# Activity-based travel demand generation using Bayesian networks*

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

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

**Abstract**

While activity-based travel demand generation has improved over the last few decades, the behavioural richness and intuitive interpretation remain challenging. This paper argues that it is essential to understand why people travel the way they do and not only be able to predict the overall activity patterns accurately. If one cannot understand the "why?" then a model's ability to evaluate the impact of future interventions is severely diminished. Bayesian networks (BNs) provide the ability to investigate causality and is showing value in recent literature to generate synthetic populations. This paper is novel in extending the application of BNs to daily activity tours. Results show that BNs can synthesise both activity and trip chain structures accurately. It outperforms a frequentist approach and can cater for infrequently observed activity patterns, and patterns unobserved in small sample data. It can also account for temporal variables like activity duration.

*Keywords:* Tour generation, activity-based, travel demand, activity choice

## 1. Introduction

Agent-based models, a specific variant of activity-based models, are becoming popular in transport and mobility planning as it allows the study of large urban systems with rich features. The agent-to-agent and agent-to-environment interactions allow for emergent phenomena that help planners identify both intended and unintended consequences. Such models require detailed descriptions of the population and their travel demand.

Existing approaches to generate the daily activity patterns is data-intensive and, consequently, often rely on inferring behaviour from small data sets. Inference includes the structure of the activity sequence, mode choice, and activity timing and location. Existing approaches make utilitarian assumptions that are neither behaviourally rich nor intuitively accurate. Many of these approaches also fail to take the dynamics into account. For example, if a person uses a car for early stages of the activity chain, s/he is likely to use it again later in the chain.

The first contribution of this paper is methodological, introducing Bayesian networks (BNs) as a viable alternative in synthesising activity and trip chains when only given a person's demographic variables. BNs present an opportunity to study the causal relationship between variables and,

*This paper has been published. Please cite as Joubert, J.W., De Waal, A. (2020). Activity-based travel demand generation using Bayesian networks. *Transportation Research Part C: Emerging Technologies*, **120**, 102804. DOI: 10.1016/j.trc.2020.102804.

ultimately, counterfactuals. The focus of this paper is not on the explanatory value of BNs but rather its novel application to travel demand synthesis. The second contribution of this paper is the performance metrics, adapted from well-established approaches in machine learning, used to evaluate the synthesised activity and trip chains. Instead of aggregate metrics such as overall frequency of specific activity types, this paper delves into more detailed metrics that answer the question "how well did we get it right for *this person*?"

The results show synthesised activity and trip chains with a high level of accuracy. Given the uncertainty and erratic behaviour of individuals with regards to their travel patterns, a BN allows one to create plausible activity chains never observed in the original data. This generative attribute is especially valuable when having to rely on small data samples.

The most common goal for artificial intelligence (AI) is the ability to perform advanced cognitive tasks at or beyond the human level of performance. Most production-level AI are highly specialised, achieving exceptional prediction performance. But it frequently focuses on one job only. Instead, we are interested in more than prediction. Our goal is to further our understanding of the nature of an AI model: why was a particular performance achieved?

Furthermore, we are interested in how the performance will change for alternative scenarios and, ultimately, how it will respond to interventions. If we want to move to the next level of AI, we need to consider and evaluate interventions and, even better, counterfactuals. BNs provide a modelling platform which can respond to stochastic and changing environments and cope sensibly with varied sources of uncertainty (Korb and Nicholson, 2011).

The next section reviews the literature on recent advances in activity-based travel demand and the timely opportunity for BNs, given its current applications. The remainder of the paper follows the workflow depicted in Figure 1. Section 3 describes the data set derived from a travel diary

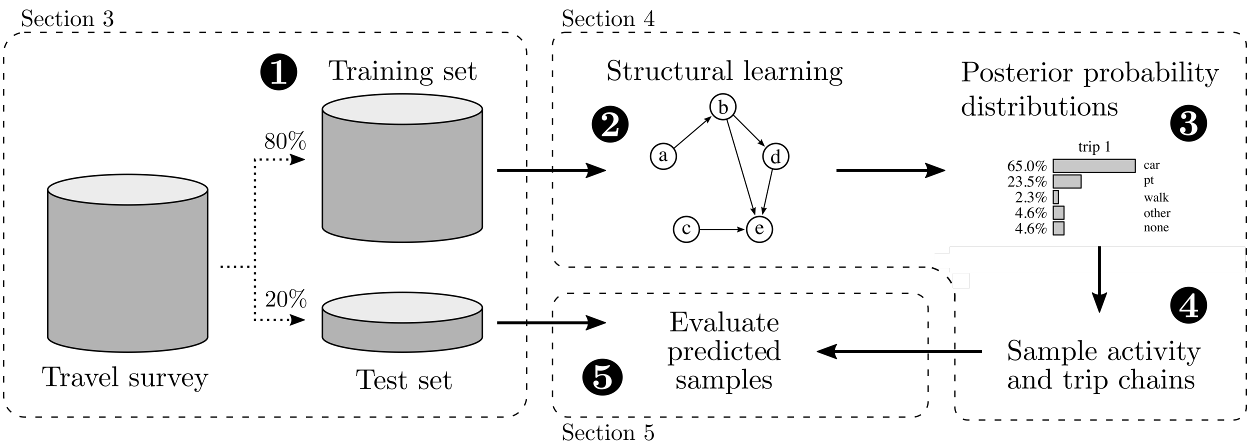



Figure 1: The workflow to generate travel demand from a Bayesian network.

in Cape Town, South Africa. Section 4 then uses the training set to learn the Bayesian network structure and estimate the posterior probabilities, ending with a description of how travel demand is predicted from the BN. Results, presented in Section 5, show how the synthesised travel demand, both activity and trip chains, compare against the test set. The paper demonstrates in Section 6 that one can include temporal variables like activity duration.

## 2. Literature review

The review here has two main components. The first portion looks at recent advances in activity-based travel demand generation, and specifically at the tour-based mode choice models. Secondly, we introduce a brief review of BNs and show how it allows addressing two shortcomings in the existing body of knowledge: the dynamic nature of mode choice; and the ability to understand causal relationships.

### *2.1. Activity-based models*

Compared to trip-based models, the activity-based approach aims to represent the travel demand of an individual over an entire 24-hour period: the sequence of activities throughout the day, its location and duration, and the mode-specific trips connecting the activities. The data burden is high and often rely on surveying the detailed movement of people, referred to as *travel diaries*, in which people report their activity participation and movement over 24 hours, or longer.

Travel diaries reveal very diverse behaviour. Each person's socio-demographic background, lifestyle and needs often influence their travel choices. Because the surveying burden to collect travel diaries is high, researchers are often left with quite small samples to infer travel behaviour. Consequently, simplifying modelling efforts sees clustering the data into homogenous daily activity patterns, and it remains an ongoing area of research (Hafezi et al., 2019). Guzman et al. (2017) also simplify activity chain generation to focus only on short home-work/study-home and home-other-home activity chains.

In their review, Rasouli and Timmermans (2014) argue that although activity-based models have matured and addressed many of the classical four-step model shortcomings, several boxes still need to be ticked. Their review distinguishes between three main classes of models emerging over the past two decades. The first is *constraint-based* models. These do not generate or predict household and individual activity patterns, but rather check if a given agenda (activities and trips) is feasible within a specific space-time context.

The second class is the popular *utility-maximising* models. Using discrete choice, these (nested) logit models consider decision-stages at multiple tiers. Work like that of Yagi and Mohammadian (2008) implements a multi-level model to generate daily activity patterns. Theirs is a system of random-utility-based, disaggregate logit and nested logit models in a decision hierarchy. First, there is the choice of the daily activity pattern, then the time-of-day choice (including activity duration), and finally the mode and destination choice. These models assume that the utility function be specified a priori by the modeller and they acknowledge that model specifications vary significantly between countries, especially between developed and developing nations. Interpreting the results and estimated coefficients is also not very intuitive.

Ma (2015) argues that despite the popularity of discrete choice models, they have inherent limitations. The most prominent of these limitations is that they are based on rational choice theory and assumes a decision-maker has perfect and complete information about all the possible alternatives. Also, it assumes the decision-maker has time and cognitive capacity to evaluate all alternatives and (deterministically) picks the one that provides the maximum utility.

The third class identified by Rasouli and Timmermans (2014) are *computational process* models. They argue that this class of models emerged to relax the strict and behaviourally unrealistic assumptions of utility-maximising models and allow for more natural decision-making under uncertainty. The rule-based approach depicts the more realistic decision heuristics of individuals. However, they too have the limitation that rules rely on a priori assumptions of the researchers and model builders. Some models in this class, like TASHA (Roorda et al., 2008), draw the daily plans and its attributes from empirical distributions.

Despite the importance of activity-based modelling systems, there is a lack of reliable evaluation criteria to check if synthesised daily plans are good representations of reality. Many modelling efforts only interpret and analyse the estimated parameters, avoiding synthesis and subsequent testing. To overcome this problem, researchers have turned to bio-informatics inspired sequence alignment methods (SAMs) that evaluate synthesised protein sequences. SAMs assess activity-based models by pairwise comparing the observed sequences extracted from the diaries and their synthesised counterparts. However, since there is much variance in an individual's activity behaviour across different days, the methods are often not a true reflection of the model performance. Or, at best, are very conservative metrics.

Liu et al. (2015) introduce a profile Hidden Markov Model (pHMM) to improve on SAM in two ways. Firstly, SAM alone cannot capture infrequent activities and, secondly, pHMM has position-specific metrics that analyse the entire length of the activity sequence and enhance the rigid pairwise comparison of SAM. The two-stage stochastic approach of a basic Hidden Markov Model starts by identifying the finite set of states for each cluster of activities at each position in the activity sequence. Then the transition probabilities are being derived. Saadi et al. (2016) combine their population synthesis with the pHMM, allowing them to use the Belgian travel survey to generate individuals with their activity sequences assigned.

Rasouli and Timmermans (2014) conclude their review noting that activity-based models suffer from both model uncertainty due to the sampling with the models, and also input uncertainty as a result of its reliance on small empirical data sets. Their call is for research to move away from utilitarian models towards models that are behaviourally rich, allowing the study of causality and incorporation of imperfect perception, which plays a significant role in human decision-making under uncertainty.

### *2.2. Bayesian Networks*

BNs have the ability to synthesise multiple information types such as empirical data, sensors, expert knowledge and literature. This makes it particularly useful in ecological modelling as emerging patterns on multiple scales can be fused into one systems-level model (Borsuk et al., 2004). This integrated model equip stakeholders with an appreciation of how policy decisions propagate across scales through the system. From a different perspective, it allows stakeholders to optimise desired outcomes, given certain ecological scenarios (Bromley et al., 2005; Düspohl et al., 2012).

Closer to the focus of this paper, BNs have only very recently been used in the domain of transport planning (Sun and Erath, 2015; Feygin and Pozdnoukhov, 2018; Borysov et al., 2019; Zhang et al., 2019). BNs are successfully applied to population synthesis. Sun and Erath (2015) generated synthetic populations for Singapore that are accurate at both household and individual level. The value lies in the joint probability distributions that are estimated, for example the age of the head of the household and their spouse. A number of studies replicate (Joubert, 2018) or build on their work. Sun et al. (2018) extend their earlier work and introduce latent variables to further refine the population synthesis procedure. Borysov et al. (2019) argue that some of the recent machine learning approaches, in population synthesis, are not very scalable and they introduce a deep generative modelling approach that promises to cater for many more variables to capture a high level of detail in terms of personal details and travel preferences. Zhang et al. (2019) augment the BN by using call detail records, allowing them to connect the individuals in the population.

Ma et al. (2017) apply BNs to model travel mode choice behaviour. They argue that, as an alternative to the decision tree approach proposed by Arentze and Timmermans (2007), BNs can explicitly estimate the causality structure between variables. This, in turn, leads to more flexible relations and more intuitive interpretation of the relationships than a decision tree structure.

If the structure of the daily activity plan is known, then specific choice dimensions can be more easily modelled. In recent work, Ma and Klein (2018) use BNs to perform location choice of discretionary activities when a space-time prism is used to limit the location choice set. They use BNs to identify the most likely heuristic the decision-maker will use to choose the detour factor and location. To our knowledge this paper is the first attempt to apply BNs to synthesise daily activity patterns.

## 3. Data

This research benefits from a travel survey that the City of Cape Town conducted during 2013. The study sampled a total of 22 332 households, accounting for 63 175 individuals and representing 2.1% of the 2011 Census total (Transport for Cape Town, 2013). The questionnaire solicited personal information about household size, age, gender, income, vehicle ownership and mobility impairments. A subset of households (2 974 after cleaning the data) in the Cape Town survey also completed a travel diary in which each member of the household reported their detailed activity chain of the previous day.

### *3.1. Demographics*

Although varied, and often inconsistent, the literature suggests a variety of variables that influence the travel behaviour of individuals. In this paper, both the literature and local knowledge of the South African context inform the choice of demographic variables. Table 1 summarises the complete encoding scheme and distinguishes between individual and household characteristics. The first variable, `gender`, is straightforward. For this paper, we distinguish between three subgroups based on the age of the individual, using the Census 2011 age categories. Those that are 5 years or younger are considered an *infant*. Individuals from 6 to 12 years are classified as a *child* and represent those typically attending primary school — people in the 13 to 23 age range fall in the *young* class. Since the secondary school phase ends for the majority of learners in the 17 to 19 age range, the first subgroup, `children`, contains all individuals in the *child* and *young* categories. This study ignores infants as they do not travel independently.

Individuals in the 24 to 45 age bracket fall in the *early-career* class and those in the 46 to 68 age bracket fall in the *late-career* class. The second subgroup in our data, `workers`, includes all individuals in the *young*, *early-career* and *late-career* class. There is, indeed, an overlap with *young* falling in both subgroups. The argument is that it is an age group that includes many individuals that leave secondary school early for the job market (`workers`) and those that complete their secondary and maybe even tertiary education (`children`). The third subgroup, `retired`, contain all individuals older than 68. The `children` subgroup has 2281 records; the `workers` has 7507, and the `retired` has 1031 records.

The third individual variable deals with the level of education the individual completed, and the fourth deals with having a driver's license. The variable `access` reflects the number of private vehicles the individual has access to, either as the owner, driver, or passenger.

An individual is considered to be employed when s/he reported to be self-employed, a seasonal worker, employed part-time or fulltime. A person's `travelForWork` variable affirms if the person's

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 block of variables relates to individual characteristics, and the second block to household characteristics.

| Code | Description | Values |
|---|---|---|
| `gender` | Gender | `Female` (4 812); `Male` (4 373) |
| `ageGroup` | Aggregated age group based on Census categories | `Infant` (20); `Child` (690); `Young` (1 591); `Early career` (3 633); `Late career` (2 283); `Unknown` (473) |
| `edu` | Completed education | `None` (999); `Primary` (1 794); `Secondary` (5 238); `Tertiary` (1 217) |
| `license` | Whether the individual has a driver's license | `Yes` (3 244); `No` (6 004) |
| `access` | The number of vehicles the person has access to | `None` (4 264); `Single` (2 541); `Multiple` (2 443) |
| `employed` | Whether the individual is employed | `Yes` (3 959); `No` (5 289) |
| `travelForWork` | Whether the individual's work is, effectively, driving or travelling | `Yes` (349); `No` (8 206) |
| `workFromHome` | Whether the person works from home | `Yes` (321); `No` (8 342) |
| `assetOne` | Calculated asset value | `Low` (2 437); `LowMiddle` (5 601); `HighMiddle` (782); `High` (428) |
| `family` | Household size | `Single` (407); `Couple` (1 388); `Small` (4 880); `Large` (2 573) |
| `housing` | Type of housing for the main dwelling | `Informal` (867); `Formal` (8 381) |

work dictates that driving is an essential part of the work activity itself, for example, being a driver. If a person's work is home-based, or s/he is allowed to perform their work from home, the `workFromHome` variable reflects that.

The first household variable, `assetOne`, represents a household's income. Since many households choose not to provide this information, the travel survey accommodates this 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 16.70 and USD 1 = ZAR 15.00 that equates to approximately EUR 192, or USD 213). The ceiling for the next income class, `LowMiddle`, is ZAR 25 600 per month (EUR 1 533 or USD 1 707); for `HighMiddle`, it is ZAR 51 200 per month (EUR 3 066 or USD 3 414); and high income is anything higher.

The unique record reflects both the household id and the member number. Consequently, one can easily calculate household size which is, in line with the Jakarta study of Yagi and Mohammadian (2008), a useful characteristic in developing countries where households tend to be larger. The couple category reflects a two-member household. Contrary to what the name suggests, this category also represents single-parent households with one child. A small family is considered to be 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 the formality of a household's main dwelling affects the travel demand.

### *3.2. Travel demand*

Travel-specific questions included locations of home and work activities, typical weekday trips, mode of travel, cost of travel and time taken per trip, including walking and waiting time. For each activity, the interviewer captured the location (at the Census main-place level) and the start and end times. For each trip, they captured the mode. After cleaning the data, the travel diary component represented the revealed activity chains of 9248 individuals from 2974 households — Table 2 reports on the encoding used for both activities and trip modes.

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

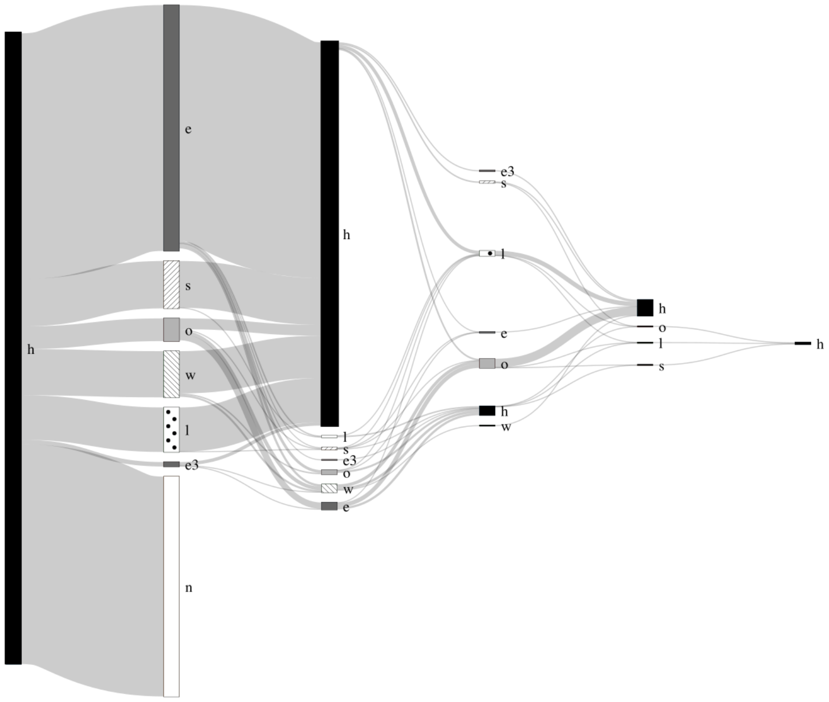
| **Activity type** | | **Trip mode** | |
|---|---|---|---|
| **Code** | **Description** | **Code** | **Description** |
| `h` | home | `n` | none |
| `w` | work | `c` | private car |
| `e` | education | `p` | public transport |
| `e3` | dropping/picking up kids at school | `w` | walk |
| `s` | shopping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

Some simplifications were made in this study to reduce the number of variable classes. Firstly, we combined primary and secondary (`e1`) with tertiary education (`e2`) into simply *education* (`e`). Secondly, recreational visits to friends (`v`) were recoded as *leisure* (`l`) activities. Thirdly, activities with a medical purpose (`m`) were combined with other activities (`o`).

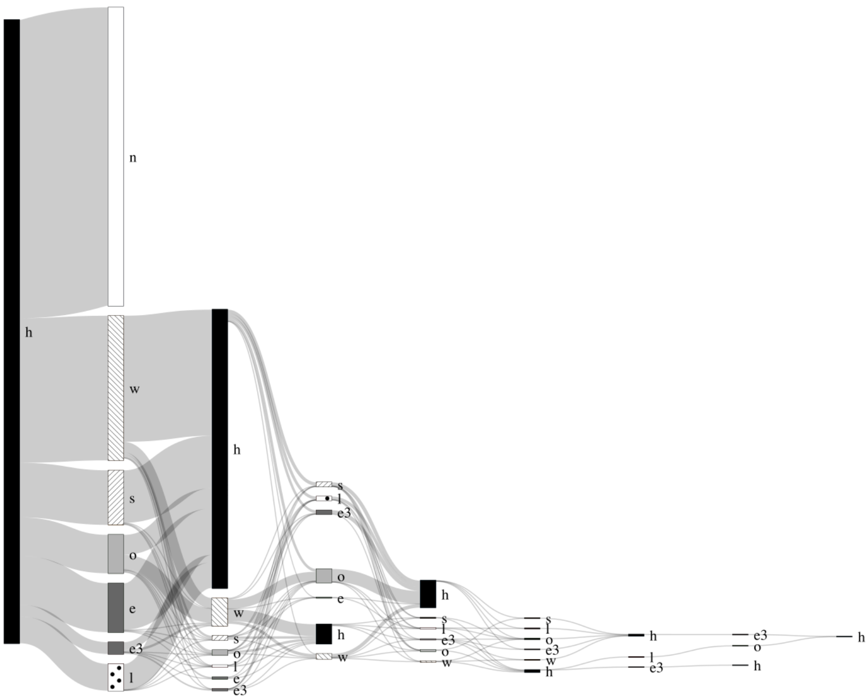
On the trip side, we joined all modes of public transport (`p`). The different public transport offerings in the City of Cape Town include the local authority's *MyCiTi* Bus Rapid Transit (BRT), the provincial commuter bus services operated by *Golden Arrow Bus Services*, the national rail services provided by *Metrorail*, and the paratransit service known in South Africa as *minibus taxis*. The minibuses are 15 to 25-seater vehicles and are the primary public transport mode, accounting for just more than 50% of all public transport trips.

As an illustration, Figure 2 uses Sankey diagrams (R Core Team, 2019; Allaire et al., 2017; Chang, 2018) to show the proportions of people having specific activity chain configurations and distinguishes between the `children` (Figure 2a) and `workers` (Figure 2b). Consider the `children` diagram in Figure 2a. One can infer from the diagram that approximately 35% of children went from home (`h`) to school (`e`). The thickness of the grey line connecting the home (`h`) bar at the far left to the education bar (`e`) second from left represent the proportion. From those that went to school, the majority then went back home (`h` in the third column of activities). One can represent the activity chains of these individuals by the string `h-e-h`. A small proportion, about 5%, left school for a variety of other activities. Quite a large proportion, ±35%, stayed at home all day as their only other activity is *none*, encoded as `n`. This worrying figure is, unfortunately, commensurate with the high youth unemployment rate in South Africa.

Comparing the two subfigures, one can observe that `children` had simpler (shorter) activity chains with, at most, 6 activities in the chain, while `workers` had, at most, 9. Lines become thinner as one progress from left to right, representing smaller proportions of people with longer




(a) Activity chains of the `children` subset.




(b) Activity chains of the `workers` subset.

Figure 2: Sankey diagram showing the activity chains. In the diagram, `n` on the first echelon refers to non-travelers.

and more complex activity chains. This is aligned with Liu et al. (2015) who confirm that, in human behaviour, one often see higher frequencies of shorter activity sequences.

### 3.3. Training and test sets

The remainder of this paper is only concerned with `workers` and leaves the travel demand prediction for the `children` and `retired` subgroups for future work. While this paper distinguishes between these three subgroups, nothing prohibits a person to extract different subgroups along spatial or demographic attributes and estimate unique Bayesian networks for each. Comparing the resulting networks may then allow decision-makers to evaluate policy implications for more targeted (sub)groups.

As is common in machine learning when one builds a data-driven model to predict, one would split the travel survey data into a training set and a test set.

Each record in the data set represents an individual: its demographic variables and travel demand. The sampling regime applied in this paper does not keep households intact but samples 80% of the records, each with equal probability, into the training set. The balance makes up the test set. Table 3 provides an overview of the representation of unique demographic profiles in the test set. When only taking into account the four variables directly influencing the travel chain, a total of 36 unique profiles were found in the test set. We only display the top 10 and bottom 5 unique profiles. Interestingly, in the top 3 demographic profiles, individuals don't have access to a vehicle, or have a driver's licence.

Table 3: Unique demographic profiles and representation in the test set.

| **Demographic Profile** | | | | |
|---|---|---|---|---|
| `access` | `licence` | `ageGroup` | `employed` | **counts** |
| None | No | Early career | Yes | 172 |
| None | No | Early career | No | 147 |
| None | No | Young | No | 135 |
| Multiple | Yes | Early career | Yes | 118 |
| None | No | Late career | No | 106 |
| Multiple | Yes | Late career | Yes | 92 |
| Single | Yes | Early career | Yes | 81 |
| Single | No | Young | No | 61 |
| Single | No | Early career | Yes | 57 |
| None | No | Late career | Yes | 52 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| None | Yes | Early career | No | 5 |
| Multiple | No | Late career | Yes | 3 |
| None | Yes | Young | Yes | 3 |
| Single | Yes | Young | Yes | 1 |
| None | Yes | Young | No | 1 |

Congruent with the workflow depicted in Figure 1, the Bayesian network structure is learned using only the training set.

## 4. Bayesian networks

Bayesian networks (BNs) are directed acyclic graphs (DAGs) containing nodes and arcs. The nodes represent quantities of interest, or variables, and the arcs quantify the relationships – which could be causal, but not necessarily – between the arcs (Jensen, 1996). During the modelling phase of a BN, the structure of a BN can be elicted from experts, or it can be machine learned (Korb and Nicholson, 2011). Similarly, parameterisation can be done by experts or machine learned. This makes BN a good modelling choice in data sparse problem domains as expert knowledge can supplement the data. Even when data are plenty, the expressive representation of conditional dependencies in BNs provide transparency and insight into model predictions.

Structural learning is unsupervised, as the ground truth is unknown. The objective of structural learning is to discover all direct probabilistic relationships in a tree structure (a graph without cycles). Two families of structural learning are constraint-based and score-based learning. Constraint-based learning aims to optimise the qualitative aspect of the BN structure and score-based learning aims to optimise the quantitative aspect.

### *4.1. Score-based structural learning*

We focus on score-based learning and use the minimum description length (MDL) as the score: that is, find the structure that minimises the MDL. MDL operates under the logic that regularities within the data can be compressed, meaning certain symbols can be used to describe the data in a more compact way than the actual data. Highly regular data can therefore be highly compressed (Grünwald, 2005). MDL consists of two parts:

$$MDL(B, D) = DL(B) + DL(D|B) \tag{1}$$

where $B$ is the model and $D$ is the observed data. $DL(B)$ represents the structural complexity (number of bits) of the suggested model and $DL(D|B)$ represents the model's capacity to encode data, which is none other than the error (Grünwald, 2005). Equation 1 states that the MDL score is the sum of the complexity of the model and the complexity of the errors, which the learning algorithm aims to minimise simultaneously. Generally, if the model is highly accurate it will need much description (as it will have many terms) but the resulting error will be small. Conversely, if the model is very simple, its description will be very short but we will need a lot of information to describe its errors. A fully unconnected network will translate to the minimum value for $DL(B)$ and a fully connected network will translate the minimum value for $DL(D|B)$. Thus obtaining minimum MDL finds the right balance between the two extremes.

A search algorithm explores the space of possible BNs to find a structure that minimises the MDL score. The challenge is that the number of possible networks explodes with the number of nodes and an exhaustive search becomes intractable. Search algorithms use operators such as arc addition, deletion, and reversal to filter across the search space in order to find the structure that optimises the MDL score. Apart from the large computational cost associated with the search, it also suffers from local minima. The Tabu algorithm (Teyssier and Koller, 2012) is a greedy search method that keeps a collection of the previously used operators in memory. These operators are *tabu* and cannot be reversed in succeeding steps. The combination of a greedy search with a deterioration of the MDL score guides the search algorithm out of the local minima. The search space can be further constrained by forbidding certain directed links. These restrictions can be based on causality, or time dependency. For example, the node `gender` cannot be conditioned by

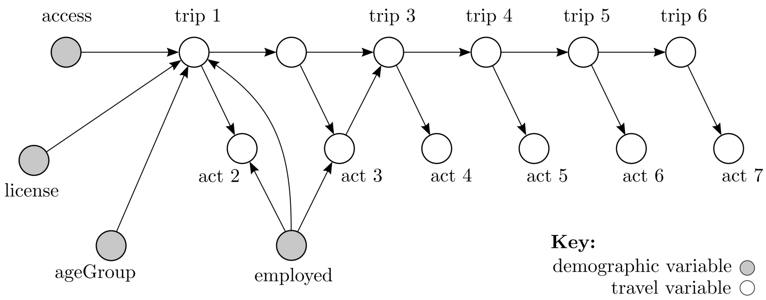



Figure 3: Learned Bayesian network structure.

the node `licence`. Having a licence does not influence the probability to be a certain gender, but it can be the other way around. Similarly, the third trip in the activity diary cannot influence the first or second trip.

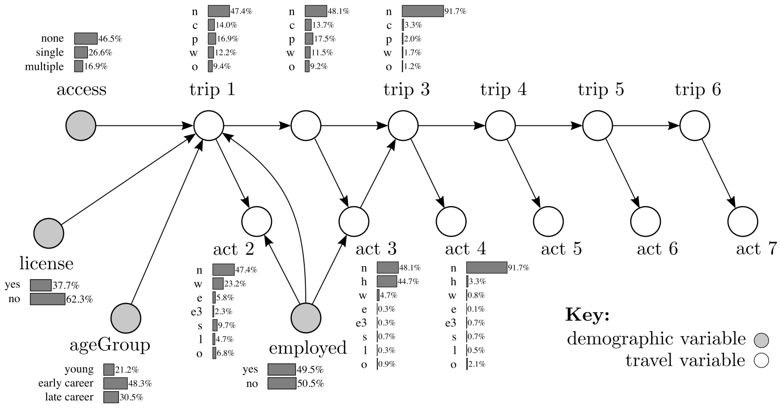
### *4.2. BN structural learning setup*

In line with the data types described in Section 3, two classes of variables were created: *demographics* and *travel demand*. The demographics class contains 11 variables pertaining to personal information and the travel demand class contains a `trip` and `activity` variable for each step in the chain. We model 6 chain steps as the last 3 steps contain only very few observations as indicated in Figures 6d and 7d. Each of the 23 variables is represented in the BN as a node. The mutual exclusive states for each node are the values listed in Tables 1 and 2. The Tabu algorithm is applied to the training set[1].

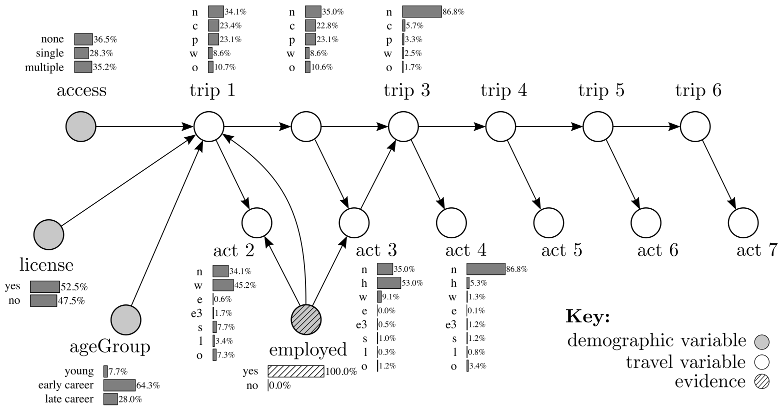
The output of structural learning is directed links between nodes with associated conditional probability distributions. Figure 3 illustrates the Bayesian network structure as machine-learned from the input data alone; without expert input. An interesting observation is that only four demographic nodes are directly linked to the travel chain: `access`, `license`, `ageGroup` and whether you are `employed` or not. `Access`, `license`, `ageGroup` influence the first trip's mode of transport, denoted by `trip 1`, and `employed` influences `act 2` and `act 3`, which describe the second and third activity types. Intuitively this makes sense: whether the person has access to a vehicle, a license and his/her age group influences what type of transport the person will take. Whether the person is employed or not is the most influential variable on the activities the person will undertake. Furthermore, they only influence the start of the travel chain, which from thereon is linked sequentially. This too is insightful as it captures the known phenomena that is often hard to capture in existing mode choice models. For example, if you went to work by car, the car is already there and you are subsequently much more likely to *again* use the car for later trips.

Figure 4a illustrates the posterior distributions associated with each node. At this stage, no evidence (also called instantiation) is entered into the network and the posterior distributions merely resemble dataset frequencies. In Figure 4b, the node `employed` is instantiated to `yes`. The evidence `employed` = `yes` gets propagated throughout the network and all other probability distributions downstream from that node are updated. This is clear in the node `act 2` in which the probability for `act 2` = `w` is updated from 23.2% (Figure 4a) to 45.2% (Figure 4b). Again,

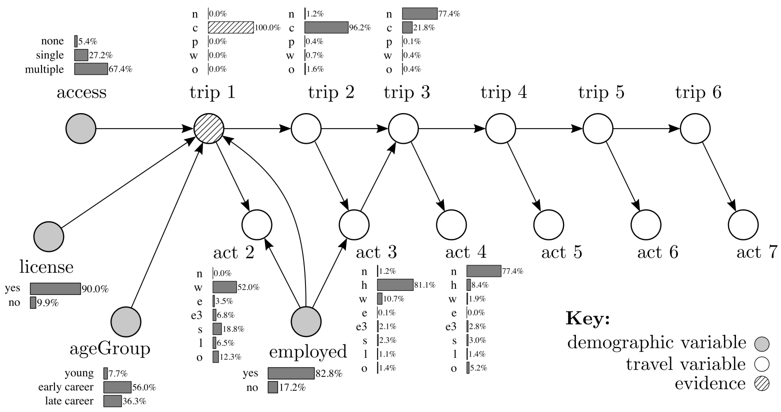
[1]All BN modelling are performed in Bayesialab v9.2.1 (http://bayesialab.com/)




(a) Initial posterior distributions.




(b) Posterior distributions with node `employed` instantiated.



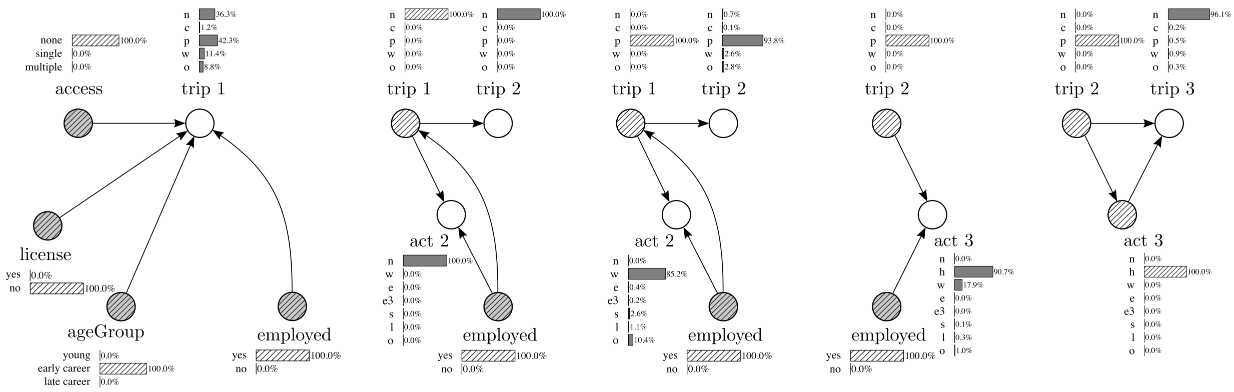

(c) Posterior distributions with node `trip1` instantiated.

Figure 4: Illustrations of BN posterior marginal posterior distributions (initial and instantiated)

this is intuitive as knowing that someone is employed increases the probability that `w` will appear as the first activity in the activity chain. In Figure 4c we instantiate the node `trip 1` with car, `c`, and it has has a similar interpretation: if it is known that `trip 1` was made by car, `c`, then the certainty that the person is employed is 82.2%. Furthermore, it is most likely (90%) that the person has a license.

### *4.3. Sampling synthetic chains from the BN*

Although inference insights as described in Section 4.2 is of great interest, it is not the main objective of this research. We are interested in sampling synthetic chains which propagated from unseen demographic profiles. The BN model provides us with posterior probability distributions to sample from in the following way. Consider the example illustrated in Figure 5 for the travel demand of an early-career female, say Lilly, who is employed but who has neither a license nor access to a private car. Instantiated variables—those for which we have evidence—are illustrated with diagonal shaded lines as per the key in Figure 4. A demographic profile is entered as evidence into the BN (Figure 5a). This evidence gets propagated throughout the rest of the network by




(a) Step 1. (b) Alternative 1 for step 2. (c) Alternative 2 for step 2. (d) Step 3. (e) Step 4.

Figure 5: Sequential sampling from the Bayesian network.

updating the posterior probabilities accordingly.

Given Lilly's demographics and the fact that all activity chains start at home, `h`, the first step is to sample the first trip. From the updated posterior probabilities there is a near-equal probability to not travel at all (36.3%) or to use public transport (42.3%). There are less probable modes like walking (11.4%) or some other mode (8.8%). If the weighted random draw sees Lilly not travelling, so `trip 1` is `n`, the posterior probabilities in Figure 5b is updated to reflect the new evidence. If Lilly is not travelling, there is a 100% that there will not be an activity 2, and also no trip 2. The activity chain therefore is completed and is only made up of the home activity, `h`. The trip chain is, similarly, denoted only by `n`.

Conversely, if Lilly's first trip is sampled to be by public transport, denoted by `p` as shown in Figure 5c, then the subsequent activities and trips look quite a bit different in terms of the updated posterior probabilities. With the new evidence of travelling by public transport, there is 85.2% chance that the activity to which Lilly travels is work, `w`, and a 10.4% chance that it is some other activity, `o`. Trip 1 also influences trip 2 and so we see there is 93.8% chance that the second trip will also be by public transport. For the remainder of the illustration, assume that Lilly's first trip was by public transport, `p`.

Having evidence that Lilly is employed *and* that the first trip is by public transport, one can sample both activity 2 and trip 2, independently from one another as they do not influence one another. That is, there is no causal relationship between them in the BN structure. Say, given its high priority, that the second activity is work, `w`, and the second trip is by public transport, `p` (Figure 5d), one can now sample the third activity. From Figure 5d we note that there is now a 90.7% chance that Lilly will return home, `h`, and a 17.9% chance that se will attend to a second work activity, `w`. There are also unlikely shopping (0.1%), leisure (0.3%) or other (1.0%) activities. Say, at this point, that one samples the third activity to be home, `h` (Figure 5e), then the subsequent trips are most likely to be to *not travel* (96.1%), `n`, and therefore it suggests the end of Lilly's day. However, there are small probabilities suggesting she may continue with other activities and travel to them via car (0.2%), public transport (0.5%), walking (0.9%), or some other mode (0.3%). If one samples the no-travel option, then the sampled activity chain for Lilly would end, and we denote the activity chain portion as `h-w-h`, and the trip chain as `p-p`.

For each individual one starts off by providing only the demographic evidence. As one progressively sample the next trip or activity, and provide that, in turn, as evidence, the posterior probabilities are updated and influences the sampling of subsequent trips or activities. That is, the evidence propagates through the BN. This allows one the opportunity to, given different pseudo random seeds, generate multiple synthetic activity chains for each individual for which you have demographic evidence. In this paper we synthesised 100 activity chains for each person in the test set.

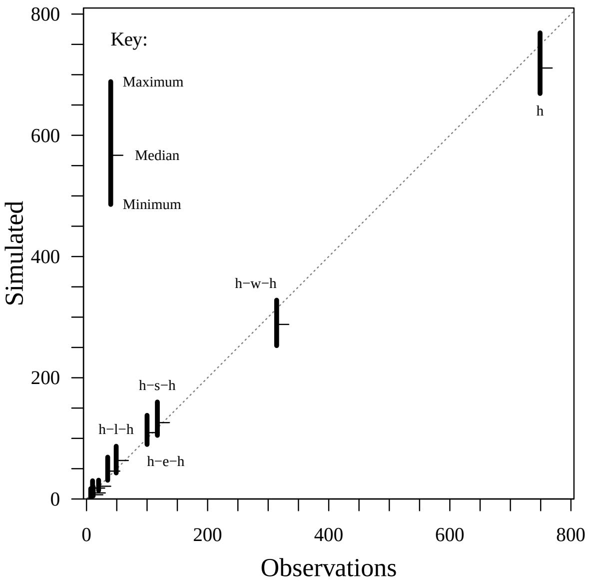
## 5. Results and discussion

In the literature the results are often aggregated. For example, Guzman et al. (2017) only reports the overall number of activity chains of a specific type, and not whether the specific *person*'s activity chain was correctly estimated.

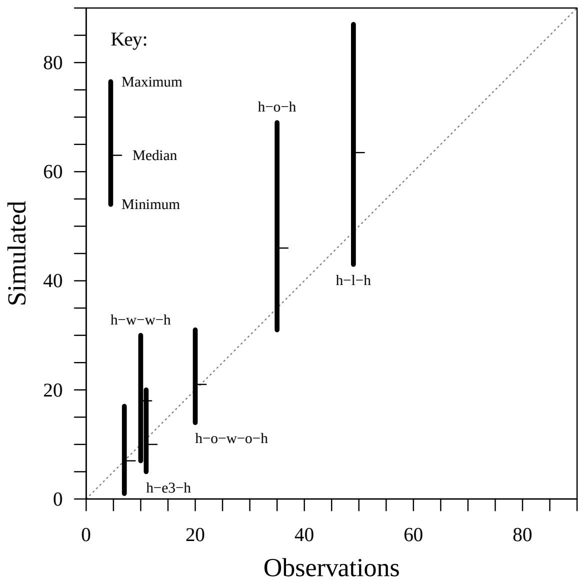
The profile approach of Liu et al. (2015) can be applied to validate activity-based transportation models. But here, too, the focus is on getting it right *overall*. How well does this person's activity chain compare to *known clusters*, i.e. typical and frequently observed clusters. In line with this practice, we also report aggregated results first and then move on to more detailed methods of evaluating the proposed BN methodology.

### *5.1. Better than random?*

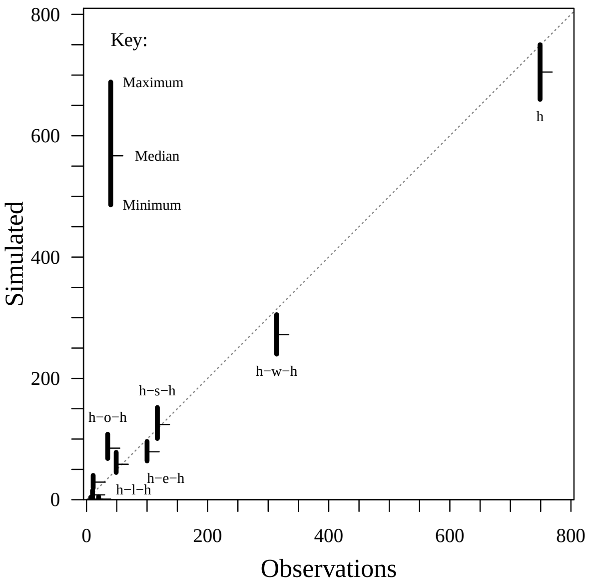
To test our approach, we split the observed activity chains into a *training set*, which represented 80% of the observations, a total of 7 507 individuals, and a *test set* representing the balance of 1 501 individuals. As a first result we also consider an aggregate comparison by taking 100 bootstrap samples, each sample having 1 501 chains, from the training set. For each sample we observe the frequency that specific activity and trip chain configurations occur and compare it against the test set. The results for the bootstrapped samples of the activity chains are shown in Figures 6a and 6b. As an example, consider the `h-w-h` activity chain. In the test set there were 314 occurrences while there were between 253 and 328 occurrences across the 100 samples, with the median being 288 occurrences. The dotted diagonal line represents ideal behaviour, so the closer the median value is to the diagonal, the better. Less frequently observed activity chains are visualised in more detail in Figure 6b.



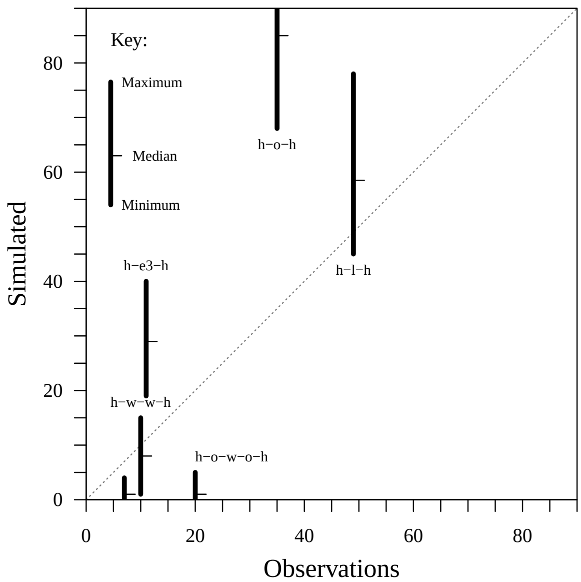

(a) Bootstrapped activity chains.




(b) Less frequent observed bootstrapped activity chains.




(c) Bayesian network activity chains.




(d) Less frequent observed Bayesian network activity chains.

Figure 6: Comparison of the frequency of occurrences for specific activity chain configurations. The observations ($x$-axis) refer to the test set while the simulated ($y$-axis) values refer to the range over the 100 samples.

Similarly, we generated 100 samples from the BN and again observed the number of occurrences for each chain configuration. The results for the activity chains are visualised in Figure 6c with the less frequently observed activity chains shown in more detail in Figure 6d.

Results for the bootstrap and BN *trip* chains are given in Figure 7. As expected, the bootstrap samples generally agree with the observed activity chains. The BN samples follow a very similar pattern. The number of people who stay home, denoted by `h` in Figure 6a and `n` in Figure 7a, is underestimated. So too are the cohort of people who only travel to work, denoted by `h-w-h`, and those who use public transport (`p`) for both the outbound and returning trip. This concurs with Figures 6d and 7d, which illustrates that the majority of respondents in the dataset has no trip, or a predictable home-work-home travel diary. These aggregated results is aligned with results reported in literature. Consequently, our results support the choice by, for example Guzman et al. (2017) and Hafezi et al. (2019), that non-travellers and those with predictable, short activity chains make up the bulk of the population. But what if we want to move beyond the *typical person* and their travel patterns on a *typical day*?

The end use of the sampling data urges us to look into more detail than the overall performance: If we want to perform agent-based modelling, we need to have confidence in modelling *all* agents, with their inherent variability, and not only the majority.

### *5.2. Detail performance analysis*

When considering detail performance, we investigate each chain independently in terms of accuracy and then the performance analysis become analogous to classification performance analysis. This allows one to turn to standard machine learning metrics to evaluate the models' output.

#### *5.2.1. Performance Metrics*

To explain the detailed metrics, some machine learning terminology is introduced for more clarity. Consider the Sankey diagram in Figure 8 that is a simplification of the complete test set and sample 1 for the activity chains. We denote with $\boldsymbol{S} = \{1, \ldots, 4\}$ the joint set of activity chain configurations observed in both the test set and the first sample. Here we simplified it to only $|\boldsymbol{S}| = 4$ configurations where $s = 1, s \in \boldsymbol{S}$ represents the stay-at-home activity chain, `h`, and $s = 2$ the home-work-home activity chain, `h-w-h`. For the first sample the aggregate results look fairly good since 47.0% of all the activity chains in the sample were `h`, compared to the 49.9% observed in the test set.

And this is where one needs to scratch deeper, below the surface. We let $x_{i,j}$ denote the number of individuals in the test set who had an activity chain $i \in \boldsymbol{S}$ while the predicted

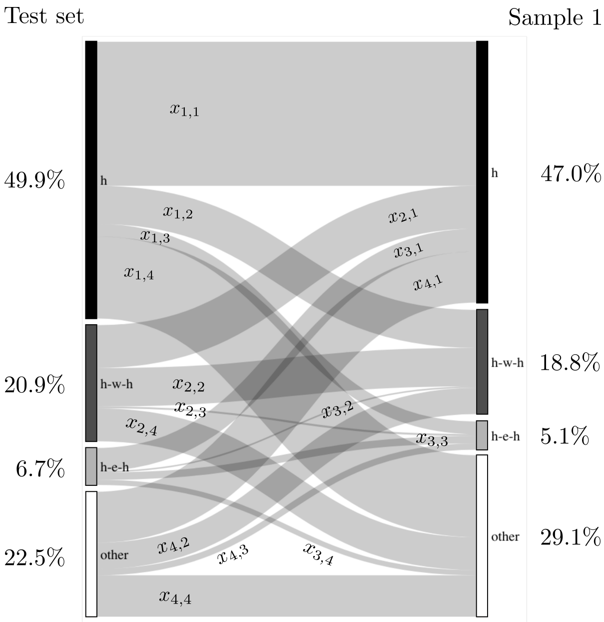
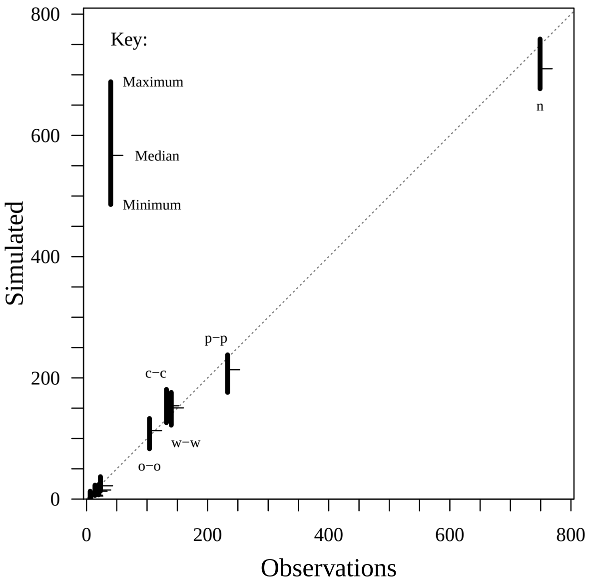
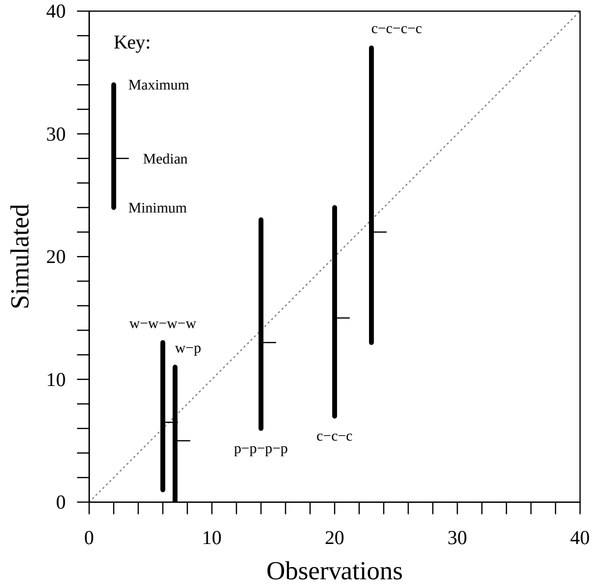


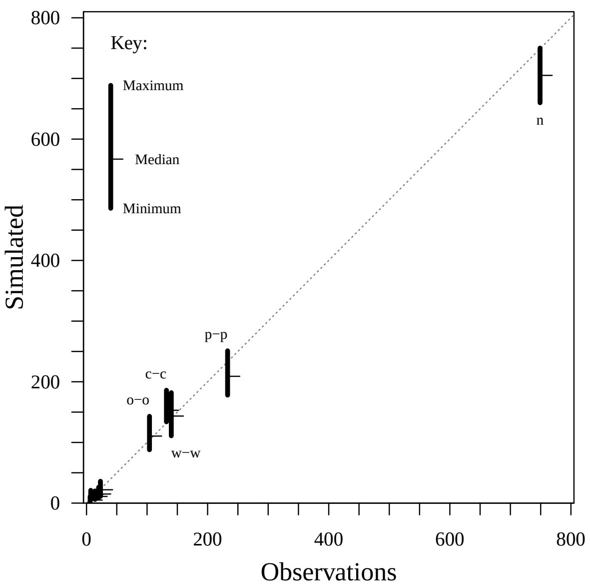

Figure 8: Illustrative Sankey diagram to explain more detailed machine learning metrics.



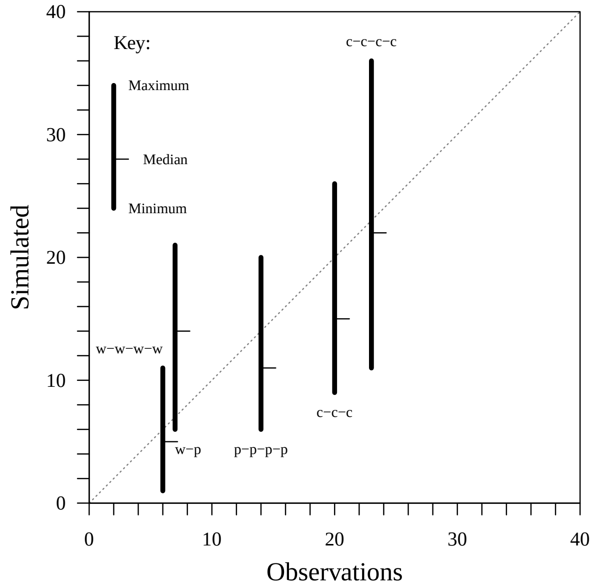

(a) Bootstrapped trip chains.




(b) Less frequently observed bootstrapped trip chains.




(c) Bayesian network trip chains.




(d) Less frequently observed Bayesian network trip chains.

Figure 7: Comparison of the frequency of occurrences for specific trip chain configurations. The observations ($x$-axis) refer to the test set while the simulated ($y$-axis) values refer to the range over the 100 samples.

activity chain for those same individuals was $j \in \boldsymbol{S}$. As an example, $x_{1,2}$ would represent the number of individuals who had an activity chain of `h` in the test set, yet we predicted that they would have a `h-w-h` activity chain in the sample.

To calculate the detailed metrics, we firstly define *true positive* (TP) to be the number of individuals that had and activity chain $i \in \boldsymbol{S}$ in both the test set and the sample, $x_{i,i}, i \in \boldsymbol{S}$. In the given example, the true positives for `h` would be $x_{1,1} = 390$, and for `h-w-h` it would be $x_{2,2} = 104$. Next we define *false positives* (FP) as the number of individuals for which we predicted an activity chain of, for example, `h`, yet we observed them to have some other activity chain that was *not* `h`. Using `h` again, as an example from Figure 8, the false positives would then be $x_{2,1}+x_{3,1}+x_{4,1} = 116+62+138 = 316$. More generally, one can express false positives for activity chain $j \in \boldsymbol{S}$ using (2).

$$FP_j = \sum_{i \in \boldsymbol{I}, i \neq j} x_{ij} \qquad \forall j \in \boldsymbol{S} \tag{2}$$

Thirdly, we define *true negatives* (TN) for an activity chain $i \in \boldsymbol{S}$ in which we are interested as the number of individuals for which we observed an activity chain *other than* $i$ in the test set, and we also predicted those individuals to have some other activity chain *other than* $i$. In the case of `h` in Figure 8 the true negatives would be $x_{2,2}+x_{2,3}+x_{2,4}+x_{3,2}+x_{3,3}+x_{3,4}+x_{4,2}+x_{4,3}+x_{4,4} = 104+5+89+4+20+14+71+17+112 = 436$. More generally one can express true negatives for activity chain $i \in \boldsymbol{S}$ using (3).

$$TN_i = \sum_{k \in \boldsymbol{S}, k \neq i} \sum_{j \in \boldsymbol{S}, j \neq i} x_{kj} \qquad \forall i \in \boldsymbol{S} \tag{3}$$

Finally, we define *false negatives* (FN) for an activity chain $i \in \boldsymbol{S}$ in which we are interested as the number of individuals for which we actually observed them to have activity chain $i$ in the test set, yet we predicted them to have some *other* activity chain in the sample. From the simplified example in Figure 8 the false negatives would then be $x_{1,2} + x_{1,3} + x_{1,4} = 103 + 34 + 222 = 359$. More generally one can express false negatives for activity chain $i \in \boldsymbol{S}$ using (4).

$$FN_i = \sum_{j \in \boldsymbol{S}, j \neq i} x_{ij} \qquad \forall i \in \boldsymbol{S} \tag{4}$$

And with these four definitions one can now introduce the metrics. Accuracy sums true positive (TP) and true negative (TN) predictions and normalise it with the total number of individuals in the test set. Reporting on accuracy alone might provide a false sense of good performance, especially when the training set is imbalanced. To report more critically on prediction performance, three questions of importance require attention:

- What percentage of your predictions were correct?

$$\text{accuracy } (A) = \frac{TP + TN}{\sum \text{test cases}} \tag{5}$$

- What percent of the positive cases did you catch?

$$\text{recall } (R) = \frac{TP}{\text{TP + TN}} \tag{6}$$

- What percent of positive predictions were correct?

$$\text{precision } (P) \;= \frac{TP}{\text{TP} + \text{FP}} \tag{7}$$

One method to combine recall and precision is the F-score, which is defined as the weighted harmonic mean of recall and precision, and calculated using (8).

$$\text{F-score } (F) = \frac{2 \times (\text{recall} \times \text{precision})}{\text{recall} + \text{precision}} \tag{8}$$

*5.2.2. Levenshtein distance*

In information theory, Hamming distance is the number of corresponding positions in two equal length strings that are different (MacKay, 2003). We need to relax the equal length assumption as chain lengths differ. The Levenshtein distance (Sarkar et al., 2016) does not require the equal length assumption and provides a metric to compare observed and sampled chains in terms of values and length. In short, the Levenshtein distance returns the number of edits required to change one string into the other string. The edits can be substitutions, deletions or insertions. Consider the following two trip chains: $\text{trip}_1$ = `c-w-w`, $\text{trip}_2$ = `w-w-p`. The Levenshtein distance between these two trips is 4 as 4 edits are required to change $\text{trip}_1$ to match $\text{trip}_2$. The edits are indicated below:

$$\begin{aligned}\text{trip}_1 &= \boxed{\texttt{c-}}\,\texttt{w-w} \\ \text{trip}_2 &= \texttt{w-w}\,\boxed{\texttt{-p}}\end{aligned}$$

The Levenshtein distance between two strings $a, b$ (of length $|a|$ and $|b|$ respectively) is given by $\text{lev}_{a,b}(|a|, |b|)$ where

$$\text{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \text{lev}_{a,b}(i-1,j) + 1 \\ \text{lev}_{a,b}(i,j-1) + 1 \\ \text{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases} \tag{9}$$

The Levenshtein ratio can be calculated by normalising the raw distance (Sarkar et al., 2016). Consider two strings $a$ and $b$. The ratio is:

$$\text{levRatio}_{\text{a,b}} = \frac{lev_{a,b}(i,j)}{|a| + |b|} \tag{10}$$

A Levenshtein ratio of 1 indicates a perfect match between the two sequences whereas 0 indicates zero overlap between the two sequences. We use the Levenshtein ratio as another proxy for accuracy. A ratio of one represents true positives (2) and all ratios smaller than one represent true negatives (3).

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).

| | | | Bootstrapping | | | | Bayesian network | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Chain** | **Occurences in test set** | **Cumulative occurence** | ***A*** (5) | ***P*** (7) | ***F*** (8) | ***LS*** | ***A*** | ***P*** | ***F*** | ***LS*** |
| h | 749 | 49.9% | 50.1% | 50.0% | 48.8% | 0.64 | 54.0% | 54.2% | 52.5% | 0.66 |
| h-w-h | 314 | 70.8% | 68.0% | 21.0% | 20.0% | 0.61 | 75.3% | 39.5% | 36.6% | 0.69 |
| h-s-h | 117 | 78.6% | 85.0% | 7.6% | 7.9% | 0.58 | 85.4% | 8.9% | 9.1% | 0.58 |
| h-e-h | 100 | 85.3% | 87.0% | 6.7% | 7.0% | 0.58 | 90.9% | 26.8% | 23.7% | 0.58 |
| h-l-h | 49 | 88.5% | 92.8% | 3.2% | 4.1% | 0.57 | 93.1% | 3.6% | 4.7% | 0.54 |
| h-o-h | 35 | 90.9% | 94.6% | 1.9% | 3.6% | 0.57 | 92.2% | 2.6% | 4.0% | 0.56 |
| h-o-w-o-h | 20 | 92.2% | 97.3% | 1.2% | 5.3% | 0.45 | 98.6% | 3.9% | 9.4% | 0.53 |
| h-e3-h | 11 | 92.9% | 98.6% | 0.9% | 9.9% | 0.52 | 97.4% | 1.2% | 6.2% | 0.48 |
| h-w-w-h | 10 | 93.6% | 98.2% | 0.7% | 7.0% | 0.51 | 98.8% | 2.1% | 12.2% | 0.59 |
| h-o-w-h | 7 | 94.1% | 99.1% | 0.6% | 12.3% | 0.52 | 99.5% | 0.0% | – | 0.59 |
| h-w-h-l-h | 5 | 94.4% | 99.6% | 0.0% | – | 0.44 | 99.4% | 0.1% | 14.3% | 0.52 |
| h-w-w-w-h | 5 | 94.7% | 99.4% | 0.9% | 23.0% | 0.43 | 99.5% | 0.2% | 20.0% | 0.51 |
| h-w-o-h | 1 | 95.0% | 99.7% | 0.0% | – | 0.43 | 99.9% | 0.0% | – | 0.53 |

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

| | | | Bootstrapping | | | | Bayesian network | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Chain** | **Occurences in test set** | **Cumulative occurence** | ***A*** (5) | ***P*** (7) | ***F*** (8) | ***LS*** | ***A*** | ***P*** | ***F*** | ***LS*** |
| n | 749 | 49.9% | 50.0% | 49.9% | 48.6% | 0.47 | 54.0% | 54.2% | 52.5% | 0.51 |
| p-p | 233 | 65.4% | 74.7% | 15.5% | 14.8% | 0.27 | 77.4% | 24.5% | 23.1% | 0.36 |
| w-w | 140 | 74.8% | 82.5% | 8.9% | 9.2% | 0.24 | 83.6% | 13.2% | 13.4% | 0.26 |
| c-c | 132 | 83.5% | 82.7% | 8.8% | 9.4% | 0.25 | 86.1% | 25.1% | 26.9% | 0.43 |
| o-o | 104 | 90.4% | 86.6% | 7.0% | 7.3% | 0.23 | 87.1% | 9.4% | 9.6% | 0.26 |
| c-c-c-c | 23 | 93.0% | 97.0% | 1.4% | 5.2% | 0.17 | 97.2% | 5.1% | 6.8% | 0.35 |
| c-c-c | 20 | 93.3% | 97.7% | 0.9% | 5.5% | 0.20 | 97.7% | 3.4% | 7.2% | 0.32 |
| p-p-p-p | 14 | 94.2% | 98.2% | 0.8% | 7.5% | 0.19 | 98.3% | 2.2% | 8.9% | 0.30 |
| w-p | 7 | 94.7% | 99.2% | 0.3% | 15.0% | 0.25 | 98.6% | 0.6% | 9.5% | 0.29 |
| w-w-w-w | 6 | 95.1% | 99.2% | 0.3% | 16.7% | 0.17 | 99.2% | 0.5% | 15.5% | 0.18 |

#### *5.2.3. Results*

The dataset contains 139 unique activity chains in the training set and 157 unique activity chains in the test set. Similarly, it contains 103 unique trip chains in the training set and 53 unique trip chains in the test set. We only consider the chains which cover 95% of occurrences in the test set and report on them in Tables 4 and 5. The tables show the accuracy ($A$), precision ($P$), F-score ($F$) and average Levenshtein ratio ($LS$) metrics for each chain and for both the bootstrapping and BN models. The reported value in each case is the mean calculated over the 100 samples. For some low occurrence chains, the F-score could not be calculated because no true positive cases exist for these instances. The BN model consistently shows better performance than bootstrapping. For the activity chain `h-e-h` which represents individuals travelling to education and back home, the BN precision score is 20% better than bootstrapping. We observed low precision and F-score metrics which might cast doubt on the model's performance. For example, for the activity `h-w-h`, the BN model precision is 39.5%. It should be taken into account that accuracy is defined as those chains which were sampled *perfectly*. An observed chain `h-w-h` which was predicted as `h-w-w` will be regarded as completely inaccurate, when in reality, the first part of the chain was indeed correct. This is where the Levenshtein ratio provides insight into the degree of overlap between the observed and sampled chains. We therefor report on the average Levenshtein similarity ratio as well.

### *5.3. Discussion*

One problem with a frequentist approach is that you can only synthesise chains that were observed in the original data. This holds for many rule-based (tree structure) and bootstrapping approaches. The BN, on the other hand, sample from posterior probabilities that allows one to synthesise activity chains that are likely or plausible, yet may not have been explicitly observed in the data. This is especially valuable, and important, when one has to understand behaviour from rich but very small data sets like travel diaries.

## 6. Introducing temporal variables

Up to this point, the paper focused on the generative ability of BNs to produce structurally sound activity and trip chains but neglected the temporal dimension. It is entirely plausible to argue that a person with an extended, 10-hour work activity in her/his activity chain will have a shorter activity chain with fewer activities. The quicker the activities, the more a person can fit into their time budget and the longer the length of the trip/activity. But is this supported by the data? Or, a more important question, can the Bayesian network approach accommodate this?

This section has two objectives. First, to introduce activity durations as temporal variables into the Bayesian network and explore if and how they affect the network structure. Secondly, we want to demonstrate that complexity increases only marginally when adding more variables.

Adding temporal variables allows one to calculate the total diary duration for utility-earning activities by merely summing over all of them. While we will only demonstrate activity duration in this section, one can in future extend this to include trip durations as well. We argue that having the total chain duration adds valuable inference capabilities and explainability to the model. For example, it allows one to estimate the overall time our illustrative Lilly spends outside her home. As a young female dependent on public transport, the out-of-home time may be a proxy for her risk to become a victim of gender-based violence and can contribute to policy decision-making.

In the travel survey, people reported the start and end-times of their activities. For the first (home) activity, they only state the time they departed and, similarly, for the last activity of their day, they only give the arrival time. Consequently, for all other activities that make up the chain, we calculated and captured the activity duration as a continuous variable. When introducing continuous variables into a BN, one must first discretise the values into bins. The software we used, BayesiaLab, has the functionality to optimise the bin intervals through several activities, one being a Genetic Algorithm. For the City of Cape Town travel survey, the optimised bins were $[0, 0.5]$, $(0.5, 2]$, $(2, 5]$ , $(5, 8]$, $(8, 10]$, $> 10$ hours. Although the discretisation was machine-learned, it surprisingly corresponded with typical activity durations.

*6.1. Structural learning*

The next step is to learn a structure from the data. We impose the same temporal restrictions introduced in Section 4. We denoted the duration of the second activity by `dur 2`, the third activity by `dur 3`, etc. Figure 9 illustrates the resulting machine-learned structure. It is noteworthy that

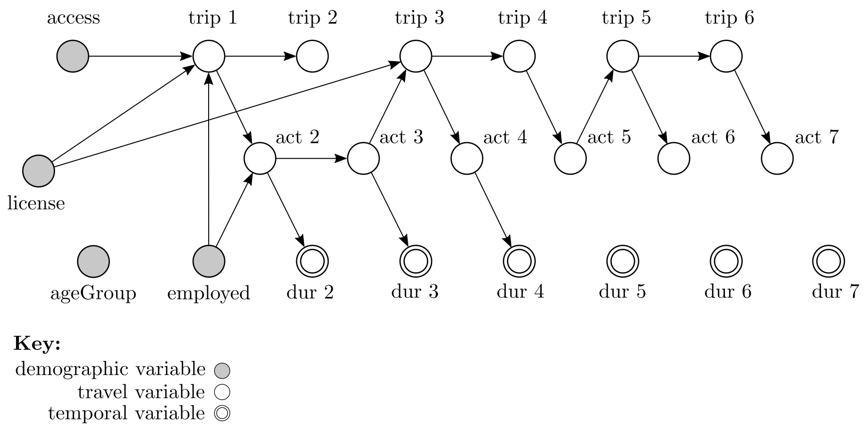



Figure 9: Learned Bayesian network structure with temporal variables included.

not all the temporal links (for example, from `act 5` to `dur 5`) manifested from the learning process. Congruent with BN theory, one can combine machine-learned and expert inputs to establish the structure by placing localised prior links between nodes. Chickering et al. (2013) refer to this method as *structural prior learning*. The temporal patterns one sees in the early stages of the chain (`act 2` and `act 3`) influenced our choice for priors, resulting in the uniform BN structure shown in Figure 10. What is insightful in the network is that the type of activity affects its duration, as one would expect. But the duration of an activity does not influence later activities. Instead, only activities influence subsequent activities in the chain, which is a change from the earlier machine-learned network in Figure 3 without the temporal variables.

Still, the posterior probabilities estimated from the data make intuitive sense. Consider Lilly again, the early career female without a license and no access to a private car. When synthesising her partial activity chain, one follows the same procedure as discussed in Section 4.3 and illustrated in Figure 5. Now, with the temporal variables included in the Figure 10 network, one might sample `trip 1` to be public transport (`p`); and then `act 2` as work (`w`) and `trip 2` to as public transport (`p`) again. Knowing both `act 2` and `trip 2` allows one to sample `act 3` to be home (`h`) again in this illustration. Figure 11a shows the evidenced variables. With `act 2` evidenced as work (`w`), we see an activity duration (`dur 2`) distribution with a median of 8–10 hours. The duration distribution for `act 3` is most likely not to exist, implying that home (`h`) will be the last activity in Lilly's chain.

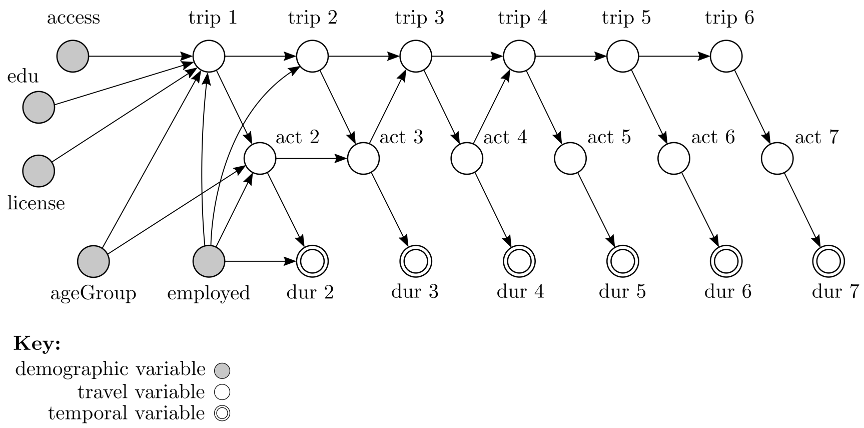
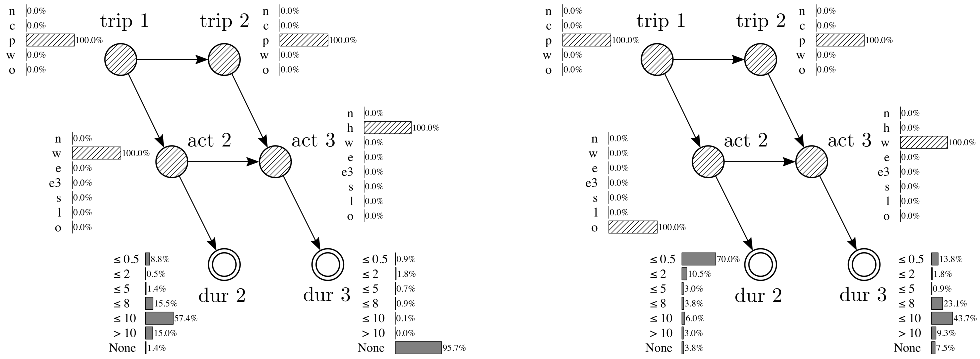



Figure 10: Bayesian network resulting from structural prior learning.




(a) Work as the first non-home activity.

(b) Other first non-home activity.

Figure 11: Illustrating how evidenced activity variables affect the posterior probability distribution of those activities' duration.

If, however, Lilly's second activity was sampled to be some other (`o`) activity, like buying her bus ticket, the posterior probability distribution of the second activity (`dur 2`) is heavily skewed towards a much shorter event (Figure 11b). Also, if her `act 3` is work (`w`), its duration distribution (`dur 3`) is most likely to be around 8–10 hours, which is plausible for a full day's work.

The more variables one add, like the temporal activity durations here or trip durations that we leave for future work, the more degrees of freedom one adds that has to be accounted for when measuring the accuracy of the synthesised chains. The purpose here was primarily to demonstrate that one indeed *can* address the important temporal aspect of activity-based travel demand, only adding marginal complexity, and without losing the intuitive explanatory power of the network. A more exhaustive network will require more data to maintain the accuracy we reported in this paper. If, like Yin et al. (2018), one has access to large call data records or geospatial positioning system (GPS) data from telemetry devices, additional temporal variables can be added with confidence.

## 7. Conclusion

This paper responded to the call of Rasouli and Timmermans (2014) that travel demand models need to become more behaviourally rich. More specifically, this paper introduced BNs, which allow for causal interpretation, to synthesise daily travel demand patterns in the form of activity and trip chains. We look at both activity sequences and the modes connecting them, i.e. trip chains.

The approach allows one to not merely follow a frequentist approach in which one only looks at the *typical* and most common activity chains. We did not do any *a priori* clustering of records based on, for example, demographics as was done in Liu et al. (2015). Instead, we included both frequent and low-frequency observations to influence the synthesis.

To evaluate the approach, this paper also moved beyond aggregate analysis. We introduced a number of machine learning metrics that have not achieved much traction in the transport domain as in more progressive fields.

One can further evaluate the BN by comparing synthesised and observed travel sequences from an unseen population, such as a survey from a different city.

Two specific areas of future research are now open. Firstly, BNs lend itself to use combinations of data-driven and expert inputs for both the structure and conditional probabilities. For example, looking at the demographics representation in Table 3, one can choose to keep the machine learned conditional probabilities for highly represented demographic profiles, and refine the lower represented profiles with expert knowledge. The opportunity here is to focus specifically on the low-frequency observations and incorporate expert input that are not easy to capture in small-scale surveys. The value for both academia and practitioners is that when equity becomes part of the overall metrics, understanding the behaviour behind the less frequently observed travel patterns may provide valuable insight to decision-makers. Specifically developing countries challenged with high levels of economic and mobility inequality can find value in such flexible approaches.

Secondly, we did not look at joint activity scheduling within households. This is in line with the concern raised by Rasouli and Timmermans (2014). But Sun and Erath (2015) and Sun et al. (2018) have shown that BNs can easily be constructed in hierarchies so that within-household decision-making can be accommodated.

### Acknowledgements

We would like to thank Lionel Jouffe, CEO of Bayesia for valuable discussions and technical support.

### References

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

Arentze, T. and Timmermans, H. (2007). Parametric action decision trees: Incorporating continuous attribute variables into rule-based models of discrete choice. *Transportation Research Part B: Methodological*, 41(7):772–783.

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.

Borysov, S., Rich, J., and Pereira, F. (2019). How to generate micro-agents? A deep generative modeling approach to population synthesis. *Transportation Research Part C*, 106:73–97.

Bromley, J., Jackson, N. A., Clymer, O., Giacomello, A. M., and Jensen, F. V. (2005). The use of Hugin® to develop Bayesian networks as an aid to integrated water resource planning. *Environmental Modelling & Software*, 20(2):231–242.

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

Chickering, D. M., Heckerman, D., and Meek, C. (2013). A bayesian approach to learning bayesian networks with local structure. *arXiv preprint arXiv:1302.1528*.

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.

Feygin, S. and Pozdnoukhov, A. (2018). Peer pressure enables actuation of mobility lifestyles. *Transportation Research Part C*, 87:26–45.

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.

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

Guzman, L. A., Gomez, A. M., and Rivera, C. (2017). A strategic tour generation modeling within a dynamic land-use and transport framework: A case study of Bogota, Colombia. *Transportation Research Procedia*, 25:2536–2551.

Hafezi, M. H., Liu, L., and Millward, H. (2019). A time-use activity-pattern recognition model for activity-based travel demand modeling. *Transportation*, 46(4):1369–1394.

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.

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

Liu, F., Janssens, D., Cui, J., Wets, G., and Cools, M. (2015). Characterizing activity sequences using profile Hidden Markov Models. *Expert Systems with Applications*, 42(13):7505–5722.

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.

Ma, T.-Y., Chow, J. Y. J., and Xu, J. (2017). Causal structure learning for travel mode choice using structural restrictions and model averaging algorithm. *Transportmetrica A: Transport Science*, 13(4):299–325.

Ma, T.-Y. and Klein, S. (2018). Bayesian networks for constrained location choice modeling using structural restrictions and model averaging. *European Journal of Transport and Infrastructure Research*, 18(1):91–111.

MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.

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.

Roorda, M. J., Miller, E. J., and Habib, K. M. N. (2008). Validation of TASHA: A 24-h activity scheduling microsimulation model. *Transportation Research Part A*, 42(2):360–375.

Saadi, I., Mustafa, A., Teller, J., and Cools, M. (2016). Forecasting travel behavior using Markov chains-based approaches. *Transportation Research Part C*, 69:402–417.

Sarkar, S., Das, D., Pakray, P., and Gelbukh, A. (2016). Junitmz at semeval-2016 task 1: Identifying semantic similarity using levenshtein ratio. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 702–705.

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

Sun, L., Erath, A., and Cai, M. (2018). A hierarchical mixture modeling framework for population synthesis. *Transportation Research Part B: Methodological*, 114:199–212.

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

Transport for Cape Town (2013). 2013–2018 comprehensive integrated transport plan. Technical report, City of Cape Town.

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

Yin, M., Sheehan, M., Feygin, S., Paiement, J.-F., and Pozdnoukhov, A. (2018). A generative model of urban activities from cellular data. *IEEE Transactions on Intelligent Transportation Systems*, 19(6):1682–1696.

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.