Explainable Bayesian Networks applied to transport vulnerability*

Alta de Waal[a,b], Johan W. Joubert[c]

[a] *Department of Statistics, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa*
[b] *Centre for Artificial Intelligence Research (CAIR), South Africa*
[c] *Centre for Transport Development, Industrial & Systems Engineering, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa*

Abstract

To deal with increasing amounts of data, decision and policymakers frequently turn to advances in machine learning and artificial intelligence to capitalise on the potential reward. But there is also a reluctance to trust black-box models, especially when such models are used to support decisions and policies that affect people directly, like those associated with transport and people's mobility. Recent developments focus on *explainable* artificial intelligence to bolster models' trustworthiness. In this paper, we demonstrate the use of an explainable-by-design model, Bayesian Networks, on travel behaviour. The model incorporates various demographic and socioeconomic variables to describe full day activity chains: activity and mode choice, as well as the activity and trip durations. More importantly, this paper shows how the model can be used to provide the *most relevant explanation* for people's observed travel behaviour. The overall goal is to show that model explanations can be quantified and, therefore, assist policymakers to truly make evidence-based decisions. This goal is achieved through two case studies to explain people's vulnerability as it pertains to their total trip duration.

Keywords: Bayesian networks, Explainable artificial intelligence, Most relevant explanation, Activity chain

1. Introduction

Advances in activity-based travel models allow the model builder to represent an urban transportation system accurately. The models richly describe the population in terms of their socioeconomic variables. The resulting activity and mode choice behaviour of individuals can allow for comprehensive, full-day mobility patterns. These accurate models make it possible for urban and transport planners to anticipate both intended and, hopefully, unintended consequence as they consider and evaluate policy interventions.

But the advances in modelling come at a price. Both Rasouli and Timmermans (2014) and Hasnine and Habib (2021) highlight these data-hungry models' plight as they review state of the art in activity-based and the associated mode-choice models, respectively. Many of the more advanced

*This paper is published. Please cite as follows: De Waal, A., Joubert, J.W. (2022). Explainable Bayesian Networks applied to transport vulnerability. *Expert Systems with Application*, **209**, 118348.
DOI: 10.1016/j.eswa.2022.118348

Email addresses: alta.dewaal@up.ac.za (Alta de Waal), johan.joubert@up.ac.za (Johan W. Joubert)

models rely on detailed travel surveys with travel diaries or high-resolution geospatial positioning system (GPS) data as the latter becomes more ubiquitous (Wand et al., 2018). The collected data is frequently only a tiny sample of the population, given the cost and survey burden.

Researchers subsequently turn to advanced pattern recognition algorithms to process and make sense of the limited data's travel behaviour. These estimated behavioural models aim to accurately capture the complex decision processes individuals go through when making choices about the activities and the trips connecting those activities. Some of these behavioural models are basic rule-based heuristics, while others are complicated and nested models relying on utility-maximising assumptions. Researchers and model estimators aim to find an acceptable trade-off between model complexity, computational burden, and how well the model reflects reality. And all these trade-offs should be subjected to what the available data will and will not support.

The end goal of these often sophisticated models is to affect change in the form of improved mobility. But how to translate the models into practice and predict future efficacy is often a daunting task for practitioners and policymakers.

Koushik et al. (2020) recently reviewed the various machine learning applications and artificial intelligence (AI) related to activity-travel behaviour. Many applications use AI for predicting travel behaviour like travel time (Abdollahi et al., 2020), route (Xinyue et al., 2018) and mode choice (Hagenauer and Helbich, 2017), to name but a few. Since the early development of expert systems in the 1970s, it soon became apparent that experts are reluctant to trust the system if they do not understand how it obtained the result (Teach and Shortliffe, 1981). Many early expert systems had comparable accuracy to domain experts (Buchanan and Shortliffe, 1985). The key to trust in an AI system is explaining its reasoning and decisions. So the focus of AI systems shifted to supporting human intelligence rather than imitating it (Lacave and Díez, 2002). For an AI system to be trustworthy, it should explain how it makes conclusions and why the findings are appropriate for domain experts to understand (Yuan et al., 2011). Regardless of the form, an essential issue in AI resolution is what to explain: a good explanation should address the task's objectives (Leake, 1995). Koushik et al. (2020) conclude their review with the concern that the (lack of) interpretability of many machine learning techniques remains an issue for transport researchers.

The contribution of this paper is two-fold. Firstly, it applies Bayesian Network (BN) theory to explain individuals' activity- and mode choice behaviour. BNs are gaining traction as a machine learning approach because it is highly interpretable, and the learned structure is the result of empirical data, expert knowledge, or some combination. It is this combination that is of value in this paper. The network presented includes demographic variables, activity and trip variables, and the temporal variables associated with activity and trip durations. This addition of temporal variables is not only novel, it also improves the performance of the BN on synthesising travel patterns.

This paper's second contribution demonstrates the BNs explanatory power, which is different to past research that used BNs for their predictive ability (Aghaabbasi et al., 2020; Zhang et al., 2019) and synthesis of populations (Sun and Erath, 2015; Joubert, 2018) and their daily activity chains (Joubert and De Waal, 2020). Lacave and Díez (2002) categorise the explanation in BNs into three groups: the knowledge base, the reasoning process and evidence. Recently, a fourth group, decisions, was suggested by Derks and De Waal (2020). In this paper, we focus on the explanation of the evidence. Can we explain why we observe certain phenomena in people's travel behaviour? Can policymakers mitigate risk for vulnerable groups in the process? To demonstrate its usefulness in policy, we show how to explain travellers' vulnerability quantitatively. While there is interest in explainability in the expert system literature, most contributions on BNs are limited to address the theoretical aspects. In this paper we provide a roadmap on how to *apply* explainability. Our application demonstrates Most Relevant Explanation, a method that produces both precise and concise explanations.

The remainder of the paper is structured as follows. Section 2 reviews the advances in and need for explainability in artificial intelligence. The section then turns more specifically to the application of AI in transport and travel behaviour. The following section, Section 3, introduces the BN model and the data used to learn its structure and posterior probabilities. The model is evaluated by means of standard machine learning evaluation metrics. Section 4 presents one metric's theory, the most relevant explanation, and applies it in Section 5 to two examples of vulnerable groups. The paper concludes in Section 6 with some remarks about the future outlook for explainable machine learning.

2. Literature review

The sophistication and so-called *democratisation* of AI systems (Garvey, 2018) in recent years makes it possible to design and deploy AI systems without human intervention. Still, decisions derived from AI systems directly affect humans' lives, which shifted the research focus from AI performance to *eXplainability* of AI methods (XAI) (Lipton, 2018). Ironically, the very first AI systems were easy to interpret. As rule-based expert systems they were developed by generating rules that linked to the features in the dataset (Buchanan and Shortliffe, 1985). The most common terms used in XAI are 'understand', 'interpret', 'explain', 'transparency' and 'trust'. These terms are often used interchangeably and not always consistently. This is where taxonomies (Barredo Arrieta et al., 2020) and workgroups such as the FAccT (Fairness, Accountability, and Transparency) conference (`https://facctconference.org/`) play important roles to establish common grounds among XAI practitioners. One of the primary goals or XAI is trustworthiness (Barredo Arrieta et al., 2020): does the model act as it is intended to under decision-making? Humans are more likely to adopt systems that are interpretable, tractable and trustworthy (Zhu et al., 2018).

Many AI models are interpretable or transparent by design. One example of transparency is a logistic regression model for which the error surface and decision boundaries are clearly understood. Also available in the suite of XAI methods are model interpretability techniques which are external to the model architecture (Barredo Arrieta et al., 2020). These include Interpretable Model-Agnostic Explanations (LIME) (Ribeiro et al., 2016), uncertainty quantification (De Waal and Steyn, 2020), what-if tools and visualisation.

Bayesian networks are interpretable by design and exhibit post-hoc explanation methods. From a mathematical point of view, a BN is a compact representation of the joint probability over all its variables (Nielsen and Jensen, 2009). This is possible because conditional independencies hold between unconnected variables in the model. Furthermore, from a knowledge representation point of view, a BN is a graphical model thereby providing a formal communication language which is easy for both humans and computers to read (Jensen, 1996). Many examples in literature exist where BNs provide insight into complex, multidisciplinary problems and facilitate expert knowledge elicitation (Düspohl et al., 2012; Borsuk et al., 2004). These insights are not only obtained through the transparent design and graphical display of BNs, but also through post-hoc explanation methods. These methods can be categorised into explanation of the model, reasoning, evidence and decisions (Derks and De Waal, 2020). We focus on explanation of evidence which is to request the state of variables given some evidence (Nielsen and Jensen, 2009). Many examples of explanation of evidence can be found in medical decision support systems, where one wants to find the most probable diagnosis, given the clinical observations (Lucas et al., 2000; Geenen et al., 2006). Two common mechanisms for finding the explanation of evidence are Maximum a Posteriori (MAP) and Most Probable Explanation (MPE) (Pearl, 1988). The problem with these two methods are that they produce hypotheses that are overspecified, meaning the explanation contains irrelevant variables. The reason for this is because they use *probability* in finding the most likely configuration of a set of variables, implying that an explanation may contain independent and/or marginally relevant variables because they have a high probability (Yuan et al., 2011). What is needed, is a method that returns precise and concise explanations. More recently, such a method was introduced by Yuan et al. (2011), namely the Most Relevant Explanation (MRE) which is able to discard less relevant variables from its explanation. MRE performs the task of returning only the essential knowledge in the explanation. Whereas the main objective of XAI is to gain trust in the system, explanation of evidence can be thought of as actionable explanation — the outcome of the explanation itself can be used in decision-making. It is clear to see how such concise explanations can be very informative and useful in decision-making and policy formulation, as oppose to only returning accurate predictions which is the case for many AI systems. The usefulness of explanation of evidence was illustrated in a recent application in forensic science where the explanation of evidence is used to evaluate contesting propositions given a set of DNA evidence (Taroni et al., 2021).

How is this relevant to transport and, more specifically, transport policy? In their review, Rasouli and Timmermans (2014) argue that while activity-based models have developed and addressed the significant shortcomings of the classical four-step model, much development is still required. One particular opportunity for improvement they identified is the ability of activity-based models to capture people's behaviour more accurately. The review highlights that studying travel behaviour using discrete choice models has several disadvantages. The utilitarian approach has its foundation in either random utility or rational choice theory. The latter assumes that decision-makers have perfect knowledge of all the choice alternatives and have enough capacity and time to review them all. Ma (2015) echoes the unrealistic assumptions. The estimated models frequently employ logistic regression, of which the interpretation is notoriously hard if going beyond the "relative sign is correct" and "relative order of magnitude" or statistical significance of determinants.

In response to the shortcomings of the mainstream travel behaviour models, more recent advances specifically consider bounded rationality that better emulates natural human decision-making and causal models of decision-making under uncertainty. Ma (2015) indicates that when it comes to understanding people's travel behaviour, various factors influence one's choice: journey, socio-demographic and spatial characteristics, and socio-psychological factors. For policymakers, it is essential to understand travel behaviour because the aim of a policy is frequently to *affect changes* in behaviour. A better understanding of the behaviour will likely result in higher efficacy and insightful anticipation of both intended and unintended consequences. Aghaabbasi et al. (2020) also use BN as a predictive model to understand what influences students to use (and use frequency) of ride-sharing options.

In previous work (Joubert and De Waal, 2020), a Bayesian network (BN) was developed to synthesise activity-based travel demand data. The results showed that Bayesian networks can successfully synthesise both activity and trip chain structures accurately and outperforms a frequentist approach. Of importance in the original study was the ability to synthesise activity chains of all travel patterns, and not only the frequently observed ones like home-work-home or home-education-home. This led to the conclusion that understanding the behaviour behind the less frequently observed travel patterns may provide valuable insight to decision-makers.

What can we learn from adding a temporal dimension? It could tell us, for example, the amount of time people spend outside their homes. For example, a person with two jobs, using public transport will most probably leave home in the early hours and return after dark. Furthermore, it could shed light on the most probable demographic profile associated with this travelling behaviour.

3. The model

This section explains the data preprocessing and its use in learning the BN structure. This paper benefits from the diary component of the City of Cape Town's 2013 travel survey, which 2974 households completed. The survey solicited information about the household size, income, mobility impairments and vehicle ownership at the household level. At the individual level, the survey includes data on each household member's age, gender, and mobility pattern in the form of an activity chain.

3.1. Demographic and travel variables

Table 1 shows the demographic variables' encoding This paper only uses a subset of the travel diary records, namely those 7 345 individuals in the working-age categories: Young, Early career and Late career. `Young` individuals are in the age range of 13 to 23; `Early career` individuals in the range 24 to 45; `Late career` in the 46 to 68 range. We argue that working-age individuals have a prominent and different behaviour pattern than school-going children and retired persons.

Since many households choose not to provide household income information, the travel survey accommodates this in the `hhInc` variable by estimating the income using an asset index from Filmer and Pritchett (2001). In 2013 terms, `Low` income refers to a monthly income of ZAR 3 200 or below (using an exchange rate of EUR 1 = ZAR 17.23 and USD 1 = ZAR 14.26 that equates to approximately EUR 186, or USD 224). The ceiling for the next income class, `LowMiddle`, is ZAR 25 600 per month (EUR 1 486 or USD 1 795); for `HighMiddle`, it is ZAR 51 200 per month (EUR 2 972 or USD 3 590); and high income is anything higher.

The unique record reflects both the household id and the member number. Consequently, one can easily calculate the household size, which is, in line with the Jakarta study of Yagi and Mohammadian (2008), a helpful characteristic in developing countries where households tend to be larger. The `Couple` category reflects a two-member household. This category, contrary to what the name suggests, also represents single-parent households with one child. A small family is 3 to 5 members, and a large family is six members or more. Finally, the housing variable distinguishes between formal and informal dwellings and aims to see if housing formality affects the travel demand.

Table 2 summarises the encoding scheme for the travel variables related to activity and mode choice. Most activity types are straightforward, except for `e3`, which might be context-specific to South Africa. General safety issues and a high reliance on a private car means that many (more affluent) people tend to drop or collect their children from school using their vehicle. As such, it is a dedicated activity type that cannot be incorporated with say education, generally, because it is much shorter. In line with previous work, this paper simplifies some variable classes. Firstly, we combined primary and secondary (`e1`) with tertiary education (`e2`) into simply *education*

Table 1: Encoding scheme for demographic variables. In the *Values* column the numbers in brackets represent the number of occurrences of the specific variable value. The first group of variables relates to individual characteristics, and the second group to household characteristics (Adapted from Joubert and De Waal (2020)).

Code	Description	Values
`gender`	Gender	`Female` (3 853); `Male` (3 465)
`age`	Aggregated age group based on Census categories	`Young` (1 562); `Early career` (3 536); `Late career` (2 247)
`edu`	Completed education	`None` (332); `Primary` (1 309); `Secondary` (4 632); `Tertiary` (1 072)
`license`	Whether the individual has a driver's license	`Yes` (2 725); `No` (4 620)
`access`	The number of vehicles the person has access to	`None` (3 416); `Single` (1 956); `Multiple` (1 973)
`employed`	Whether the individual is employed	`Yes` (3 582); `No` (3 763)
`travelForWork`	Whether the individual's work is, effectively, driving or travelling	`Yes` (308); `No` (6 535)
`workFromHome`	Whether the person works from home	`Yes` (283); `No` (6 649)
`hhInc`	Calculated asset value as the proxy for household income	`Low` (1 912); `LowMiddle` (4 439); `HighMiddle` (636); `High` (358)
`hhSize`	Household size	`Single` (332); `Couple` (1 018); `Small` (3 953); `Large` (2 042)
`housing`	Type of housing for the main dwelling	`Informal` (745); `Formal` (6 600)

(`e`). Secondly, recreational visits to friends (`v`) were recoded as *leisure* (`l`) activities. Thirdly, we reclassified activities with a medical purpose (`m`) as other activities (`o`).

For mode choice, we group all forms of public transport together, which, in South Africa, represents the majority-share paratransit mode referred to as minibus taxis, formal bus, commuter rail and bus rapid transit. Bus rapid transit is likely underrepresented in the survey data as it was in its infancy at the survey time.

Figure 1 uses a Sankey diagram (R Core Team, 2019; Allaire et al., 2017; Chang, 2018) to depict activity chains' structure as we observe them in the travel diary. The height of a bar connecting two adjacent activity types indicates the proportion of observations. For example, a large number of working-age individuals stay at home with no travel activity. One sees this as the thick connection between home (`h`) as the first activity and none (`n`) as the second activity. Unfortunately, this is in line with the country's high (official) unemployment rate of 32.5% in December 2020. The unofficial rate is even higher. The reader can also see a thick connector between home (`h`) and work (`w`), of which the majority then connects to home (`h`) again at the next echelon. The rapidly diminishing widths mean that most persons have relatively short activity chains, which is in line with the literature. Researchers often choose to only focus on the main activity chain configurations: home-work-home and home-education-home.

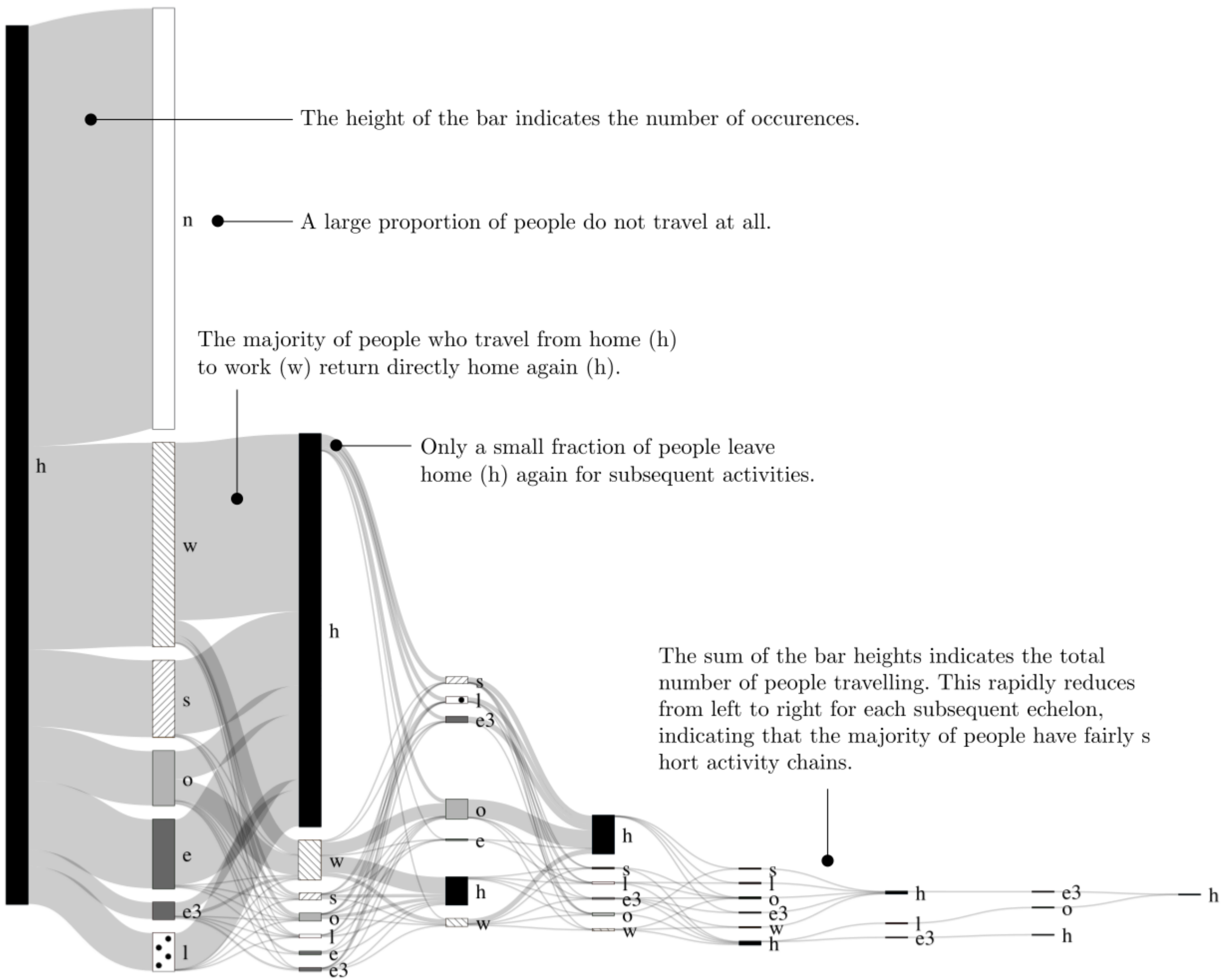

Figure 1: Sankey diagram showing the activity chains of the working age survey respondents. In the diagram, `n` on the first echelon refers to non-travellers (Adapted from Joubert and De Waal (2020)).

Table 2: Encoding scheme for activity types and trip modes.

Activity type		**Trip mode**	
Code	**Description**	**Code**	**Description**
`h`	home	`n`	none
`w`	work	`c`	private car
`e`	education	`p`	public transport
`e3`	dropping/picking up kids at school	`w`	walk
`s`	shopping	`o`	other
`l`	leisure		
`o`	other		

3.2. The discretisation of temporal variables

In addition to the demographic and travel variables described in the previous section, we introduce variables of two temporal dimensions: activity and trip durations. In the diary, respondents report the start and end times of each trip taken. Assuming a day starts and ends at midnight, one can derive the activity and trip durations. While there were a few exceptions, which we removed for this paper's work, most activity chains start and end at home. As opposed to the demographic and travel variables, the time stamped variables are continuous and need to be discretised. Even though BNs can accommodate continuous Gaussian distributions, we decided on the discretisation of temporal variables for the two reasons. Firstly, the synthesis of temporal variables (as oppose to the other chain variables in the network) is not an objective of this paper's model. Therefore, an approximation of these variables is good enough and lowers the computational burden. Secondly, discretisation into time intervals which are intuitive provide valuable insight to the explanations. For example, it is plausible that an activity with duration $t > 8.5$ could constitute a *long* work day.

This paper applied k-means clustering to discretise the durations (Hahsler et al., 2021), using three clusters for the shorter trip durations and four for the longer activity durations. The input data for the activity clustering includes all activity durations, irrespective of their type or positions in the chain. The same holds for trip durations. Table 3 summarises the clustering results. Each discrete duration variable has an additional `none` category to indicate, for example, if an

Table 3: Clustering results to discretise activity and trip durations.

	Duration, t (hours)	
Cluster	**Activity**	**Trip**
1	None	None
2	$0.0 < t \leq 1.5$	$0.0 < t \leq 0.5$
3	$1.5 < t \leq 5.0$	$0.5 < t \leq 1.0$
4	$5.0 < t \leq 8.5$	$t > 1.0$
5	$t > 8.5$	–

activity only consisted of a home-work-home chain, that there was no third trip and hence no associated duration either. For reasons that will become apparent later in this paper, we include two summative variables: one for the daily trip duration and the other for the daily activity duration.

3.3. Structural learning

Once the data is prepared, the BN model can be constructed. When constructing a Bayesian network structure as a directed acyclic graph, the model builder can either machine learn it from data alone, expert knowledge alone, or some combination. Given the described data above, we used a score-based learning algorithm and the minimum description length (MDL) as the score which finds the structure that minimises the MDL (Grünwald, 2005). Once the structure is established, the probabilities are calculated using an EM algorithm (Jensen, 1996).

An exhaustive search over the entire space of possible BNs quickly becomes intractable as the number of possible networks explodes with the number of. The Tabu algorithm (Teyssier and Koller, 2012) is a greedy search method that keeps a collection of previously used operators such as an addition and deletion in memory. These operators are tabu and cannot be reversed in succeeding steps. In addition to the Tabu algorithm, restricting the direction of some arcs also limits the exploding search space. For example, demographic variables can influence travel and temporal variables, but not the other way around.

Travel and temporal variables also cannot affect prior travel and temporal variables and therefore, assumes forward dynamics. The structural learning of the network is unsupervised as the ground truth is not known. The objective is to discover all direct probabilistic relationships in a tree structure (a graph without cycles). In this paper, *BayesiaLab* v9.2.1 (`http://bayesialab.com/`) provides the computational infrastructure to perform the structural learning of the Bayesian network. A similar network structure could also be obtained with libraries in R (R Core Team, 2019; Scutari, 2010). The learning algorithm is score-based and utilised the minimum description length (MDL) as the score. Figure 2 illustrates the resulting network structure.

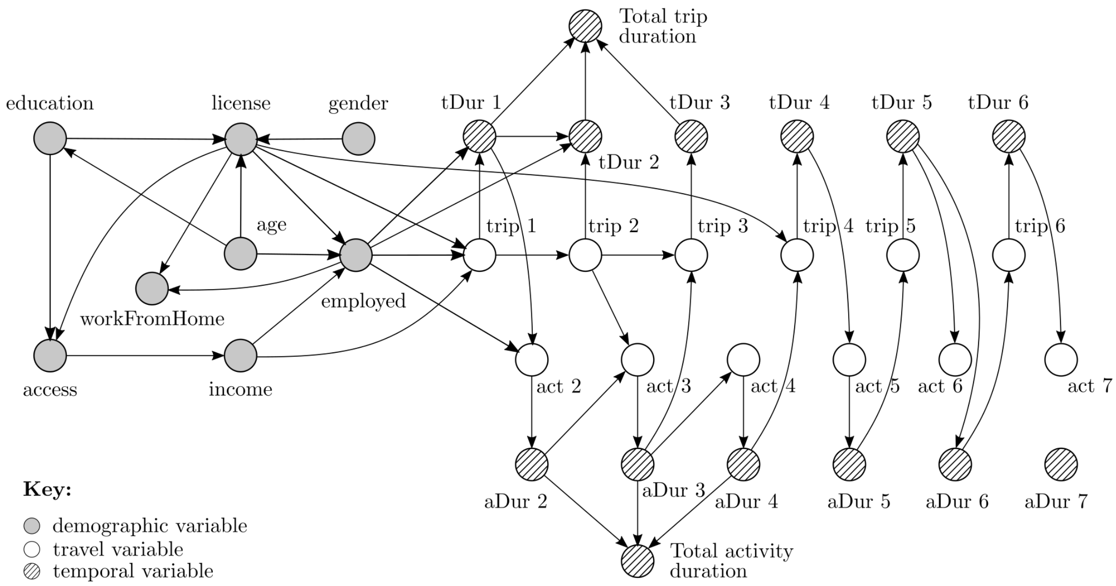

Figure 2: Learned Bayesian network structure with temporal variables included.

The figure suggests that there are recurring patterns, the topic of the following subsection.

3.4. Time-stamped models

When working in a domain that evolves, one can model a discrete component for each time unit to capture recurring patterns (Nielsen and Jensen, 2009). Guo et al. (2019) use the BNs to study the dynamics of people's mobility decisions over time as they consider major life events. In this paper, this unit of time refers to a specific activity (trip purpose) and the subsequent trip and its purpose. For example, a child at school (origin activity), walking (mode) to home where 'walking' represents the trip mode.

We base the recurring patterns on both the learned structure (Figure 2) and Joubert and De Waal (2020). One distinct emerging pattern is the direct links between subsequent trips. This relationship between consecutive trips addresses an essential aspect often identified as a gap in current literature models: choosing a mode early in the trip chain influences later mode choices. For example, if you travel by car to work in the morning, the vehicle has to get back home in the evening.

Secondly, the trip mode influences trip duration. Intuitively this makes sense, as we know a public transport trip takes longer than its private car counterpart. Similarly, the type of activity affects its course. Knowing an activity is of type 'work' allows us to predict that it will take longer than a shopping activity.

In contrast to typical time-stamped models (Nielsen and Jensen, 2009), the number of observations in our travel survey drastically decreases for more extended activity and trip chains. Figure 1 illustrates the diminishment of observations as more and more respondents conclude their activities for the day. Therefore, there are more observations to support the patterns observed earlier in the chain. We acknowledge this diminishing evidence, and we aim to conserve these patterns throughout the chain. The flexibility of BNs to combine data and expert knowledge allows us to adjust the machine-learned structure to support expert inputs. Figure 3 illustrates the time-sliced component inferred. Finally, Figure 4 shows the BN structure resulting from both the data and expert knowledge.

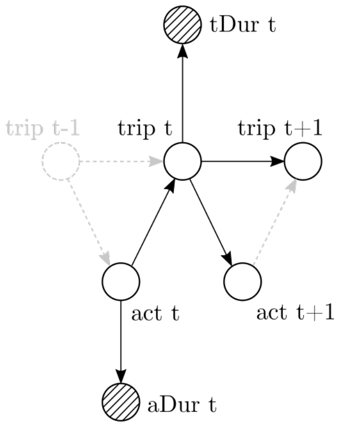

Figure 3: Time-sliced component.

3.5. Model evaluation

To evaluate the model, we investigate each chain independently in terms of accuracy, precision and F-score. We compare the results with those established in Joubert and De Waal (2020) (see Tables 4 & 5). The third column in the tables indicates the cumulative occurrences of the activity

Table 4: Activity chain results. The activities are denoted as `h` (home); `w` (work); `s` (shopping); `e` (education); `l` (leisure); `o` (other); and `e3` (dropping off or picking up children at school).

Chain	Occurrences in test set	Cumulative occurrences	Joubert and de Waal (2020)			Dynamic Bayesian network		
			A	**P**	**F**	**A**	**P**	**F**
`h`	726	49.4%	50.1%	50.0%	48.8%	51.9%	51.4%	50.8%
`h-w-h`	275	68.1%	68.0%	21.0%	20.0%	75.4%	33.9%	33.7%
`h-s-h`	107	75.4%	85.0%	7.6%	7.9%	84.1%	5.6%	6.4%
`h-e-h`	107	82.7%	87.0%	6.7%	7.9%	87.3%	9.2%	8.8%
`h-l-h`	70	87.5%	92.8%	3.2%	4.1%	91.4%	4.8%	4.5%
`h-o-h`	46	90.6%	94.6%	1.9%	3.6%	93.9%	6.1%	6.3%
`h-w-w-h`	20	92.0%	98.2%	0.7%	7.0%	98.0%	0.0%	–
`h-o-w-o-h`	19	93.3%	97.3%	1.2%	5.3%	98.1%	0.0%	–
`h-e3-h`	12	94.1%	98.6%	0.9%	9.9%	98.5%	0.0%	–
`h-o-w-h`	8	94.6%	99.1%	0.6%	12.3%	98.4%	0.0%	–
`h-e3-h-e3-h`	7	95.1%	–	–	–	99.5%	0.0%	–

Table 5: Trip chain results. The trips are denoted as `n` (none); `p` (public transport); `w` (walk); `c` (car); and `o` (other).

Chain	Occurrences in test set	Cumulative occurrences	Joubert and de Waal (2020)			Dynamic Bayesian network		
			A	**P**	**F**	**A**	**P**	**F**
`n`	726	49.4%	50.0%	49.9%	48.6%	51.9%	51.4%	50.8%
`p-p`	193	62.6%	74.7%	15.5%	14.8%	79.2%	23.0%	23.9%
`w-w`	154	73.0%	82.5%	8.9%	9.2%	83.4%	18.3%	17.6%
`c-c`	142	82.7%	82.7%	8.8%	9.4%	85.4%	24.6%	24.6%
`o-o`	109	90.1%	86.6%	7.0%	7.3%	86.4%	6.7%	6.5%
`c-c-c-c`	23	91.7%	97.0%	1.4%	5.2%	97.2%	5.0%	4.7%
`c-c-c`	15	92.7%	97.7%	0.9%	5.5%	98.4%	0.0%	–
`p-p-p-p`	9	93.3%	98.2%	0.8%	7.5%	98.0%	0.0%	–
`p-p-p`	8	93.9%	–	–	–	98.8%	0.0%	–
`w-w-w-w`	8	94.4%	99.2%	0.3%	16.7%	99.0%	0.0%	–
`c-c-c-c-c`	5	94.8%	–	–	–	99.5%	0.0%	–
`o-o-o`	5	95.1%	–	–	–	99.3%	0.0%	–

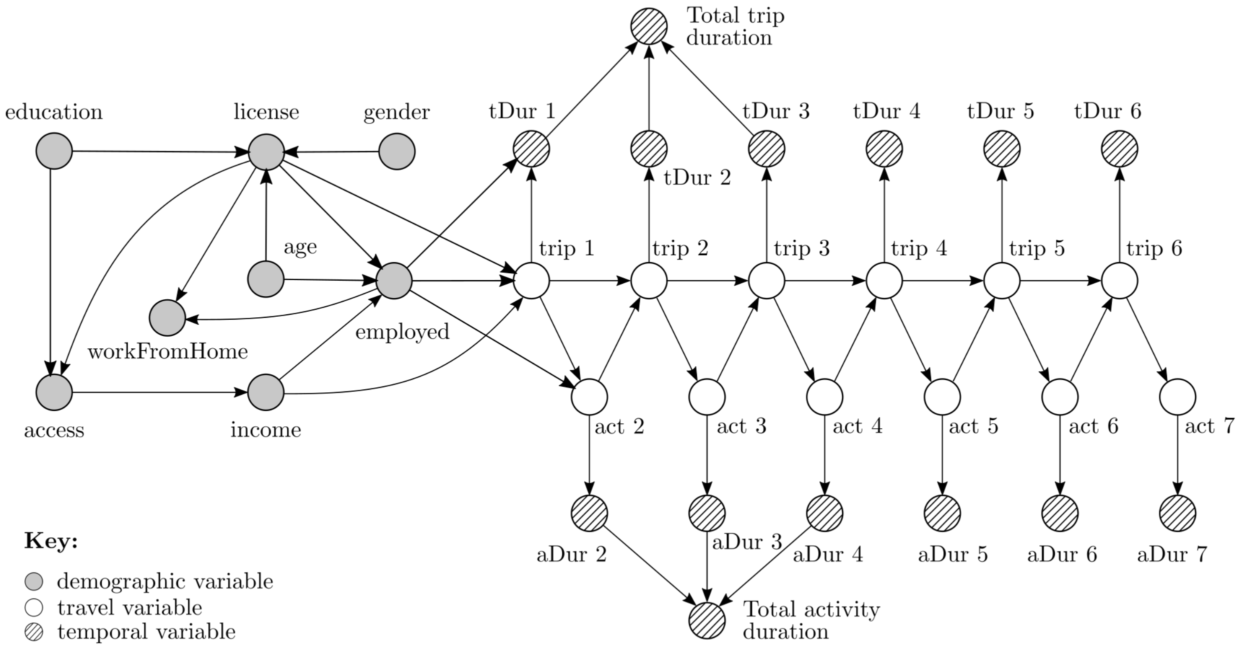

Figure 4: Time-stamped Bayesian network.

or trip chain in the test set. We only calculate the performance metrics of unique occurrences up to ±95% coverage of the test set. It can be seen that the Dynamic BN compare better that the baseline model of Joubert and De Waal (2020) for many of the chain and trip configurations, except for configurations with very few occurrences in the test set. Consequently, the inclusion of temporal variables results in *better* activity and trip chain synthesis performance than the baseline model. This concludes the first objective of the paper.

The temporal variables are stochastic and therefore a metric such as root mean square error (RMSE) between the observed and predicted test cases is appropriate. The calculated RMSE for total activity chain duration is 6.50 hours, and 1.32 hours for total trip duration. These values are very high and can be explained as follows. Many of the observations in the test set contains zero durations because the individuals did not travel. If you synthesise an activity and trip chain for an individual who, in the test set, did not travel, the duration discrepancy is significant. Also, temporal variables are discretised as explained in Section 3.2. When synthesising an activity or trip duration, a value is randomly sampled that is uniformly distributed between the bin extremes. Fewer bins have the benefit of capturing more observations per bin, making the model better specified. Yet, fewer bins are also wider, aggravating the impact of the estimation error when sampling from the wide bin. The objective of the model, however, is not to synthesise or predict temporal variables, but to utilise them as categorical variables for the purpose of *what-if* analysis in the next section. In the case of temporal variables, we move beyond the machine learning objectives of the model towards the second objective of the paper, namely that of explainability which specifically addresses concerns of policy and decision makers.

4. Explanation of evidence in Bayesian networks

One potential concern one would like to address in policy formulation, is the time that people spend out of home. For example, a person working two jobs is likely to spend more than eight hours out of home, thereby potentially leaving or arriving at home in the dark. Furthermore, this duration may increase if a person makes use of public transport, in stead of driving his own car. Lastly, it can be of value to know if a person's last trip is in walk-mode, which might increase the security risk if this takes place during night time.

Contrary to Joubert and De Waal (2020) where the BN was used for synthesising activity chains, this paper is concerned with explaining why some variables are in their particular states, given a set of target variables. This type of explanation falls in the category of active explanations as set out in Section 2. The task at hand is to find the best explanatory hypothesis to explain the observation (also referred to as the evidence set).

In BNs terms, this means that rather than updating beliefs given evidence, we are interested in finding the most probable values, or instantiation, for a set of variables (Korb and Nicholson, 2011). A theoretical framework for explanations in BNs can be found in Derks and De Waal (2020). The paper introduces a taxonomy of explainability in Bayesian networks into the categories of explainability of the model, reasoning, evidence and decisions. Existing machinery includes explaining evidence such as Maximum a Posteriori assignment (MAP) (Pearl, 1988), and most probable explanation (MPE) (Kwisthout, 2011). MAP is the most general method used to calculate the most probable instantiation of a set of n variables by finding the assignment for the n variables that maximises $P(X_1 = x_1, \ldots, X_n = x_n|\boldsymbol{E})$ where $\boldsymbol{E}$ is the evidence (Korb and Nicholson, 2011). The output of MAP is the full joint probability table for many variables and its calculation has been shown to be at least NP complete (Park, 2002). Instead of finding the instantiations of all n variables, MPE finds the instantiation of all the non-evidence nodes. MPE is calculated similarly to probability updating (Nielsen and Jensen, 2009; Korb and Nicholson, 2011). MAP and MPE are full instantiations of the target variables. The limitation of them as explanation methods is that they tend to generate explanations that are or too complex. Users want accurate explanations, but they do not want to be burdened with unnecessary details (Yuan et al., 2011). This begs the question: What constitutes a good explanation? Yuan et al. (2011) argue that a good explanation should consist of two basic properties, namely preciseness and conciseness. *Precise* addresses the accuracy criterion in the sense that the explanation should decrease the surprise value of the set of target variables to be explained. *Concise* addresses the challenge of overspecified explanations and aims to prune less relevant variables from the explanation.

4.1. Generalised Bayes Factor

The next step is to decide on a relevance measure to evaluate the quality of an explanation. Measuring preciseness is intuitive and given that a BN is a probabilistic graphical model, it makes sense that the measure should be probabilistic in nature. Obvious choices are probability and likelihood of the evidence. Likelihood of evidence has been used in calculating most probable explanation. The problem is that a likelihood measure has the property called *irrelevant conjunction* which means that adding an irrelevant fact to the explanation does not change its likelihood (Rosenkrantz, 1994). Likelihood thus addresses preciseness, but not conciseness.

We turn to the Bayes factor, which provides a measure to compare two competing models (Bernardo and Smith, 2009), similarly to testing a full vs. reduced model in classical statistics. They assess evidence for each of the competing models corresponding to a specific hypothesis.

Consider two competing models, M_1 and M_2 with corresponding hypothesis H_1 and H_2. The posterior odds in favour of model M_1 over M_2 is:

$$BF_{12} = \underbrace{\frac{P(M_1|D)}{P(M_2|D)}}_{\text{Posterior odds}} = \underbrace{\frac{P(D|M_1)}{P(D|M_2)}}_{\text{Likelihood ratio}} \times \underbrace{\frac{P(M_1)}{P(M_2)}}_{\text{Prior odds}} \tag{1}$$

The likelihood ratio provides a measure of whether observing D increased or decreased the odds on H_1 relative to H_2 by updating the prior odds (Bernardo and Smith, 2009): $BF_{12} > 1$ signifies that given D, H_1 is now more relatively plausible as oppose to $BF_{12} < 1$ which signifies that the relative plausibility of H_2 has increased. Although these values are relative, a guidance scale was proposed by Jeffreys (1961) in order to determine the significance of the Bayes factor value. The scale signifies $1 < BF_{12} < 3$ as barely worth mentioning, $3 < BF_{12} < 10$ as substantial, $10 < BF_{12} < 30$ as strong, $30 < BF_{12} < 100$ as very strong and $BF_{12} > 100$ as decisive.
If $P(M_1) = P(M_2)$ and the parameter spaces $\Theta_1 = \Theta_2$, then the Bayes factor reduces to the likelihood ratio. The logarithms of the ratios in Equation (1) is referred to as the weight of evidence (WOE) (Good, 1985):

$$\log \text{posterior odds} = \text{weight of evidence } H_1 : H_2 + \log \text{prior odds} \tag{2}$$

A very early application of WOE dates back to the *Banburismus* method devised by Alan Turing at Bletchley Park where the accumulating weights of evidence were a measure in favour for the most likely settings of the Enigma machine (Alexander, 1945).
Moving closer to using the Bayes factor as a relevance metric for explanation in BNs, let us define the following parameters:

$\boldsymbol{M} \triangleq$ a set of target variables in a BN.
$\boldsymbol{e} \triangleq$ the set of evidence on the remaining variables (excluding the $\boldsymbol{M}$).
$\boldsymbol{x} \triangleq$ explanation for the given set of evidence.

Where the Bayes factor is used to compare two hypotheses, Yuan et al. (2011) defined the generalised Bayes factor in order to compare multiple hypotheses.

4.2. Most Relevant Explanation

Yuan et al. (2011) discuss several properties of GBF, of which the key property is that it is able to weigh the relative importance of multiple variables and then only include the most relevant ones in the explanation. The result of this is a definition of a method called *Most Relevant Explanation* (MRE) which uses the GBF as relevance metric to find explanations for a given set of evidence (Yuan et al., 2011):

Definition 4.1. *Let $\boldsymbol{M}$ be a set of target variables, and $\boldsymbol{e}$ be the partial evidence on the remaining variables in a Bayesian network. The most relevant explanation finds an explanation $\boldsymbol{x}$ for $\boldsymbol{e}$ that maximises the generalised Bayes factor score $GBF(\boldsymbol{x};\boldsymbol{e})$, i.e.,*

$$MRE(\boldsymbol{M};\boldsymbol{e}) = \underset{\boldsymbol{x},\emptyset\subset\boldsymbol{X}\subseteq\boldsymbol{M}}{\arg\max}\ GBF(\boldsymbol{x};\boldsymbol{e}). \tag{3}$$

In many cases, one might be interested in not only the singleton explanation which maximises the GBF, but also other competing explanations. This provides additional information and confidence in the best explanation. For example, if the difference between the top and second explanation is very small, the explanations might be very sensitive to input parameters (Yuan et al., 2011). The most obvious approach is to find the top k explanations based on the highest GBF scores. The drawback of this approach is that one might find explanations that are supersets of other explanations and thereby not provide the user with useful alternatives to choose from. In order to produce explanations that are diverse and representative, Yuan et al. (2011) defined filters to remove strongly and weakly dominated explanations from the top k solution set.

4.3. The search space for finding explanations

BNs are a compact representation of a joint probability distribution over a domain, which is beneficial to MRE calculations as MRE can utilise the conditional independencies inherent in a BN to seek explanations more efficiently (Yuan et al., 2011). One can furthermore limit the search space by manually excluding nodes and/or define the search space to only include ancestors of the node or only include the observable Markov Blanket. The Markov Blanket of a node, say A, includes the ancestors, descendants and the nodes sharing a descendant with A (Nielsen and Jensen, 2009). Even when limiting the search space for explanations, the search can be time consuming, especially when the BN contains nodes with multiple states. The software utilised for all BN computations, *BayesiaLab*, contains a built-in MRE feature. This feature uses a genetic algorithm for searching the top-k explanations.
In this section we motivated k-MRE as a method for finding precise and concise explanations for a given set of evidence in a BN. Given the exclusion of explanations which are neither strongly or weakly dominated, the user is presented with a minimal set of explanations which are also diverse and representative. The literature of how to interpret explanations in BNs are few and far between and mostly limited to text book studies. One of the objectives of this paper is to provide guidelines on how to find explanations of evidence in a BN and we suggest the following steps:

1. Define the evidence set.
2. Define the search space (which variables to exclude from the search space)
3. Choose the number of explanations to include in the solution set.
4. Choose to filter strong and/or weakly dominated explanations.

In the next section, we follow these steps in two systematic examples.

5. Case Study

5.1. Motivation

This section demonstrates the quantification of explainability in a policy setting. For this illustration, we use the total time travelling as a proxy for a person's vulnerability. There are three plausible arguments for this choice. Firstly, whereas activity duration may result from one's own will to spend a particular time on an activity, travel time is a function of many attributes out of the decision-maker's control: the circuituity (windiness) of the available bus route; delayed departure time of paratransit; or the number of transfers required to connect the origin and destination. Secondly, in most behavioural research, travelling is assigned a negative utility: it exposes one to the risk of accident or simply the loss of positive utility earned from participating in an activity. In a country like South Africa, the risk of travelling is higher given the country's crime rates and, in particular, gender-based violence. Thirdly, travel time reduces the time an individual spends at home, impeding the ability to accumulate sufficient sleep. Research on sleep deprivation is increasing, and Chaput et al. (2018), in their review article, argue that "[s]leep is not a waste of time and should receive the same level of attention as nutrition and exercise in the package for good health". Consequently, its consideration by and value to policymakers is plausible.

5.2. Experimental Setup

In this subsection, we define the parameters for the proposed steps from Section 4.3, we first define the evidence set.

5.2.1. Define the evidence set

Whereas an extremely long activity duration could be by choice, an extreme long trip is likely *not* by choice and could potentially hold a security risk, increasing one's vulnerability. The purpose of these two case studies is to explain the total trip duration of a person and it is therefore added to the evidence set. In addition to the total trip duration, we also add a socioeconomic profile as evidence. We selected two representative profiles, and will introduce them in the later subsections.

5.2.2. Define the search space

We include demographic variables that are not part of the evidence set, as well as travel variables that relate to activity and mode choice. We omit activity and trip duration variables since they offer little relevant value as policy levers. For example, policymakers have little control over the choice of how long people choose to shop or visit with friends. In the case of trip duration, it is already accounted for in the provided evidence of total trip duration.

5.2.3. Number of explanations in the solution set

As the third step in the proposed methodology, we choose to seek *three* explanations for the individuals' total trip duration. While this number is fairly arbitrary, we argue that since the network was machine learned, the lower ranking explanations are more likely to be sensitive to model parameters.

5.2.4. Explanation filter

Forth and finally, we filter both strongly and weakly dominated explanations. This implies that the resulting explanations are diverse and representative: no explanation is a superset of any other explanation. These are set through parameters configurations in the software of choice, *BayesiaLab*.

5.3. Case study 1: Lilly

The first of the two case studies is based on a fictitious young lady that we will call Lilly. She is a young employed female. Her *evidence set* therefore has a `Female` value for the `gender` variable, an `EarlyCarrer` value for the `age` variable, and a `Yes` value for being `employed`.
If we want to explain the total duration of her trips, we will consider three distinct categories, each associated with a unique evidence set. An *extreme* duration is considered to be longer than 6 hours in a 24-hour cycle, a *long* duration between 1.5 hours and 6 hours, and a *short* duration less than 1.5 hours.

The three most *relevant explanations* found by the k-MRE feature in *BayesiaLab* for each evidence set is given in Table 6.

Table 6: Most relevant explanations for Lilly's total trip time.

Total trip duration	**Solution**	Variables in the search space								**Generalised Bayes Factor**
		Demographic variable				**Travel variable**				
		`income`	`education`	`license`	`workFromHome`	`act2`	`trip1`	`trip2`	`trip3`	
Extreme	1	–	–	–	–	`w`	`p`	–	`p`	152.31
	2	–	–	–	–	`w`	`p`	–	`o`	25.70
	3	–	–	–	`Yes`	–	`p`	–	`o`	23.23
Long	1	–	–	–	–	`w`	–	–	–	11.15
	2	–	–	`No`	`Yes`	–	–	–	`p`	8.07
	3	–	`Secondary`	–	–	–	–	–	`p`	5.95
Short	1	–	–	–	–	`w`	–	–	–	6.74
	2	–	`Tertiary`	–	`Yes`	–	`w`	–	–	4.73
	3	`HighMiddle`	–	–	`Yes`	–	–	–	`w`	4.42

In the extreme trip duration scenario, k-MRE attributes the total trip duration to Lilly going to work directly from home (`act2` being `w`) and using public transport for her first and third trips (`trip1` and `trip3` being `p`). With a GBF of 152.31, this is a decisive explanation.
The second explanation confirms the dependency on public transport, even though it has a significantly lower GBF of 25.70 compared to the first explanation's GBF of 152.31. While the third explanation suggests that the place of work is not as relevant (Lilly can either work from home or at a formal place of employment), it still attributes the extreme trip duration to the presence of a third trip, the first of which was made using public transport. Both the second and third explanations can be considered strong evidence. The presence of work (`w` for the `act2` variable) in many of the explanations is understandable as we specifically gave evidence that Lilly is employed.
For the scenario where Lilly has a long trip duration, the presence of public transport is still prominent as the mode choice for the third trip. The difference between the GBF values are smaller and while the first explanation is considered to be strong evidence, the second and third explanations are substantial. In this scenario we see the first evidence of Lilly's level of completed education to be a determining factor.
What is notable when Lilly has a short total trip duration is that she was able to walk. While this only features in the second and third explanations, the GBF scores among the three top explanation differs only slightly. Also noteworthy is that to explain Lilly's lower vulnerability (short total trip time), she is more likely to be well educated and be exposed to a higher household income.
The trend over the three scenarios suggests that the privilege of good education and access to income allows one the choice of living closer to one's activities. This is particularly relevant in urban environments like South Africa that is still marred by its Apartheid history where low-income earners reside on the periphery of the cities in townships that exhibit limited accessibility (Ziemke et al., 2018). While policymakers may have anecdotal evidence for these facts, the application of MRE provides data-driven confirmation if they aim to confront vulnerability in society.

5.4. Case study 2: Jean

In the second case study, the individual is referred to as a more gender-neutral Jean since gender is not specified. The only demographic variable in the *evidence set* here is that the person is young (`Young` as `age` variable value) and unemployed (`No` as `employed` variable value). The objective is to omit the `gender` variable as evidence and see if it actually features in the explanation.

To address vulnerability, we use the same three evidence categories for total trip duration as we have done for Lilly. The three most relevant explanations found by the k-MRE feature in *BayesiaLab* for each evidence set is given in Table 7. Throughout most of the explanations, the

Table 7: Most relevant explanations for Jean's total trip time.

Total trip duration		Variables in the search space							**Generalised Bayes Factor**
		Demographic variable			**Travel variable**				
	Solution	`income`	`education`	`license`	`act2`	`trip1`	`trip2`	`trip3`	
Extreme	1	–	–	–	`e`	`p`	–	`p`	163.44
	2	–	–	–	`e`	–	`p`	`p`	150.81
	3	–	`Secondary`	–	`e`	–	–	`p`	143.16
Long	1	–	`Secondary`	–	`e`	–	–	`p`	21.13
	2	`Low`	–	–	`e`	–	–	`p`	20.82
	3	–	–	–	`e`	`p`	–	`p`	19.13
Short	1	–	`Secondary`	`No`	–	`w`	–	–	8.40
	2	–	–	`No`	–	`w`	`w`	–	8.36
	3	`Low`	`Secondary`	–	`e`	–	–	–	8.31

presence of education (`e` for the `act2` variable) as the first activity after leaving home is prominent. This is plausible given the demographic evidence for both age and employment.

The strength of all three explanations for the extreme trip duration is decisive with GBF scores exceeding 100). The presence of public transport features prominently, especially for the third trip (`trip3`). Not only does this suggest there *indeed is* a third trip, but also that either earlier trips include may include public transport, or involves multiple transfers to complete the overall journey.

For long trips, the pattern of public transport, especially as the third trip, remains persistent. The GBF values indicate strong evidence (10–30) even though the values are much closer together than for Lilly.

Consistent with Lilly's case study is that short trips are associated with walk (`w`) as chosen mode. Here the lack of a license (variable `license` having a `No` value) is presented in two explanations, as is the fact that Jean has completed at least part of her/his secondary schooling. In two cases we also see that a `Low` household `income` is included in the explanations, once for a long trip and once for a short trip.

5.5. Discussion of results

What is relevant to policymakers is that both case studies suggest that lower vulnerability (proxied by total trip duration) can be explained by walking as opposed to using public transport. This has significant implications, especially in countries (like South Africa) where the lower income population, who are more dependent on public transport, reside further away from the economic centres. Investing in improved public transport (higher frequencies or more services) may rightfully reduce the total trip duration for users. But such improvements may well be fiscal suicide for a country because the lower-income people can already not afford to pay the full cost (or even a representative portion) of the service offering, for two reasons. Firstly, maintaining a network serving outlying areas is much more expensive than the more traditional network that sees denser networks and higher frequencies *close to* the economic center. Secondly, the higher operating cost must be born by lower-income individuals spread over a larger area with lower densities. Consequently, such a network will remain highly reliant on substantive subsidies.

The explanations provided by MRE are insightful as they can also challenge bias. For example, while one may have attributed shorter total trip distance with car ownership and having `access` to one or more vehicle in a household, such a bias is not supported by the evidence. At least not for this data set.

Using data-driven evidence makes a strong case for the alternative policy approach: take the economic activities to the people. While Lilly, with her access to higher household income, may afford the choice to relocate closer to her activities of choice, Jean is less able to. But supporting economic development in the peripheral areas not only addresses the heavy burden of unemployment pervasive in (South) Africa and many developing nations, it also provides employment and other activity options *within walking distance* to the lower income people. As such, economic development and *not* public transport infrastructure may indeed yield better returns on social wellbeing and reducing citizens' vulnerability.

6. Conclusion

The dataset in this paper has served multiple analyses and decision-support efforts. In a previous paper the dataset was used to motivate BNs as a competitive method to synthesise activity chain data and, in doing so, predict people's daily travel behaviour. Since a well-tested BN was developed for this purpose, we decided to further investigate the powerful analyses methods available for BNs. After adding a temporal component to the original BN, we demonstrated how actionable explainability might contribute to policy formulation by means of the k-MRE method. We also provided guidelines on how to find explanation of evidence and interpret the outcome.

One of the limitations of k-MRE, is that it can be sensitive to modelling choices. This is specifically true if the BN is machine-learned and contain multiple states for each variable. Both these conditions are true in our case. To explore this further, one can supplement MRE with a sensitivity analysis which investigate how sensitive the model is to small parameter changes.

Future work includes two specific directions. Firstly, given the insightful explanations from this dataset, it is easy to envisage the insight that might be gained from a more specifically-designed data set or survey. A tailored data-gathering effort may yield more specific insight to a target policy question at hand.

Secondly, we used a combination of open-source and proprietary software for the development of the model and explanation calculations. There is an opportunity to develop open-source code to make the complete XBN workflow more accessible. Especially since BN libraries such as `bnlearn` in *R* already exist (Scutari, 2010).

References

Abdollahi, M., Khaleghi, T., and Yang, K. (2020). An integrated feature learning approach using deep learning for travel time prediction. *Expert Systems with Applications*, 139:112864.

Aghaabbasi, M., Shekari, Z. A., Shah, M. Z., Olakunle, O., Armaghani, D. J., and Moeinaddini, M. (2020). Predicting the use frequency of ride-sourcing by off-campus university students through random forest and Bayesian network techniques. *Transportation Research Part A: Policy and Practice*, 136:262–281.

Alexander, C. H. O. (1945). Cryptographic history of work on the German Naval Enigma. *The National Archives, Kew, Reference HW*, 25(1).

Allaire, J., Gandrud, C., Russell, K., and Yetman, C. (2017). *networkD3: D3 JavaScript Network Graphs from R*. R package version 0.4.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., and Herrera, F. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115.

Bernardo, J. M. and Smith, A. F. (2009). *Bayesian theory*, volume 405. John Wiley & Sons.

Borsuk, M. E., Stow, C. A., and Reckhow, K. H. (2004). A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modelling*, 173(2-3):219–239.

Buchanan, B. G. and Shortliffe, E. H., editors (1985). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading, MA.

Chang, W. (2018). *webshot: Take Screenshots of Web Pages*. R package version 0.5.1.

Chaput, J.-P., Dutil, C., and Sampasa-Kanyinga, H. (2018). Sleeping hours: what is the ideal number and how does age impact this? *Nature and Science of Sleep*, 10:421–430.

De Waal, A. and Steyn, C. (2020). Uncertainty measurements in neural network predictions for classification tasks. In *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, pages 1–7. IEEE.

Derks, I. P. and De Waal, A. (2020). A taxonomy of explainable Bayesian networks. In *Southern African Conference for Artificial Intelligence Research*, pages 220–235. Springer.

Düspohl, M., Frank, S., and Döll, P. (2012). A review of Bayesian networks as a participatory modeling approach in support of sustainable environmental management. *Journal of Sustainable Development*, 5(12):1.

Filmer, D. and Pritchett, L. H. (2001). Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India. *Demography*, 38(1):115–132.

Garvey, C. (2018). A framework for evaluating barriers to the democratization of artificial intelligence. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Geenen, P., Elbers, A., Van der Gaag, L., and Loeffen, W. (2006). Development of a probabilistic network for clinical detection of classical swine fever. In *Proceedings of the Eleventh Symposium of the International Society for Veterinary Epidemiology and Economics*, pages 667–669.

Good, I. (1985). Weight of evidence: A brief survey. *Bayesian statistics 2*, 2:249–270.

Griinwald, P. (2005). Introducing the minimum description length principle. *Advances in minimum description length: Theory and applications*, page 3.

Guo, J., Feng, T., and Timmermans, H. J. P. (2019). Time-varying dependencies among mobility decisions and key life course events: An application of dynamic Bayesian decision networks. *Transportation Research Part A: Policy and Practice*, 130:82–92.

Hagenauer, J. and Helbich, M. (2017). A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Systems with Applications*, 78:273–282.

Hahsler, M., Buchta, C., Gruen, B., and Hornik, K. (2021). *arules: Mining Association Rules and Frequent Itemsets*. R package version 1.6-7.

Hasnine, M. S. and Habib, K. N. (2021). Tour-based mode choice modelling as the core of an activity-based travel demand modelling framework: a review of state-of-the-art. *Transport Reviews*, 41(1):5–26.

Jeffreys, H. (1961). Theory of probability (3rd edt.) oxford university press. *MR0187257*, 432.

Jensen, F. V. (1996). *Introduction to Bayesian Networks*. Springer-Verlag, Berlin, Heidelberg, 1st edition.

Joubert, J. W. (2018). Synthetic populations of South African urban areas. *Data in Brief*, 19:1012–1020.

Joubert, J. W. and De Waal, A. (2020). Activity-based travel demand generation using Bayesian networks. *Transportation Research Part C: Emerging Technologies*, 120:102804.

Korb, K. B. and Nicholson, A. E. (2011). *Bayesian artificial intelligence*. CRC Press, Boca Raton, FL, 2 edition.

Koushik, A. N. P., Manoj, M., and Nezamuddin, N. (2020). Machine learning applications in activity-travel beheavior research: a review. *Transport Reviews*, 40(3):288–311.

Kwisthout, J. (2011). Most probable explanations in bayesian networks: Complexity and tractability. *International Journal of Approximate Reasoning*, 52(9):1452–1469.

Lacave, C. and Díez, F. J. (2002). A review of explanation methods for Bayesian networks. *The Knowledge Engineering Review*, 17(2):107–127.

Leake, D. B. (1995). Abduction, experience, and goals: A model of everyday abductive explanation. *Journal of Experimental & Theoretical Artificial Intelligence*, 7(4):407–428.

Lipton, Z. C. (2018). The mythos of model interpretability. *Queue*, 16(3):31–57.

Lucas, P. J., de Bruijn, N. C., Schurink, K., and Hoepelman, A. (2000). A probabilistic and decision-theoretic approach to the management of infectious disease at the icu. *Artificial Intelligence in Medicine*, 19(3):251–279.

Ma, T.-Y. (2015). Bayesian networks for multimodal mode choice behavior modelling: a case study for the cross border workers of Luxembourg. *Transportation Research Procedia*, 10:870–880.

Nielsen, T. D. and Jensen, F. V. (2009). *Bayesian networks and decision graphs*. Springer Science & Business Media.

Park, J. D. (2002). MAP complexity results and approximation methods. In Darwiche, A. and Friedman, N., editors, *UAI '02, Proceedings of the 18th Conference in Uncertainty in Artificial Intelligence, University of Alberta, Edmonton, Alberta, Canada, August 1-4, 2002*, pages 388–396. Morgan Kaufmann.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rasouli, S. and Timmermans, H. (2014). Activity-based models of travel demand: promises, progress and prospects. *International Journal of Urban Sciences*, 18(1):31–60.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "Why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144.

Rosenkrantz, R. (1994). Bayesian confirmation: Paradise regained. *The British Journal for the Philosophy of Science*, 45(2):467–476.

Scutari, M. (2010). Learning Bayesian Networks with the `bnlearn` R package. *Journal of Statistical Software*, 35(3):1–22.

Sun, L. and Erath, A. (2015). A Bayesian network approach for population synthesis. *Transportation Research Part C*, 61:49–62.

Taroni, F., Garbolino, P., and Aitken, C. (2021). A generalised Bayes' factor formula for evidence evaluation under activity level propositions: Variations around a fibres scenario. *Forensic Science International*, 322:110750.

Teach, R. L. and Shortliffe, E. H. (1981). An analysis of physician attitudes regarding computer-based clinical consultation systems. *Computers and Biomedical Research*, 14(6):542–558.

Teyssier, M. and Koller, D. (2012). Ordering-based search: A simple and effective algorithm for learning bayesian networks. *arXiv preprint arXiv:1207.1429*.

Wand, Z., He, S. Y., and Leung, Y. (2018). Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society*, 11:141–155.

Xinyue, X., Liping, X., Haiying, L., and Qin, L. (2018). Learning the route choice behavior of subway passengers from AFC data. *Expert Systems with Applications*, 95:324–332.

Yagi, S. and Mohammadian, A. (2008). Modeling daily activity-travel tour patterns incorporating activity scheduling decision rules. *Transportation Research Record*, 2076:123–131.

Yuan, C., Lim, H., and Lu, T.-C. (2011). Most relevant explanation in Bayesian networks. *Journal of Artificial Intelligence Research*, 42:309–352.

Zhang, D., Cao, J., Feygin, S., Tang, D., Shen, Z.-J., and Pozdnoukhov, A. (2019). Connected population synthesis for transportation simulation. *Transportation Research Part C*, 103:1–16.

Zhu, J., Liapis, A., Risi, S., Bidarra, R., and Youngblood, G. M. (2018). Explainable AI for designers: A human-centered perspective on mixed-initiative co-creation. In *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 1–8. IEEE.

Ziemke, D., Joubert, J. W., and Nagel, K. (2018). Accessibility in a post-apartheid city: Comparison of two approaches for accessibility computations. *Networks and Spatial Economics*, 18(2):241–271.