A DSS for Foreign Goldsmiths Students

Task

The main goal for my Decision Support System (DSS) is to create a transport recommendation system for foreign students who come to study at Goldsmiths University, in London.

The purpose of this project is to assist foreign students how to navigate their way from their local residence into Goldsmiths' campus, situated in New Cross Gate, in the borough of Lewisham.

Background Knowledge as a Bayesian Network

My personal motivation is that 12 years ago, when I first came to London, I was a Goldsmiths foreign student myself. At the time I didn't know the city at all and also wanted an efficient and cheap way to get myself to class everyday. As a foreign student in an expensive city such as London, I was living on a budget, so I needed to get around in a cheap and efficient way. I lived about 3 miles away from campus, so walking was not an option.

I had 3 possible ways to get to campus:

a) **Train**, the quickest way, since it's a short walk from New Cross Gate station to Goldsmiths building complex. But riding the train everyday was expensive for a foreign student like myself and I certainly couldn't afford it every day. Also at the time the train service would suffer several cancellations throughout the day, so it wasn't always reliable

b) **Bus**, was a good option since the bus stop was closer to the house where I lived compared to the train station, so on rainy days I wouldn't get wet as if I'd used the train. It was also 50% cheaper than the train, so it was more attractive to my student's budget. But the bus route was prone to traffic jams, particularly early in the morning when the first lecture would start at 9AM.

c) **Bicycle**, was another available choice. It was inexpensive (apart from the initial first investment in purchasing a second hand bicycle), and a really pleasant ride in the nice spring weather, but I dreaded it in the cold winter months, since I am originally from a South American country and not used to harsh winters. Also there was a huge hill in my route, so it would take about the same time to go by bicycle than by the other two previous options. Lastly, as a female student, I was also weary of using the bicycle on frequent basis because of female cyclists are more likely to have a road accident than male ones as Greater London according to Transport For London statistics.

What I attempt to do in this brief description of my commute as a Goldsmiths student, is to quantify uncertainty for different concurring and competing variables such as: weather conditions, time of day for travel, monthly travel budget, traffic conditions, train services frequencies, geographic characteristics of the commute route, gender and age.

This way of quantifying uncertainty is the basis of my Bayesian Network. Out of all the competing variables, I seek to identify which ones should be the main nodes for my recommendation system.

Potential Users

Foreign students at Goldsmiths University who live in the Greater London area between 1 and 10 miles from the New Cross Gate campus. I am choosing foreign students, who live within a 10 miles radius so as to have 3 possible means of transport: train, bus and bicycle. In order to identify these nodes, it is imperative to think about who will be the main users for my BN, in other words how would I characterise the foreign students at Goldsmiths and their needs in the next section.

Conceptual Design:

The questions about foreign students that will inform my development process are:

- How close do they live from the university?
- How frequently do they need to be on campus on a weekly basis?
- How much does it cost per trip to get to campus?
- What percentage of their monthly budget can be dedicated to transport?
- Are they full-time or part-time students?
- What are the available means of transport to get to campus?

Network Components

From these questions, I will identify the main variables in Table 1. that can be used as a components for informing decisions about the best way to travel to Goldsmiths campus.

Table 1. Network Nodes

Node	Expected Values	Rationale
Age	0 = Above 30 years old 1= Below 30 years old	I choose this threshold to represent younger students who are more likely to want to ride a bicycle. Older students might be able to afford accommodation that is closer to campus. Younger students are more likely to be full-time students.
Gender	0 = Female 1 = Male	The fact that a student is male or female might condition their choice of travel option. Gender might also inform how close they want to live from campus, so their residence
Residence	0 = Within 3 miles 1 = Greater than 3 and smaller than 10 miles	Location of their home will obviously affect which type of means of transport is available and also the closer to campus the more likely to ride a bicycle

Weather	0 = above or equal to 15C 1 = below 15C	Weather, as guidance I am considering that any weather above 15 C is considered to be mild, and below that is likely to put off people from riding a bicycle, particularly those who live further away.
Type of Student	0 = Full-time 1 = Part-time	I consider that part-time students are more likely to have a part-time job and therefore will have less time for commuting and be more likely to rely on the transport system.
Means of transport	0 = Train 1 = Bus 2 = Bicycle	The options are conditioned by the type of transport available in the Goldsmiths campus area: trains servicing New Cross and New Cross Gate stations, bus stops within 0.1 mile of campus and bicycles.

For programming simplicity in this network, I am only including 2 possible gender options, although I am aware that gender is not a binary answer.

Fig.1 - DAG

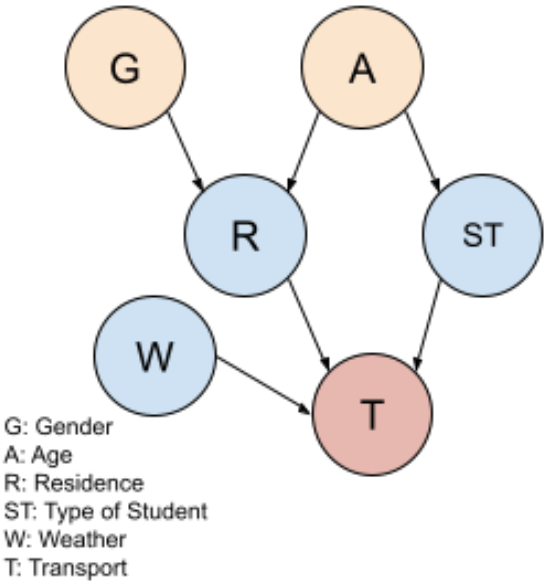


Figure 1 illustrates the Direct Acyclic graph (DAG) with the nodes in the Bayesian Network and the direction of relationships.

Input collection & Output

The input data can be simply collected via a user interface, hosted perhaps in the university's portal. Users can answer configuration questions such as:

1) Select if your gender: Male, Female
2) Enter DOB: Above or 30, Below 30.
3) Enter your residence's postcode: [SW2 2QT] - background calculations will input the distance to campus and bin the user in either of these buckets: above 3 miles, within 3 miles. This can be easily done with Python libraries such as Geopandas, Haversine and Geopy
4) Are you a full-time or part-time student?

Weather conditions will query against the Met Office's weather forecast API, each time a question is asked to the network, so no need to ask that question.

The results will read:

1) Your recommended means of transport is: [Train] and a link to Transport for London's app so they can identify the best route.
2) Your recommended means of transport is: [Bus] and a link to Transport for London's app so they can identify the best route.
3) Your recommended means of transport is: [Bicycle] and a link to cyclists' app so they can identify the best route.

Existing solutions & potential integrations

Currently there are many transport recommendations applications such as Google Maps, Waze, Cyclers, TFL, CityMappers, etc. However none of these systems are customisable enough to meet the needs and profile of foreign students who face particular challenges at time of travel.

I expect this recommendation system can be best integrated with Google Maps as this product has an open source API that the BN can query against. Google Maps can also provide live status of train services and bus routes being diverted for roadworks. However this DSS can also be used as a stand-alone that would just return a recommendation for a transport system and then the user would need to go to another application to check for more details, i.e. the actual bicycle route on Cyclers.

It also can be customised by inputting the student's weekly class schedule and/ or the allocated monthly budget, affecting expensive transport options like train towards the end of the month.

Constraints

As per recommended guidance, I am limiting this BN to 6 nodes, but it is possible to expand this network as per above reference to include other nodes such as: monthly transport budget, day of the month and week, student schedule, availability of bus and train services depending on residence, existence of cycle routes from residence to campus, etc.

Stakeholders

The main stakeholders will be Goldsmiths University and their webmaster, who would be responsible for hosting this solution in the student portal and evaluating if it's suitable or not.

Risks

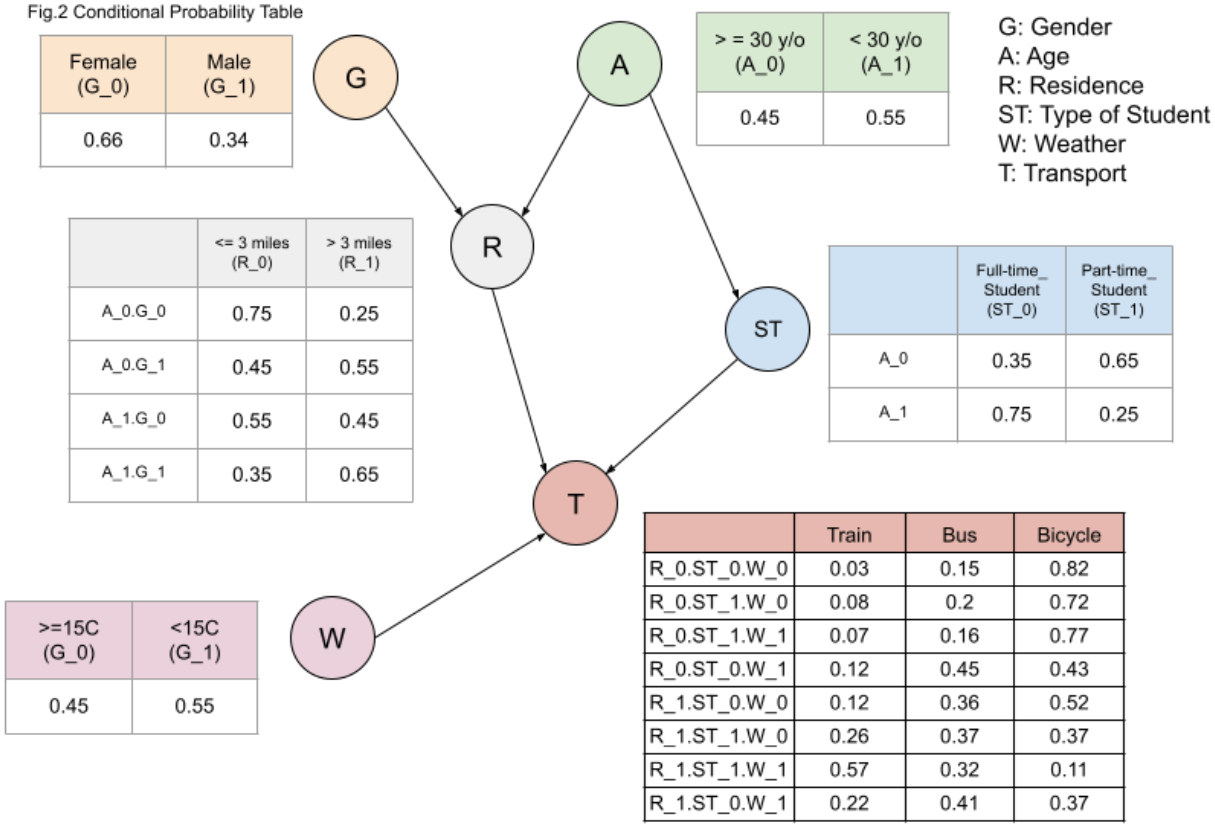
Potential risks have to do with different orders:

1) **Technical:** the DSS would need to get status updates from Transport for London about possible train cancellation services or buses being diverted because of roadworks
2) **Compliance:** the data collection and storage from the input about gender, age, type of student and residence would need to be done in a compliant way with General Data Protection Regulation (GDPR)

Conditional Probability Table (CPT)

In order to build the CPT, I am using a heuristic methodology based on personal and fellow foreign students' inputs. Figure 2 shows the CPT for each node and stage of the BN.

Fig.2 Conditional Probability Table



Female (G_0)	Male (G_1)
0.66	0.34

> = 30 y/o (A_0)	< 30 y/o (A_1)
0.45	0.55

	<= 3 miles (R_0)	> 3 miles (R_1)
A_0.G_0	0.75	0.25
A_0.G_1	0.45	0.55
A_1.G_0	0.55	0.45
A_1.G_1	0.35	0.65

	Full-time_ Student (ST_0)	Part-time_ Student (ST_1)
A_0	0.35	0.65
A_1	0.75	0.25

>=15C (G_0)	<15C (G_1)
0.45	0.55

	Train	Bus	Bicycle
R_0.ST_0.W_0	0.03	0.15	0.82
R_0.ST_1.W_0	0.08	0.2	0.72
R_0.ST_1.W_1	0.07	0.16	0.77
R_0.ST_0.W_1	0.12	0.45	0.43
R_1.ST_0.W_0	0.12	0.36	0.52
R_1.ST_1.W_0	0.26	0.37	0.37
R_1.ST_1.W_1	0.57	0.32	0.11
R_1.ST_0.W_1	0.22	0.41	0.37

Assumptions made for each variable:

- Gender: I defined the probability of female students at 0.66 based on a recent Goldsmiths diversity report that claims that 66% of students identify as female
- Age: I assume that over half of foreign students are less than 30 years old, however I set this probability at 0.55 since foreign students are more likely to be postgraduates so there's a high chance as well that they would be over 30.
- Residence: I am awarding a higher probability of students living within a 3 mile radius from campus being higher than farther away

- Student Type: I am assigning a higher probability of students who are above 30 years old being part-time students as they could be doing a postgraduate programme whilst working (as is my case)
- Weather: I am assuming that the temperature will be more likely below 15C during academic months (excluding August and September). I calculated the average temperature in London to be 14.5C as per these calculations.
- Transport, based on previous assumptions, I then put together the probabilities for mode of transport (Train, Bus or Bicycle) in terms of the 3 previous nodes: Student Type, Residence and Weather. Assumes that students who live less than 3 miles from campus would prioritize bicycle and bus.

Improvements

This DSS can be improved through integrations with 3rd party apps such as:

- Citymapper: to evaluate the 'quality' of bicycle route to boost the probabilities of using a bicycle depending on the student's residence
- Google Calendar: can assist in evaluating at what time of the day the student is travelling to campus, using as input the time for the first lecture the student needs to attend. If the student needs to be in campus at rush hour, then the model will prioritize using a bicycle or train to avoid getting stuck in traffic. Also in times of exams, it will prioritize the mode of transport that is faster and less likely to have delays.
- If-This-Then-That (IFTTT): this popular application can take in the probability result from DSS and then open the suggested route on Google Maps.
- National Rail, can provide the time that the next train headed towards New Cross Gate or New Cross is leaving.

PART II

Based on the DSS I built, I will ask 2 types of questions: prediction (2) and diagnostic (1).

A) **Prediction:** *what is the preferred transport method for a Female, part-time, above 30 years old student when the weather is below 15C?*
Answer: **Bus** with 44% probability

B) **Prediction:** *what is the preferred transport method for a Male student, who is above 30 years old and lives within 3 miles when the weather is above 15C?*
Answer: **Bicycle** with 78% probability

C) **Diagnostic:** *What type of student would prefer to travel by train when the weather is below 15C who is above 30 years old?*
Answer: This student has a 77% probability of being a **part-time** student.
Word Count: 1738 (excludes tables and figures)

References:

- Urwin, R., 2015. *Why are female cyclists more vulnerable to London's lorries?*. [online] Standard.co.uk. Available at: <https://www.standard.co.uk/lifestyle/london-life/why-women-seem-to-be-more-vulnerable-around-traffic-blackspots-in-london-10341420.html> [Accessed 21 December 2021].
- Goldsmiths, University of London. 2022. Diversity at Goldsmiths. [online] Available at: <https://www.gold.ac.uk/studyabroad/diversity/#:~:text=Students%20from%20all%20backgrounds%20are,21%25%20have%20a%20declared%20disability> [Accessed 2 January 2022].
- Pgmpy.org. 2022. Bayesian Network — pgmpy 0.1.15 documentation. [online] Available at: <https://pgmpy.org/models/bayesiannetwork.html> [Accessed 2 January 2022].
- Russell, S. and P. Norvig Artificial Intelligence: A Modern Approach. (Harlow: Pearson, 2016) 3rd edition.
- Bratko, I. Prolog programming for artificial intelligence. (Addison-Wesley, 2012) 4th edition.