nearby road network nodes	1
nearbyEdges	number of nearby road network edges	1
nearbyEdgeClass	number of nearby road network edge classes (i.e. road classes)	1
nearMotorway	whether near motorway or motorway_link road	1
nearTrunk	whether near trunk or trunk_link road	1
nearPrimary	whether near primary or primary_link road	1
nearSecondary	whether near secondary or secondary_link road	1
nearTertiary	whether near tertiary or tertiary_link road	1
nearUnclassified	whether near unclassified road	1
nearResidential	whether near residential road	1
nearLivingStreet	whether near living_street road	1
nearbyBusStop	number of nearby bus stops	2
nearbyCrossing	number of nearby crossings	2
nearbyGiveWay	number of nearby give way signs	2
nearbyMotorwayJunction	number of nearby motorway junctions	2
nearbySpeedCamera	number of nearby speed cameras	2
nearbyStop	number of nearby stop signs	2
nearbyTrafficSignals	number of nearby traffic signals	2
nearbyTurningCircle	number of nearby turning circles	2
nearbySustenance	number of nearby amenities of sustenance	3
nearbyEducation	number of nearby amenities of education	3
nearbyTransportation	number of nearby amenities of transportation	3
nearbyFinancial	number of nearby amenities of finance	3
nearbyHealthcare	number of nearby amenities of healthcare	3
nearbyEntertainment	number of nearby amenities of entertainment, arts & culture	3
nearbyPublicservice	number of nearby amenities of public service	3
nearbyFacilities	number of nearby amenities of facilities	3
nearbyKindergarten	number of nearby kindergartens	3
nearbySchool	number of nearby schools	3
nearbyDoctors	number of nearby doctors	3
nearbyHospital	number of nearby hospitals	3
nearbySocialfacility	number of nearby social facilities	3
nearbyParking	number of nearby car parks	3
nearAllotments	whether near land use type of allotments	4
nearAnimalKeeping	whether near land use type of animal keeping	4
nearBasin	whether near land use type of basin	4
nearBrownfield	whether near land use type of brownfield	4
nearCemetery	whether near land use type of cemetery	4
nearCommercial	whether near land use type of commercial	4
nearConstruction	whether near land use type of construction	4
nearEducation	whether near land use type of education	4
nearFarmland	whether near land use type of farmland	4
nearFarmyard	whether near land use type of farmyard	4
nearFlowerbed	whether near land use type of flowerbed	4
nearForest	whether near land use type of forest	4
nearGarages	whether near land use type of garages	4
nearGrass	whether near land use type of grass	4
nearGreenfield	whether near land use type of greenfield	4
nearGreenhouseHorticulture	whether near land use type of greenhouse horticulture	4
nearIndustrial	whether near land use type of industrial	4
nearMeadow	whether near land use type of meadow	4
nearOrchard	whether near land use type of orchard	4
nearPlantNursery	whether near land use type of plant nursery	4
nearRailway	whether near land use type of railway	4
nearRecreationGround	whether near land use type of recreation ground	4
nearReligious	whether near land use type of religious	4
nearResidentialLU	whether near land use type of residential	4
nearRetail	whether near land use type of retail	4
nearTrafficIsland	whether near land use type of traffic island	4
nearVillageGreen	whether near land use type of village green	4
isAccident	whether there is accident	5

Attribute Table of Traffic Intersections after Data Enrichment (5 classes, 61 items)

4 Case Study

Part 1: Network Analysis – Node Level

- PageRank Centrality: computes a ranking of the nodes in the graph based on the number and weight (accident linear density) of incoming edges.
- Nodes connecting more roads and more high accident linear density roads get higher centrality -> candidate locations.

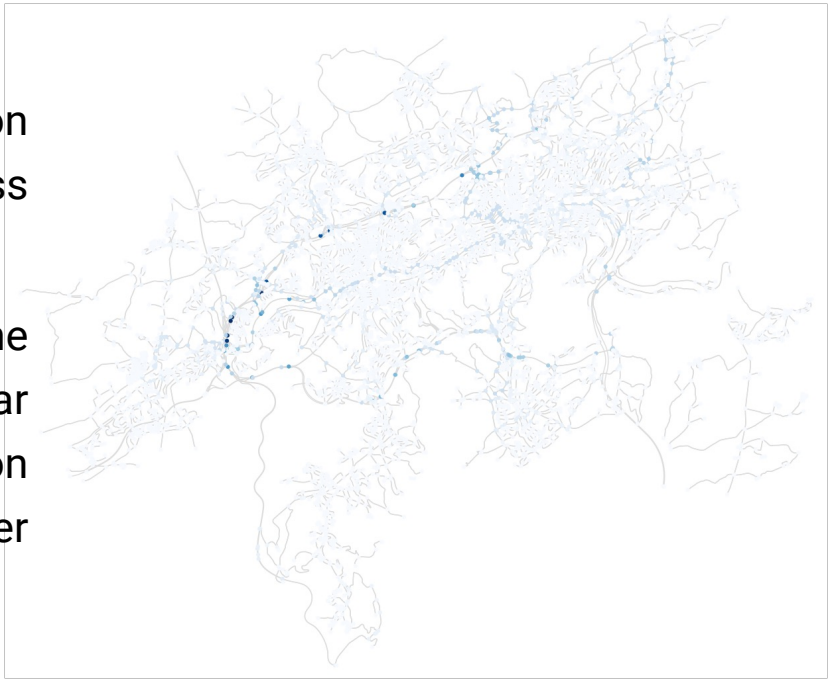


PageRank Centrality Result

4 Case Study

Part 1: Network Analysis – Node Level

- Betweenness Centrality: nodes lying on more shortest paths get higher betweenness centrality.
- With inversed accident linear density as the cost of each edge, edges with higher linear density get lower cost, therefore lying on more shortest paths and getting higher centrality -> candidate locations.

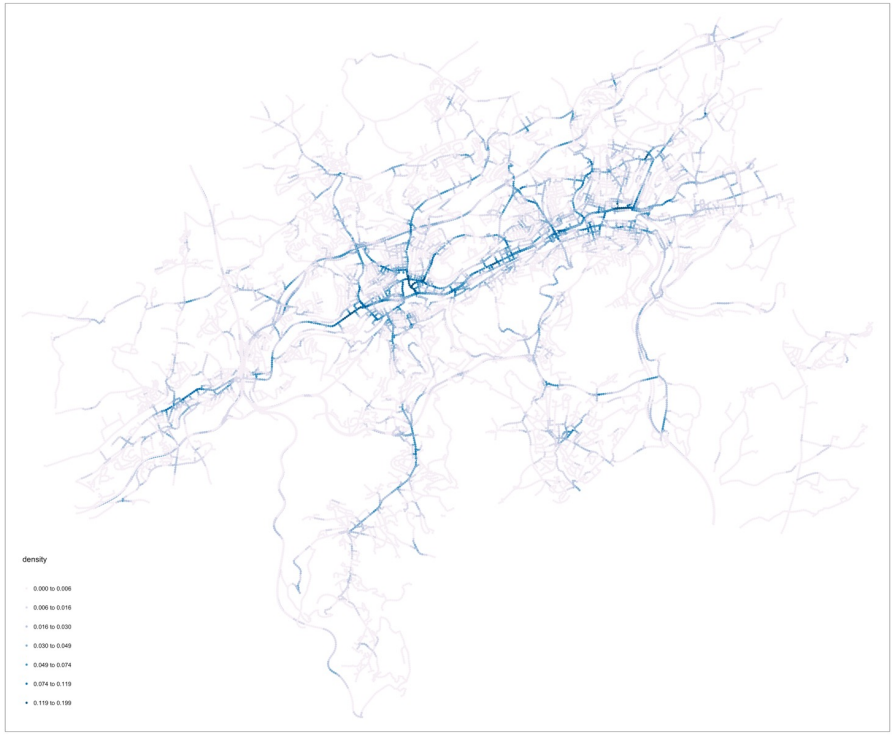


Betweenness Centrality Result

4 Case Study

Part 1: Network Analysis – Lixel Level

- Network KDE is the updated version of classical KDE for the density of events occurring in a network space.
- The kernel density of the nearby traffic accidents of each lixel (50m long) is calculated.
- Lixels with more nearby traffic accidents get higher kernel density -> candidate locations.

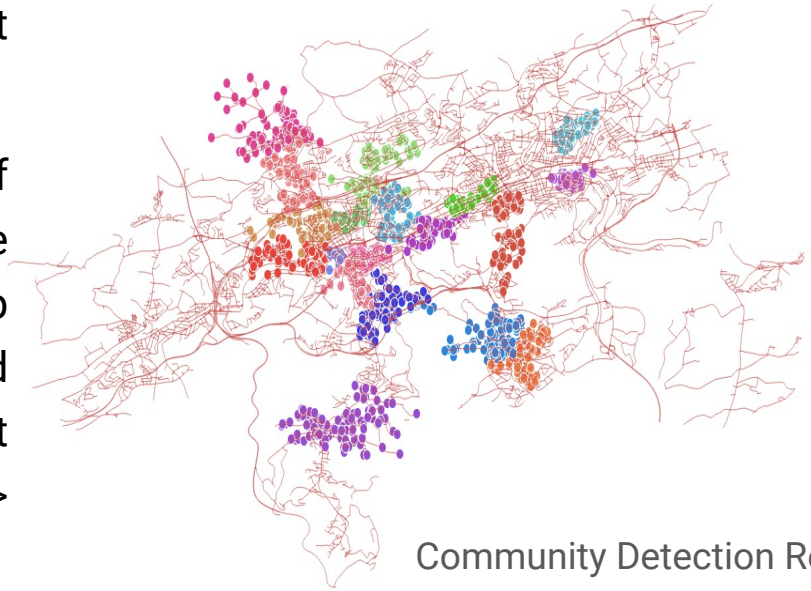


Network KDE Result

4 Case Study

Part 1: Network Analysis – Community Level

- Louvain Algorithm[8] is applied to detect the high accident risk community.
- The higher the accident linear density of one edge, the larger the weight of the edge, thus the stronger the relationship between two nodes of the edge. And community detection method can detect those nodes with strong relationships -> candidate locations.
- Limitation of methods -> to be discussed later

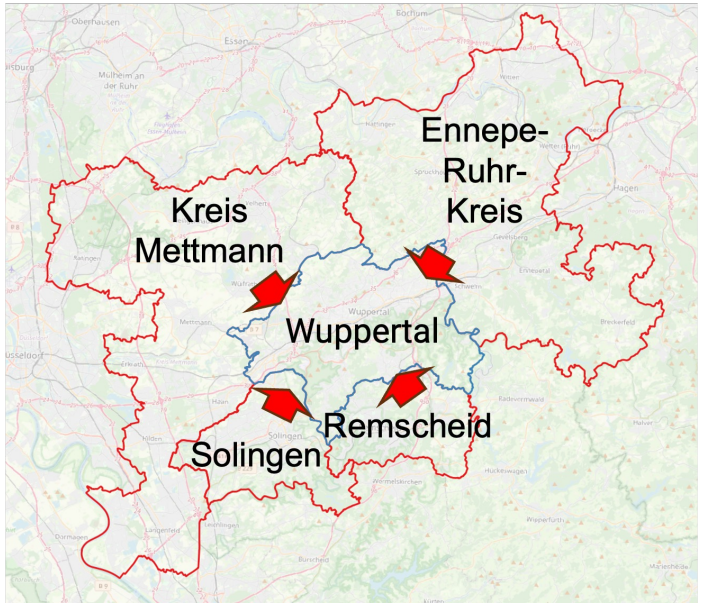


Community Detection Result

4 Case Study

Part 2: Risk Analysis – Model Training

- Train the model with the data of nearby four cities, and predict the risk of traffic intersections in Wuppertal.
- Model: Random Forest Classifier
- Features: geographic features of each traffic intersection
- Label: traffic accident risk high or low (<= 1 accident/year)



Schematic Graph of Training/Testing Areas (Basemap: OSM)

	Description	Number	Features	Labels
Training Set	four nearby cities/counties	946	946 * 60	946 * 1
Testing Set	Wuppertal	345	345 * 60	345 * 1

Division of Training Set and Testing Set

4 Case Study

Part 2: Risk Analysis – Model Training

Category	Items	No
Intersection	nearbyNodes, nearbyEdges, nearbyEdgeClass	3
Nearby Road Network	nearMotorway, nearTrunk, nearPrimary, nearSecondary nearTertiary, nearUnclassified, nearResidential, nearLivingStreet	8
Nearby Traffic Facility	nearbyBusStop, nearbyCrossing, nearbyGiveWay, nearbyMotorwayJunction, nearbySpeedCamera, nearbyStop, nearbyTrafficSignals, nearbyTurningCircle	8
Nearby Amenity	nearbySustenance, nearbyEducation, nearbyTransportation, nearbyFinancial, nearbyHealthcare, nearbyEntertainment, nearbyPublicservice, nearbyFacilities, nearbyKindergarten, nearbySchool, nearbyDoctors, nearbyHospital, nearbySocialfacility, nearbyParking	14
Nearby Landuse	nearAllotments, nearAnimalKeeping, nearBasin, nearBrownfield, nearCemetery, nearCommercial, nearConstruction, nearEducation, nearFarmland, nearFarmyard, nearFlowerbed, nearForest, nearGarages, nearGrass, nearGreenfield, nearGreenhouseHorticulture, nearIndustrial, nearMeadow, nearOrchard, nearPlantNursery, nearRailway, nearRecreationGround, nearReligious, nearResidentialLU, nearRetail, nearTrafficIsland, nearVillageGreen	27

Detailed List of Features

- Features: nearby geographic features of traffic intersections – 5 categories, 60 items

4 Case Study

Part 2: Risk Analysis – Risk Prediction

		Prediction		
		Low-Risk	High-Risk	
Fact	Low-Risk	155	22	87.6%
	High-Risk	73	95	56.5%
				72.5%

Confusion Matrix of Prediction Results

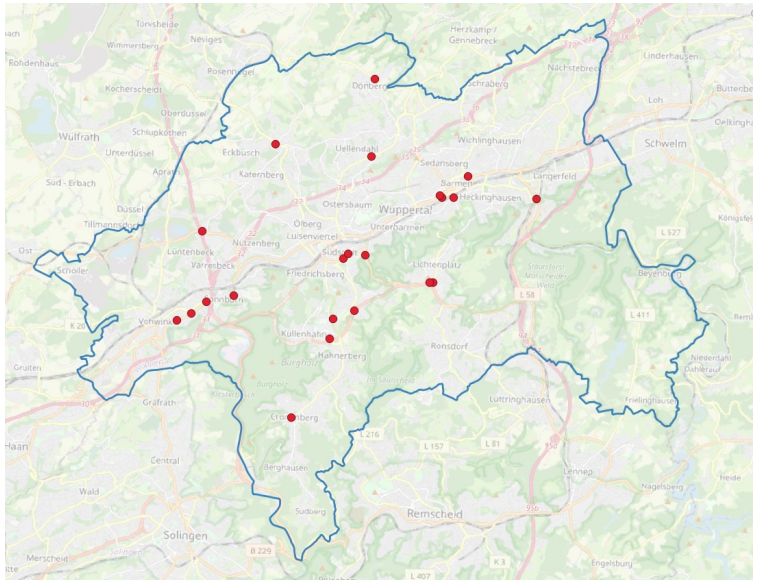
- The prediction accuracy is high for low-risk intersections, relatively low for high-risk intersections

4 Case Study

Part 2: Risk Analysis – Risk Prediction

		Prediction	
		Low-Risk	High-Risk
Fact	Low-Risk	155	22
	High-Risk	73	95

Confusion Matrix of Prediction Results



Geographic Distribution of Predicted Potential Near-Accident Hotspots (Basemap: OSM)

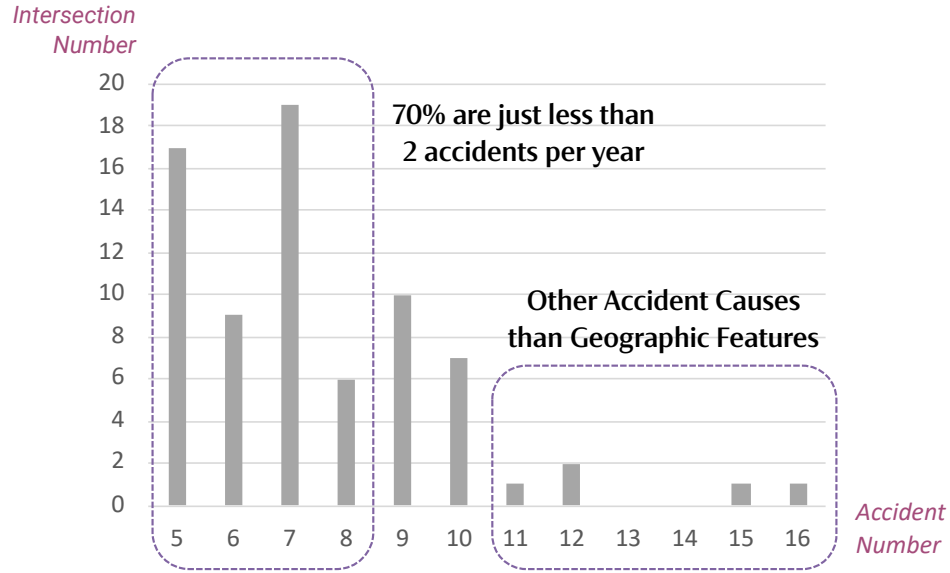
- Those actual low-risk but predicted high-risk ones can be recognized as potential near-accident hotspots.

4 Case Study

Part 2: Risk Analysis – Risk Prediction

		Prediction	
		Low-Risk	High-Risk
Fact	Low-Risk	155	22
	High-Risk	73	95

Confusion Matrix of Prediction Results

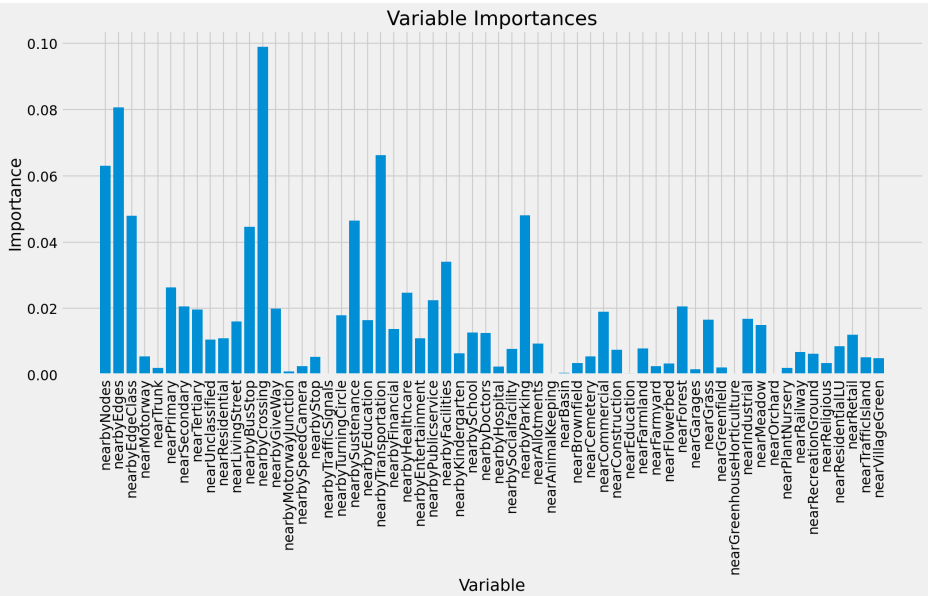


Histogram of Predicted Low-Risk but Actual High-Risk Traffic Intersections

- Those actual high-risk but predicted low-risk ones concentrate on the slightly high-risk part of the histogram.

4 Case Study

Part 2: Risk Analysis – Risk Prediction



Feature	Importance	Description
nearbyCrossing	0.10	number of crossing
nearbyEdges	0.08	number of road segment
nearbyNodes	0.06	number of intersection
nearbyTransportation	0.06	number of transportation amenities
nearbyEdgeClass	0.05	number of road class
nearbyParking	0.05	number of parking lots
nearbyBusStop	0.04	number of bus stops

Importances of Geographic Features from Trained Model

- Importance results are optimistic, similar to human cognition -> model should be reliable

Part 3: Rule Analysis – Data Clustering

- Method 1: cluster accidents by their attributes.
- Method 2: community detection on the accident similarity network[9][10]: accidents as nodes, links built between similar accidents (sharing more than a certain number of identical attributes), and similarity as the weight of links.

Column Name	Content
UJAHR	Year of Accident
UMONAT	Month of Accident
USTUNDE	Hour of Accident
UWOCHENTAG	Day of the Week
UKATEGORIE	Road Accidents Involving Personal Injury
UART	Kinds of Accidents
UTYP1	Type of Accidents
ULICHTVERH	Light Conditions
IstRad	Accident with Bicycle
IstPKW	Accident with Passenger Car
IstFuss	Accident with Passenger
IstKrad	Accident with Motorcycle
IstGkfz	Accident with Goods Road Vehicle
IstSonstige	Accident with Other
USTRZUSTAND	Road Surface Conditions

List of Attributes Used for Data Clustering
(German Accident Atlas, 2021)

Part 3: Rule Analysis – Data Clustering

- Carry out community detection on the accident similarity network, and the Louvain algorithm is used to extract the internally similar clusters of accident data.
- 6 communities are detected, but the separation between communities is blurred.

Extracted Communities

Part 3: Rule Analysis – Association Rule Analysis

- Association Rules[11][12]: *if X, then Y*

  e.g. if there is an accident, then geographic feature Y is nearby

- Common Indicator:

  support - proportion of X and Y occurring at the same time

  e.g. proportion of accidents that near geographic feature X

- Limitation of dataset -> to be discussed later

Part 3: Rule Analysis – Association Rule Analysis

Category	Items	No
Nearby Road Network	nearMotorway, nearTrunk, nearPrimary, nearSecondary nearTertiary, nearUnclassified, nearResidential, nearLivingStreet	8
Nearby Traffic Facility	nearBusStop, nearCrossing, nearGiveWay, nearMotorwayJunction, nearSpeedCamera, nearStop, nearTrafficSignals, nearTurningCircle	8
Nearby Amenity	nearSustenance, nearEducation, nearTransportation, nearFinancial, nearHealthcare, nearEntertainment, nearPublicservice, nearFacilities, nearKindergarten, nearSchool, nearDoctors, nearHospital, nearSocialfacility, nearParking	14
Nearby Landuse	nearAllotments, nearAnimalKeeping, nearBasin, nearBrownfield, nearCemetery, nearCommercial, nearConstruction, nearEducation, nearFarmland, nearFarmyard, nearFlowerbed, nearForest, nearGarages, nearGrass, nearGreenfield, nearGreenhouseHorticulture, nearIndustrial, nearMeadow, nearOrchard, nearPlantNursery, nearRailway, nearRecreationGround, nearReligious, nearResidentialLU, nearRetail, nearTrafficIsland, nearVillageGreen	27

Detailed List of Geographic Features

- Nearby geographic features of accident locations – 4 categories, 57 items

  -> Boolean Value (near rather than nearby)

4 Case Study

Part 3: Rule Analysis – Association Rule Analysis

No.	Support	ItemSets	Length
1	0.804406	(nearResidentialLU)	1
2	0.656445	(nearResidential)	1
3	0.570408	(nearResidentialLU, nearResidential)	2
4	0.477821	(nearTransportation)	1
5	0.417684	(nearCrossing)	1
6	0.392676	(nearParking)	1
7	0.392676	(nearTransportation, nearParking)	2
8	0.379280	(nearSecondary)	1
9	0.376600	(nearTransportation, nearResidentialLU)	2
10	0.342662	(nearCrossing, nearResidentialLU)	2
11	0.335814	(nearTransportation, nearResidential)	2
12	0.333135	(nearForest)	1
13	0.317059	(nearSecondary, nearResidentialLU)	2
14	0.313188	(nearBusStop)	1
15	0.302173	(nearTransportation, nearResidentialLU, nearParking)	3
16	0.302173	(nearResidentialLU, nearParking)	2
17	0.301876	(nearCommercial)	1
18	0.292647	(nearCrossing, nearResidential)	2
19	0.292647	(nearPrimary)	1
20	0.292349	(nearGiveWay)	1
21	0.281929	(nearResidential, nearParking)	2
22	0.281929	(nearTransportation, nearResidential, nearParking)	3
23	0.277464	(nearTertiary)	1
24	0.270914	(nearTransportation, nearResidentialLU, nearResidential)	3
25	0.265853	(nearResidentialLU, nearBusStop)	2

ARA Results of the Whole Dataset

No.	Support	Itemsets
1	0.615385	(nearResidentialLU)
2	0.538462	(nearForest)
3	0.407692	(nearResidential)
4	0.400000	(nearMotorway)
5	0.392308	(nearPrimary)
6	0.376923	(nearCrossing)
7	0.369231	(nearTransportation)
8	0.276923	(nearParking)
9	0.269231	(nearCommercial)

Type 1: Accidents Involving Trucks

No.	Support	Itemsets
1	0.866460	(nearResidentialLU)
2	0.810559	(nearResidential)
3	0.566770	(nearTransportation)
4	0.472050	(nearParking)
5	0.428571	(nearCrossing)
6	0.372671	(nearBusStop)
7	0.368012	(nearSecondary)
8	0.315217	(nearSustenance)
9	0.315217	(nearGiveWay)

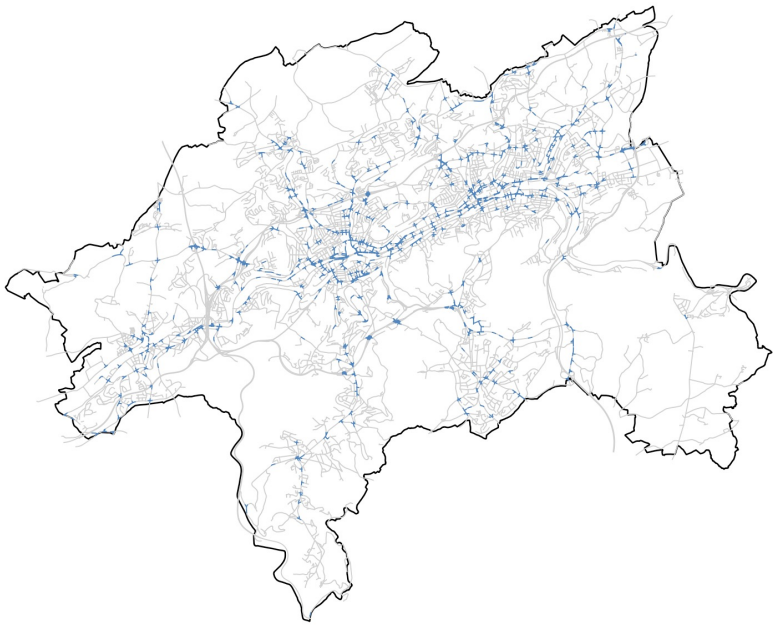
Type 2: Accidents Involving Passengers

ARA Results of Certain Types of Accidents

4 Case Study

Part 3: Rule Analysis – Association Rule Analysis

- Clusters of Community Detection do not exhibit significant geographic features compared to the whole dataset -> bad clustering.
- Both the trained RF model and the association rule analysis results show a high correlation between traffic accidents and crossings, therefore those parts of the road network that are close to crossings can be regarded as -> candidate locations.



Parts of the Road Network that are Close to Crossings

5 Discussions

From Analysis Results to Candidate Sensor Locations

Analysis Methods	Study Objects		Analysis Results	Candidate Locations	Category
Network Analysis	Road Network	Node	High Centrality Nodes	Hot Intersections	Historical Accident Hotspots
		Lixel	High Accident Density Lixels	Hot Road Segments	
		Community	High Accident Risk Communities	Hot Intersections and Road Segments	
Risk Analysis	Traffic Intersection		Predicted High-Risk Traffic Intersections	High-Risk Traffic Intersections	Potential Accident/ Near-Accident Hotspots
Rule Analysis	Accident Location		Accident-Associated Geographic Features	Roads near Associated Geographic Features	

- The results of different methods are not directly comparable, three methods are oriented to different objects, and have different characteristics. -> select as needed.

5 Discussions

Relationship between Traffic Accidents and Geographic Features

- The correlation/association between traffic accidents (risk) and nearby geographic features are explored both in risk analysis and rule analysis, their results show partially strong consistency.

Feature	Importance	Description
nearbyCrossing	0.10	number of crossing
nearbyEdges	0.08	number of road segment
nearbyNodes	0.06	number of intersection
nearbyTransportation	0.06	number of transportation amenities
nearbyEdgeClass	0.05	number of road class
nearbyParking	0.05	number of parking lots
nearbyBusStop	0.04	number of bus stops

No.	Support	ItemSets	Length
1	0.804406	(nearResidentialLU)	1
2	0.656445	(nearResidential)	1
3	0.570408	(nearResidentialLU, nearResidential)	2
4	0.477821	(nearTransportation)	1
5	0.417684	(nearCrossing)	1
6	0.392676	(nearParking)	1
7	0.392676	(nearTransportation, nearParking)	2
8	0.379280	(nearSecondary)	1
9	0.376600	(nearTransportation, nearResidentialLU)	2
10	0.342662	(nearCrossing, nearResidentialLU)	2
11	0.335814	(nearTransportation, nearResidential)	2
12	0.333135	(nearForest)	1
13	0.317059	(nearSecondary, nearResidentialLU)	2
14	0.313188	(nearBusStop)	1

Results of Trained Model and ARA

5 Discussions

Relationship between Traffic Accidents and Geographic Features

- The correlation/association between traffic accidents (risk) and nearby geographic features are explored both in risk analysis and rule analysis, their results show partially strong consistency.
- Some features show strong inconsistency like residential land use types and residential roads. They already exist in large numbers in the city.

Feature	Importance	Description
nearbyCrossing	0.10	number of crossing
nearbyEdges	0.08	number of road segment
nearbyNodes	0.06	number of intersection
nearbyTransportation	0.06	number of transportation amenities
nearbyEdgeClass	0.05	number of road class
nearbyParking	0.05	number of parking lots
nearbyBusStop	0.04	number of bus stops

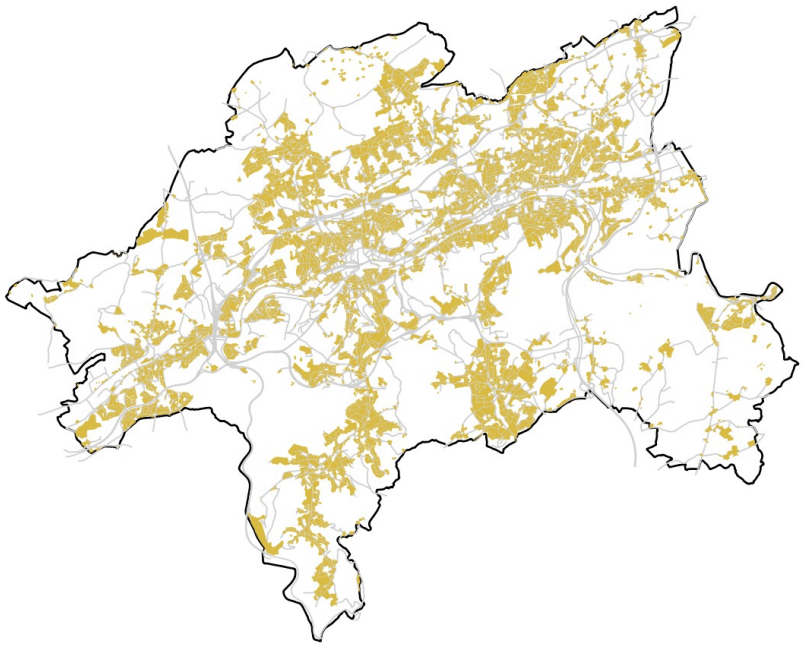
No.	Support	ItemSets	Length
1	0.804406	(nearResidentialLU)	1
2	0.656445	(nearResidential)	1
3	0.570408	(nearResidentialLU, nearResidential)	2
4	0.477821	(nearTransportation)	1
5	0.417684	(nearCrossing)	1
6	0.392676	(nearParking)	1
7	0.392676	(nearTransportation, nearParking)	2
8	0.379280	(nearSecondary)	1
9	0.376600	(nearTransportation, nearResidentialLU)	2
10	0.342662	(nearCrossing, nearResidentialLU)	2
11	0.335814	(nearTransportation, nearResidential)	2
12	0.333135	(nearForest)	1
13	0.317059	(nearSecondary, nearResidentialLU)	2
14	0.313188	(nearBusStop)	1

Results of Trained Model and ARA

5 Discussions

Relationship between Traffic Accidents and Geographic Features

- The correlation/association between traffic accidents (risk) and nearby geographic features are explored both in risk analysis and rule analysis, their results show partially strong consistency.
- Some features show strong inconsistency like residential land use types and residential roads. They already exist in large numbers in the city.
- Geographic features are not the cause of traffic accidents but can help to locate accident hotspots as their symptoms.



Geographic Distribution of Residential Land Use Types in Wuppertal (Data Source: OSM)

5 Discussions

Current Limitations and Where to Continue

- In network analysis, the main difficulty is to find effective centrality measures and community detection algorithms with reasonable practical meanings.

  ➔ Current network analysis is limited, go deeper with designing customized centralities and algorithms native to this problem.

- Limited by dataset (only negative values, no positive values), the current rule analysis is incomplete (only from accident to geographic feature, no opposite rules).

  ➔ Combine risk analysis and rule analysis, and apply them to traffic intersection data (both positive values – low-risk and negative values – high-risk).

- Analysis has not been iterated based on the results (e.g. crossings).

  ➔ Aggregate crossings as traffic intersections for further analysis.

6 Conclusions

Summary

- This thesis is an exploratory study of a class of spatial decision-making processes. It aims to explore the location strategy of traffic sensors in the application scenario of traffic safety. Specifically, it explores the possible solutions for detecting or predicting traffic accident and near-accident hotspots on the road network.
- This thesis proposed three possible solutions in its methodology, they all use open data sources, therefore they are all transferable. They are based on different analysis methods, aiming at different study objects, and they acquire different types of results.
- With the case study, the effectiveness of the methodology is initially verified. Network analysis methods can effectively identify the historical accident hotspots on the road network. The correlation between accident risk and geographic features can be effectively explored, modelled, and used for risk prediction.

6 Conclusions

General Guideline

- Select by target locations: road intersection, traffic intersection, road segment or traffic community?
- Select by data availability: there is historical traffic accident data or no historical data? Single year or multiple years?
  - Network analysis requires multiple years of historical data.
  - Prediction model and association rules can be transferred from A to B.

6 Conclusions

Outlook

- The process of applying advanced traffic sensors will not stop to achieve smart and sustainable transportation. Therefore, the discussion about their location strategy have to continue, extend, and adapt to the new scenarios.
- The future work of solving TSLP from the perspective of traffic safety can continue in many ways. More effective network analysis methods can be developed, and more attributes of the road network can be used in the process. The association relationship between accident risk and geographic features can be further and completely analyzed, and it can be implemented on a larger geographic scale. More work should be carried out to explore other methods used for solving TSLP from the perspective of traffic safety, like optimization methods and heuristic methods.

7 Bibliography

[1] Zhao, J., Xu, H., Liu, H., Wu, J., Zheng, Y., & Wu, D. (2019). Detection and tracking of pedestrians and vehicles using roadside LiDAR sensors. *Transportation research part C: emerging technologies*, 100, 68-87.

[2] Zhang, J., Xiao, W., Coifman, B., & Mills, J. P. (2020). Vehicle tracking and speed estimation from roadside lidar. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 5597-5608.

[3] Zhang, Z., Zheng, J., Xu, H., & Wang, X. (2019). Vehicle detection and tracking in complex traffic circumstances with roadside LiDAR. *Transportation research record*, 2673(9), 62-71.

[4] Wu, J., Xu, H., Zheng, Y., & Tian, Z. (2018). A novel method of vehicle-pedestrian near-crash identification with roadside LiDAR data. *Accident Analysis & Prevention*, 121, 238-249.

[5] Owais, M. (2022). Traffic sensor location problem: Three decades of research. *Expert Systems with Applications*, 118134.

[6] Liu, Z., Luo, P., Murphy, C. E., & Meng, L. (2023). The Potential of VGI and Traffic Flow Generation Model for Solving the Traffic Sensor Location Problem. *Abstracts of the ICA*, 6, 145.

[7] Gelb, J. (2021). spNetwork: A Package for Network Kernel Density Estimation. *R journal*, 13(2).

[8] Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10), P10008.

7 Bibliography

[9] Kumar, S., & Toshniwal, D. (2015). A data mining framework to analyze road accident data. *Journal of Big Data*, *2*(1), 1-18.

[10] Lin, L., Wang, Q., & Sadek, A. W. (2014). Data mining and complex network algorithms for traffic accident analysis. *Transportation Research Record*, 2460(1), 128-136.

[11] Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. *In Proceedings of the 1993 ACM SIGMOD international conference on Management of data* (pp. 207-216).

[12] Agrawal, R., & Srikant, R. (1994, September). Fast algorithms for mining association rules. In Proc. *20th int. conf. very large data bases, VLDB* (Vol. 1215, pp. 487-499).

[13] Cao, Q., Li, Z., Ma, Y., & Li, H. (2022). Optimal layout of heterogeneous sensors for traffic accidents detection and prevention. *IEEE transactions on instrumentation and measurement*, 71, 1-13.


ITC
UNIVERSITY OF TWENTE.
TECHNISCHE UNIVERSITÄT DRESDEN
Technical University of Munich
TU WIEN
TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology