TRAFFIC PREDICTIONS BASED ON PROBABILISTIC GRAPHICAL METHOD

M.SC. Thesis Defense – 24th Aug 2020

Mohan Raju Jayashankaramma Shankar Raj | s6040489

Thesis Assessment Board:

Prof. Dr. IR. Alfred Stein (Chair),
Dr.IR. M. Van Keulen (External)
Dr. Frank Badu Osei,
Dr. Michael Yang

Procedural Advisor:

Drs. J.P.G Bakx (Wan)

INTRODUCTION

- Managing traffic congestion requires accurate knowledge of the causes of congestion on a given road network as well as their relative significance and respective solutions.
- Previous work focused on estimating the relative effect of different causes of congestion, with statistical analysis such as linear regression methods.
- The study developed a method to build a BN model according to historical traffic data and demonstrated the BN-based traffic analysis.

RESEARCH IDENTIFICATION

To identify leading causes for traffic congestion, estimating traffic flow & lastly analysing street networks for Amsterdam and Rotterdam using probabilistic graphical method.

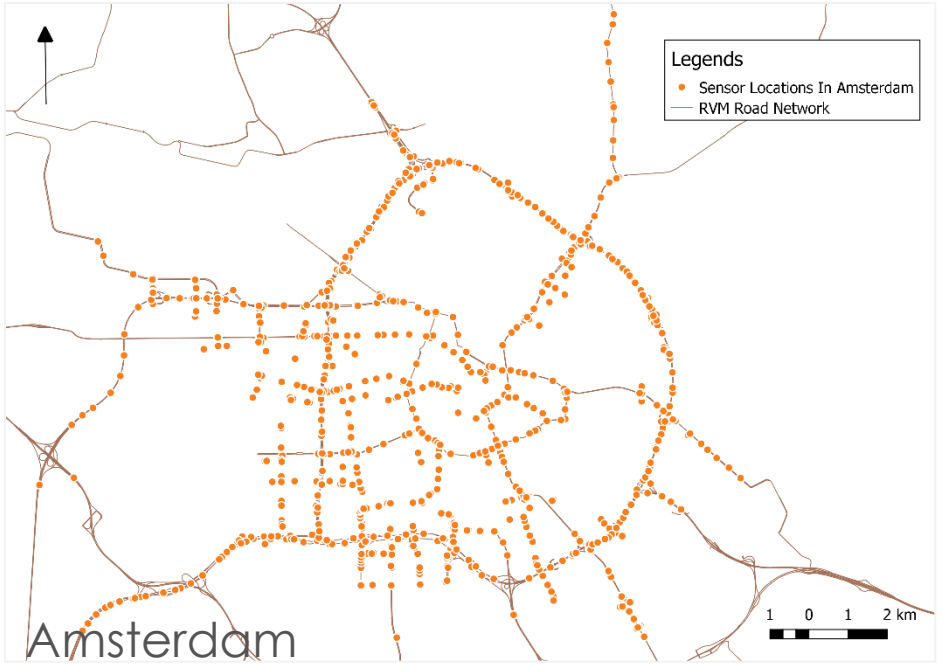
What are the techniques to identify leading causes for congestion in a road networks, and how are the short-term predictions made?

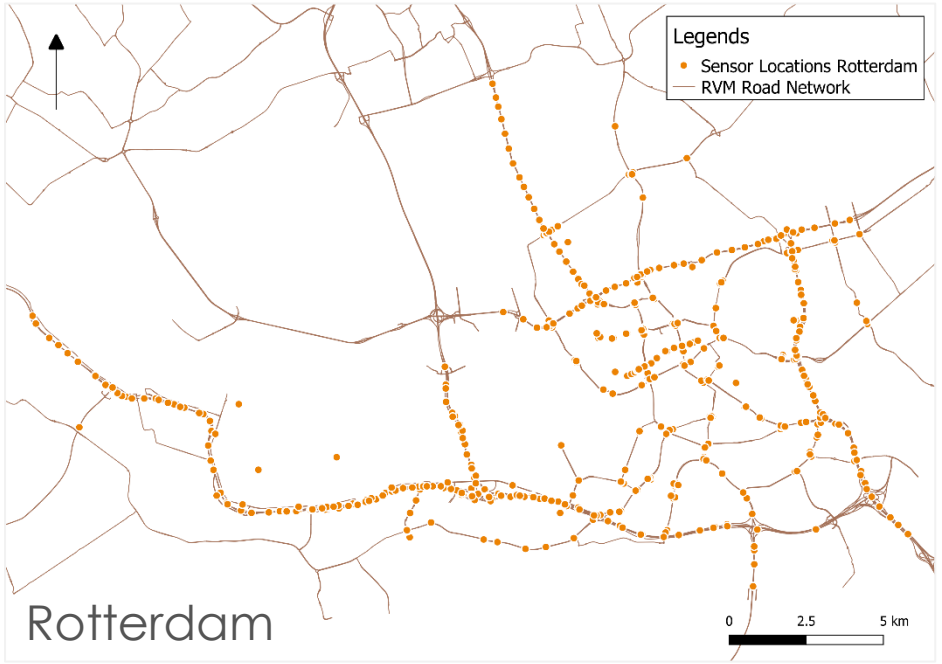
How accurate and reliable are the short-term traffic predictions made using Bayesian Networks?

What are the characteristics of road segments of Amsterdam and Rotterdam road networks?

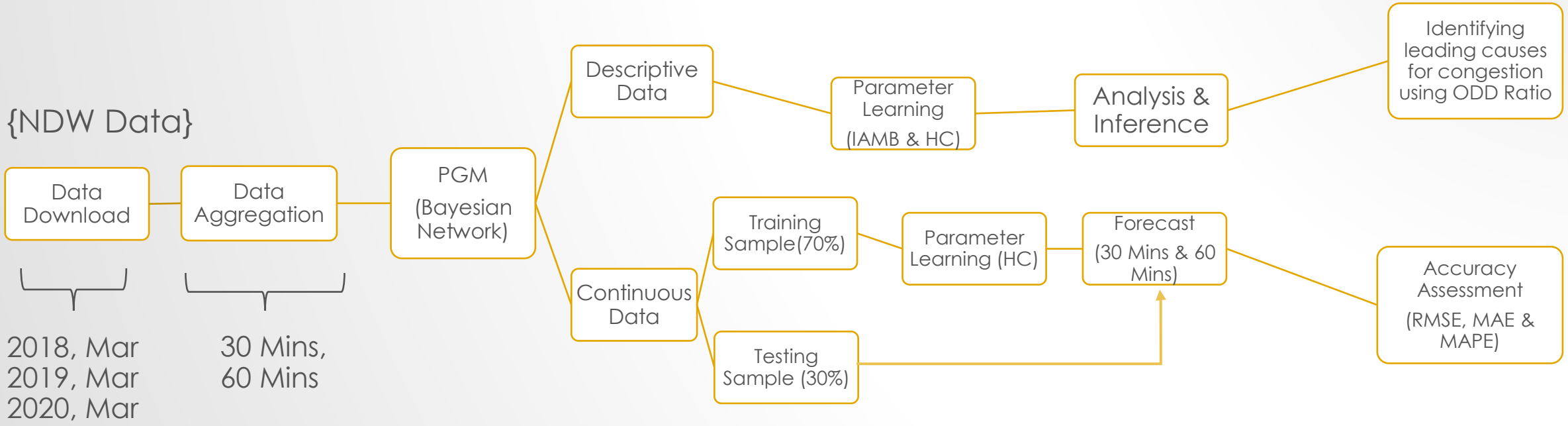
DATA AND STUDY AREA

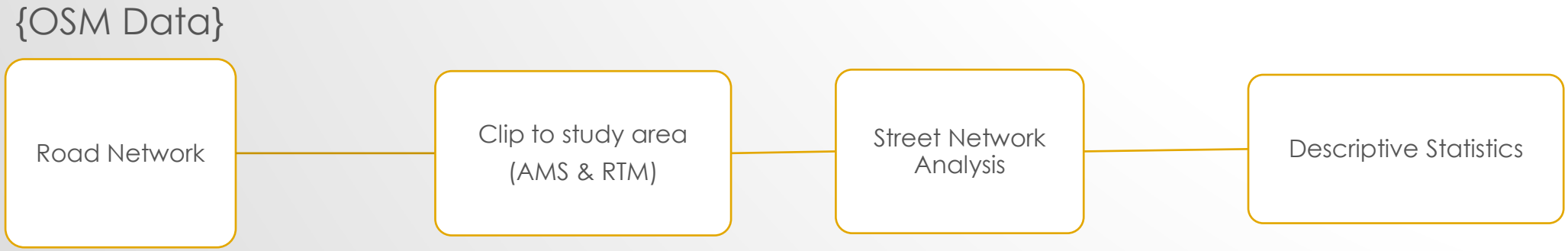
- **Historical Data** – *National Databank Wegverkeersgegevens* (NDW)
- **Street Network Data** – Open street maps (OSM)
- **Study area**: RVM Road networks of Amsterdam and Rotterdam





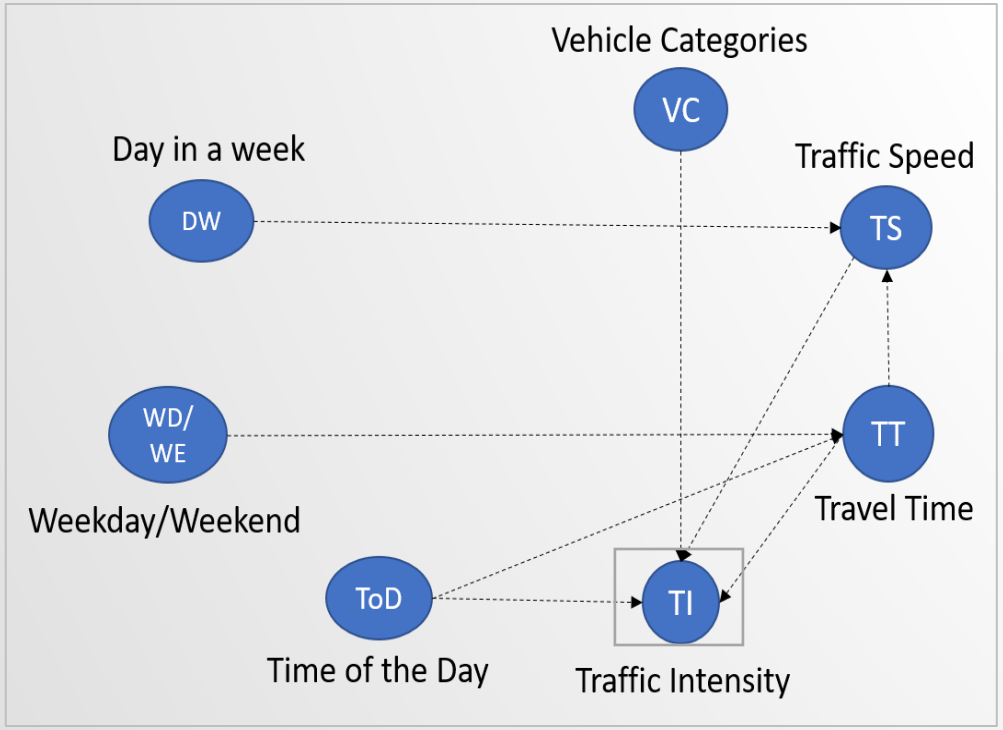
METHODOLOGY



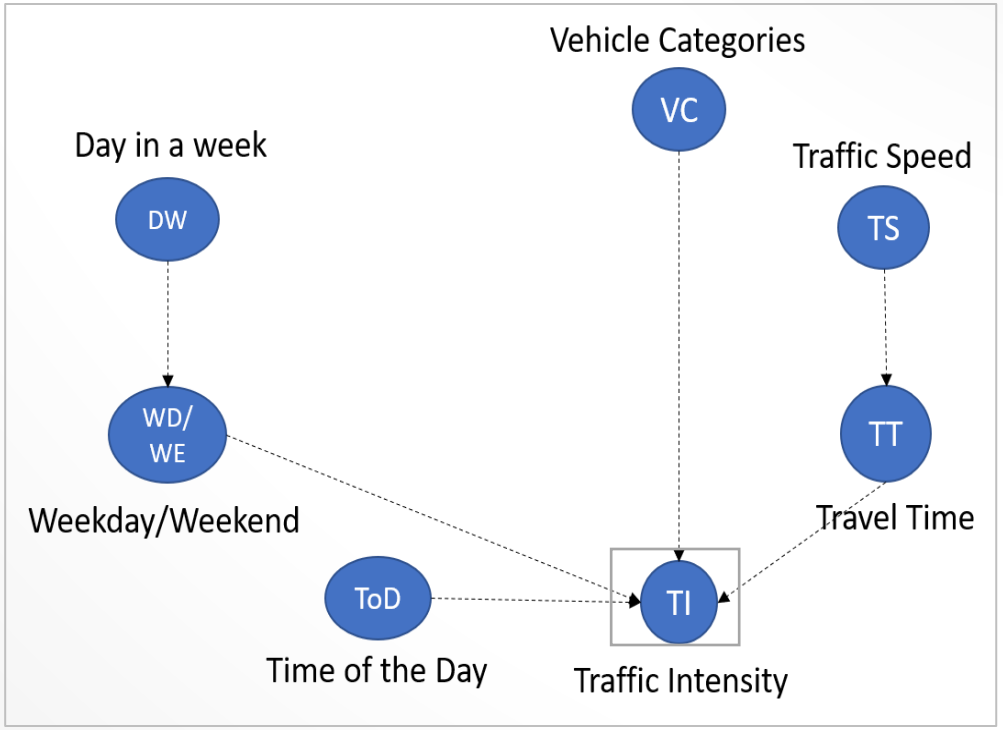


BAYESIAN MODEL (NETWORK STRUCTURE)

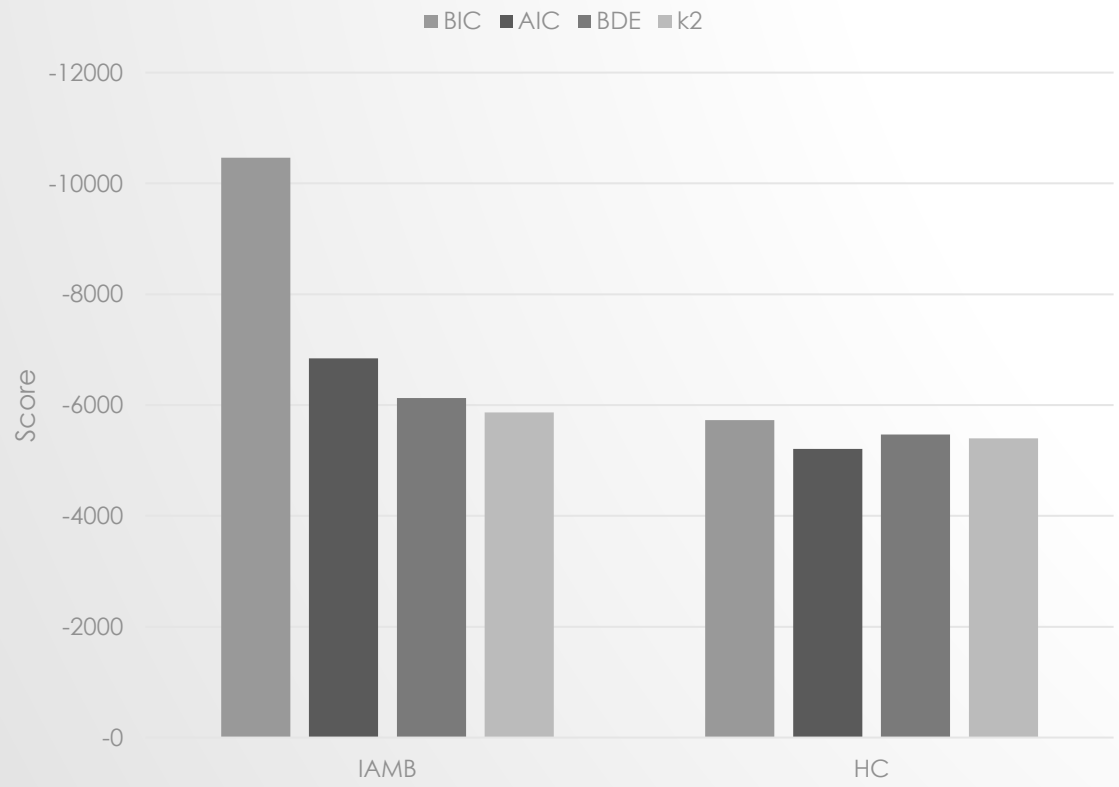
IAMB ALGORITHM



HILL CLIMBING ALGORITHM

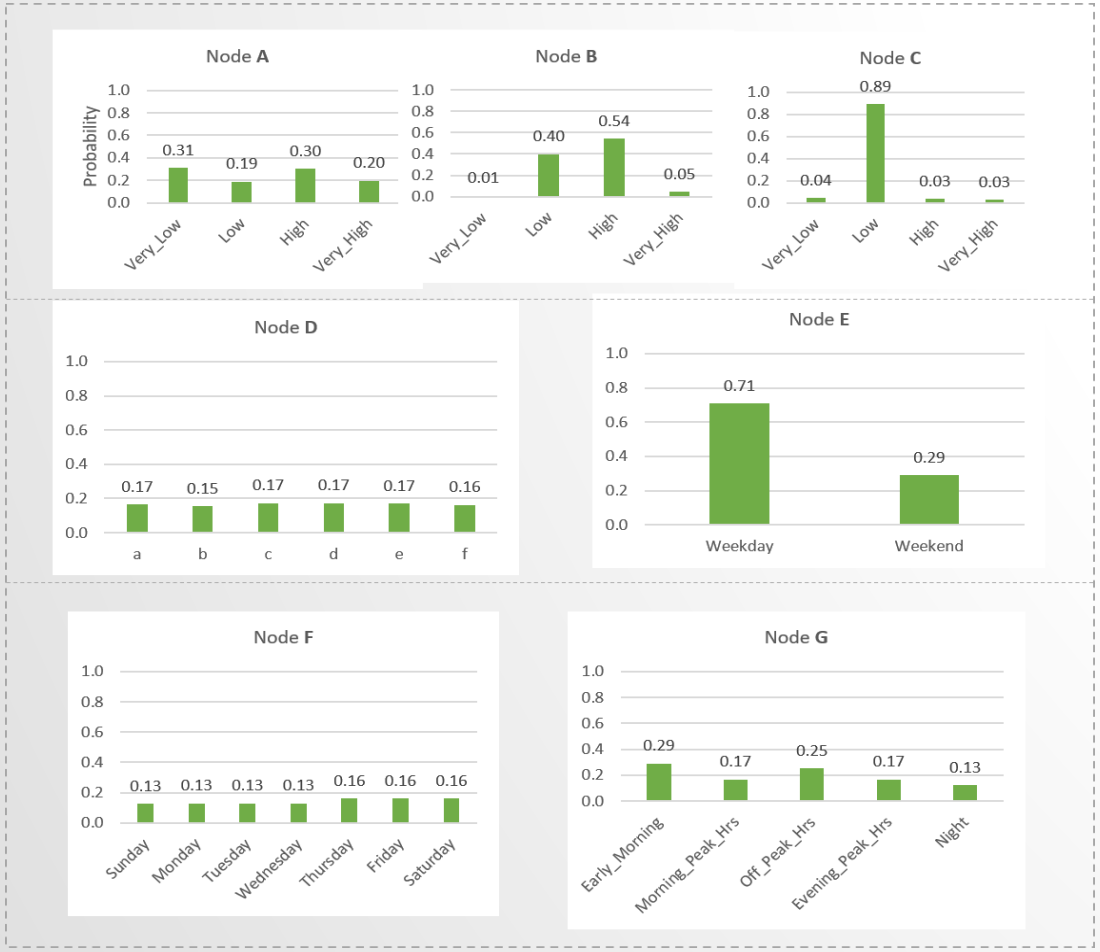


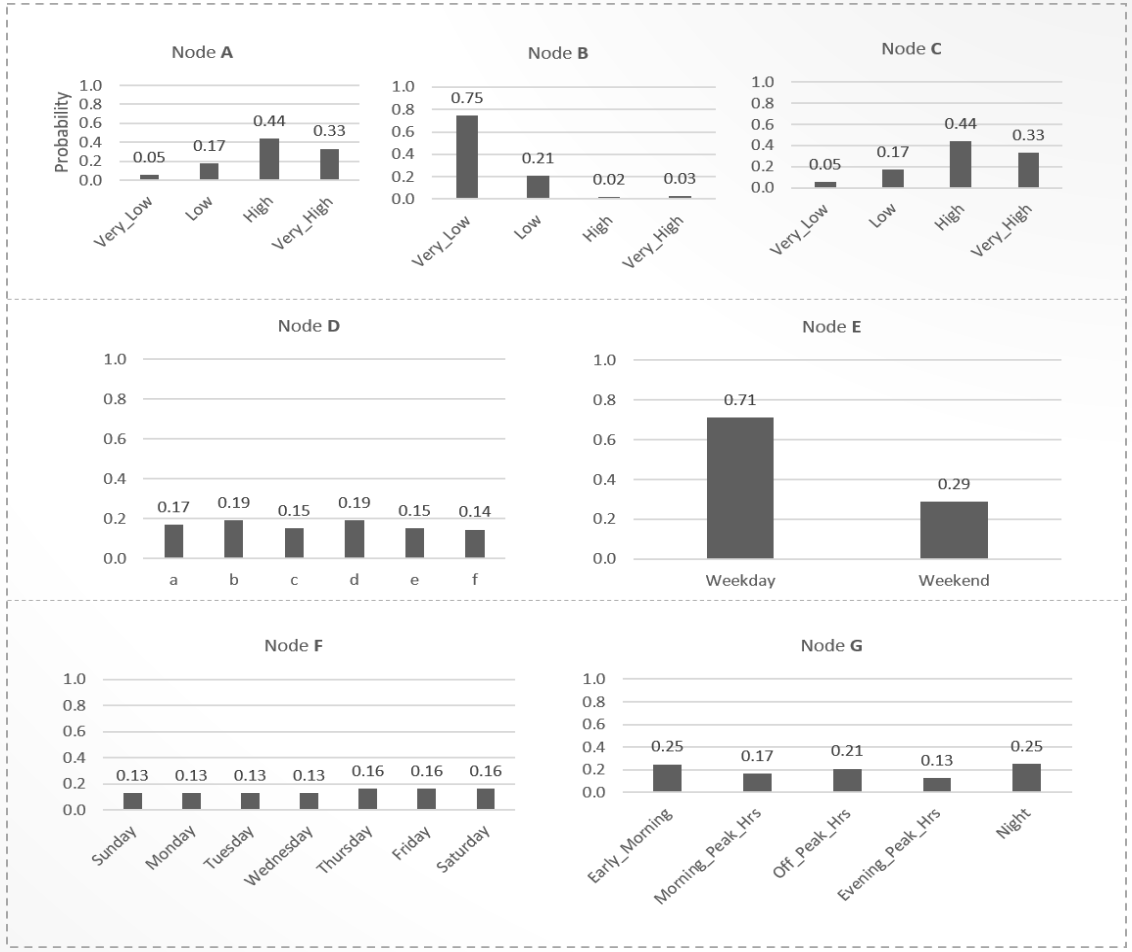
MODEL SUITABILITY



k-fold cross-validation for Bayesian networks		
Target learning algorithm	IAMB	Hill-Climbing
Number of folds	10	10
Loss function	Log-Likelihood Loss (disc.)	Log-Likelihood Loss (disc.)
Expected loss	7.234212	6.804658

RESULTS (PRIOR PROBABILITY)





Representing marginal probability distribution for 7 nodes in the designed BN model **(left (AMS) & right (RTM), 2018)**.

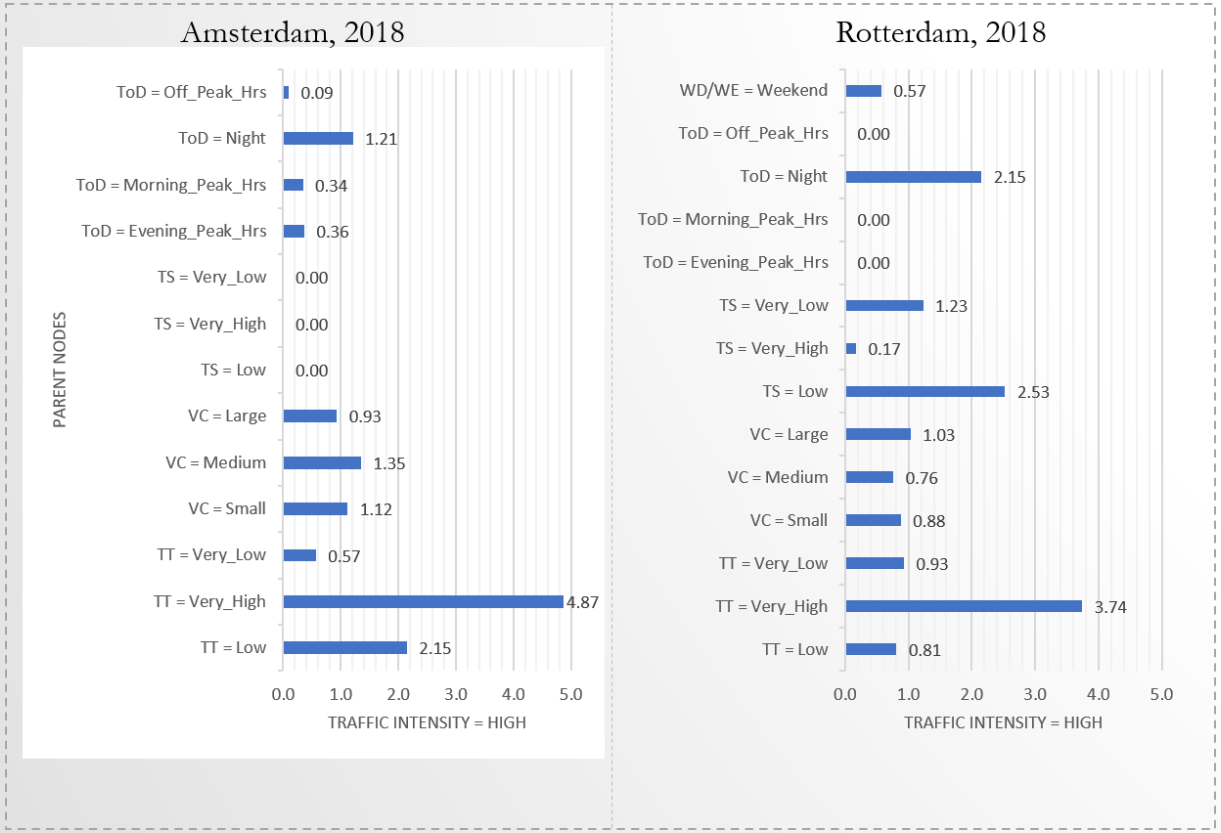
Node **A**- Intensity, **B**- Travel Time, **C**- Traffic Speed, **D**- Vehicle Categories, **E**- Weekend or Weekday, **F**- Day in week & **G**- Time of the day

RESULTS (POSTERIOR PROBABILITY)

- Influence of 'Time of the Day' on '*Traffic Intensity*'
- Influence on *Traffic Intensity* during 'Weekdays & Weekends'
- Influence of 'Vehicle Size' on '*Traffic Intensity*'
- Influence of 'Travel Time' on '*Traffic Intensity*'

What are the major factors Influencing Traffic Intensity?

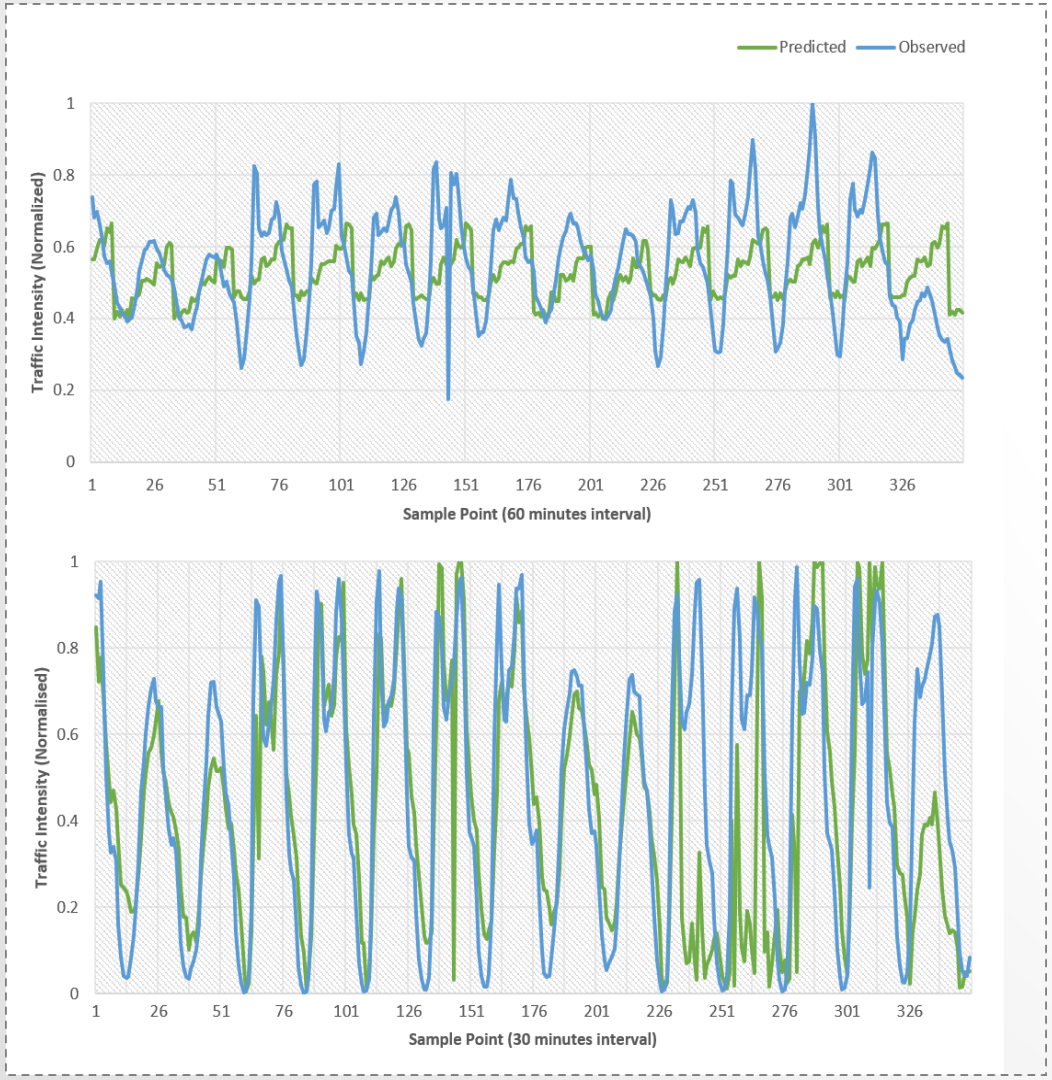
RESULTS (CONGESTION DIAGNOSIS)



OR between Contributing nodes and Traffic Intensity = "High" | AMS & RTM - 2018

Identifying major causes for increased 'Traffic Intensity' by measuring the odd ratio

TRAFFIC FLOW PREDICTION

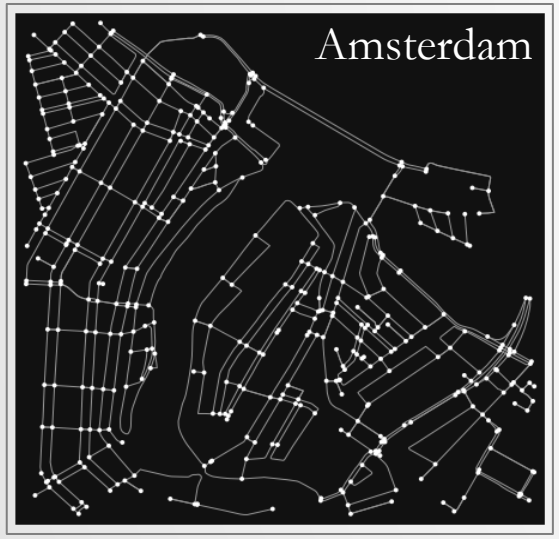


Traffic intensity forecasting (60- & 30-minutes interval)

Model	Measurement	30 Minutes		60 Minutes	
HC		Amsterdam	Rotterdam	Amsterdam	Rotterdam
	RMSE	0.784	0.116	0.920	0.346
	MAE	0.586	0.866	0.609	0.947
	MAPE	58.19	23.44	349.79	41.85

Comparison of RMSE, MAE & MAPE for Amsterdam & Rotterdam

ANALYZING STREET NETWORK



Indicators	Amsterdam	Rotterdam
Area covered	193.5 sq km	161.6 sq km
Number of nodes in a network (n)	11520	11245
Number of edges in a network (m)	26580	25387
Average node degree	4.6	4.5
Intersection count	9995	9682
Intersection density (per km2)	51.64	59.88
Average streets per node	2.93	2.91
Average street length(m)	102.58	97.7

ANALYZING STREET NETWORK(CONTD.)

Monochrome images of Amsterdam (left) and Rotterdam(right) street network (One square mile)

CONCLUSION

- The study has combined discrete and continuous data and provided a framework using Bayesian network for estimating the traffic congestion by using a data-driven approach.
- This study proposed a BN analysis approach to modeling the probabilistic dependency structure between causes of congestion and analyzing the probability of traffic congestion given various scenarios.
- Accuracy and reliability are measured through RMSE MAE and MAPE values. The model performs better with 30 minutes interval data compared to 60 minutes.
- Based on the road network statistics - Amsterdam as many nodes, edges and networks are in semi-structured type and Rotterdam has fewer number of edges and nodes with square typed structure.

RECOMMENDATIONS AND FUTURE WORKS

- The study made use of 4-5 parameters of the data from the sensor data that are Speed, Intensity, Travel Time and Vehicle characteristics. To make the BN model more robust, which yields good results, other variables such as specific lane, driving lane, the direction of the road and bicycle data can be included.

- The BN model completely depended on single-source data. To enhance the model, other sources like rainfall and temperature date etc. can be incorporated. Data from various sources will have an impact on the prediction model.

- The current BN model is a static & non-spatial one. In future, the entire can be made dynamic by adding time data and spatial components.

THANK YOU