Ride Sharing at FEUP

Catarina Ferreira
Faculty of Engineering
University of Porto, Portugal
up201506671@fe.up.pt

Gil Teixeira
Faculty of Engineering
University of Porto, Portugal
up201505735@fe.up.pt

Tiago Neves
Faculty of Engineering
University of Porto, Portugal
up201506203@fe.up.pt

***Abstract*—Everyday hundreds of students and professors go to FEUP using various means of transportation. The mean of transportation used impacts the environment in a serious way, although most of the time none of us think about it. This project aims to simulate the way a closed community comes to the same destination currently, using data collected from a survey, and simulate the impact that different matching techniques have regarding the emissions of harmful gases to our atmosphere. The matching techniques used are random, closest to starting point and a "real life" algorithm that combines using the relationships between the rider and its passengers (friends, colleagues of the same year, colleagues of the same course and other students) and the detour distance added to the rider's regular path. Our simulation uses the software NetLogo and the model Traffic Grid Goal which simulates traffic moving in a city grid.**

***Index Terms*—survey data analysis, factor analysis, clustering, ride sharing, matching algorithm, pollutant gases emissions, transportation, NetLogo**

I. Introduction

A. Context

Each year, the Portuguese population spends more than 7.8 millions of tons of fuel, of those 7.8, approximately 4.7 are spent in diesel and a bit more than 1 million are spent in petrol [12].

In 2017 in Porto Metropolitan Area the occupancy tax of a car was 1.56 in average and 1.36 for aggregates with only one person. It's also interesting to notice that 47.3% of students spend more than €100 per month on fuel [13].

The focus of this project is the segment of the population that we most resemble with, students, more precisely FEUP students. Students in Porto Metropolitan Area are the second most mobile segment, the first being the working class [13].

In this project the alternative of ride sharing will be tested to reduce our fuel consumption and, by consequence, our ecological footprint.

B. Problem definition

In this project the main problem we are tackling is how to reduce the emissions of pollutant gases with the use of ride sharing in a closed community in which users are commuting to same destination but with different starting points.

As a secondary goal we will also analyse, the different matching algorithms effect on the traffic and the "happiness" of the rider to give the ride to the people matched to him.

C. Expected contributions

We expect that this project will help verify some relations between the use of private transportation and ride sharing and how that affects the environment.

We also expect to sensitize people to the effects that something as simple as going to work/school every day using private means of transportation has on the environment and how they could decrease their ecological footprint if they were open to the idea of ride sharing. Our project could contribute to:

- Universities/Colleges since with our project we would know some incentives that motivate students in that demographic and the effects of practicing ride sharing;
- Raise awareness to the benefits of ride sharing;
- Academical purposes, for students that want to learn about NetLogo and the Traffic Grid Goal model.

D. Paper structure

In this paper we have the literature review in section II in which we present some papers done regarding carpooling and the impact of carpooling in the environment. We also did a Gap analysis using the papers presented and our own model.

In section IV we dive into our process of Data Collection and define our model. In the model, besides the characteristics of the model, we also enumerate and define the matching algorithms we thought about and that are going to be used in our simulation.

Section V first goes into a detailed description of all the data cleaning and processing the original Data Set went through to obtain useful information to be used in our model, such as Factor Analysis and Clustering algorithms.

In this section we also explain how the variables extracted from the survey answers are used and discuss our three matching algorithms which are Random allocation, Closest starting point and Real Life.

Continuing in this section we explain the shortest path algorithm, an algorithm that is used at the end of all the matching algorithms, to decide the shortest path the user could take passing through the nodes that were selected in the matching phase to pick passengers up and we also discuss the improvements made to the initial Traffic Grid Goal model.

Last section is section VI in which we describe the results of our simulation runs and present charts in which we can see the difference between the matching algorithms and our worst case scenario, which is the scenario used to model the answers we got on our survey.

II. Literature Review

The area of modelling and simulation allows to simulate a real system, using computers as a substitute for physical experiments. A model must be complex enough that simulates what it's trying to represent well enough but simple so that it's understandable and possible to perform the needed experiments. The simulation allows to evaluate the performance of a proposed theory. One of the methodologies used is agent-based modelling (ABM). This technique consists of individuals called agents. Each agent has a set of rules that he is obliged to follow. Furthermore, they operate concurrently and interact within an environment. NetLogo is a platform for Agent Based Model development that allows testing of hypotheses of real-life scenarios. (It is authored by Uri Wilensky and developed at the CCL). It already contains examples of various models. One of those models is Traffic Grid Goal which simulates traffic moving in a city grid [1].

A. *Impact of carpooling on the environment*

In [4], a study was done by a student to measure the total ecological footprint of the Ohio State University. Three parameters were chosen to do the Ecological Footprint Analysis: Energy, Transportation and Waste. This paper showed that transportation was the greatest cause to the Ecological Footprint with 72% of the total.

B. *Carpool matching models*

In an article [6] published by a team of researchers was presented an enhanced model that simulates urban traffic by adding carpooling functionalities. The objective of this project was to evaluate the benefit of carpooling using different scenarios with several parameters. With this simulations it was concluded that carpooling can have a bigger impact in highly populated cities than in the smaller ones and that it's essential that the amount of drivers be several times greater than the number of users waiting to be served. Also, the integration of the public transports services could strongly benefit carpooling system since many routes cannot be frequently covered by the drivers belonging to the carpooling system alone. The matching algorithm is as following, when a user requests ride, the system searches among the drivers that passes the user's current position and destination. These locations are added to the driver's destinations-list. If there is not a driver that can fulfil the desired path, then the user disappears after a period of time.

Within the context of supporting carpooling in large companies, [7] presented a framework for matching employees who are candidates for ride sharing. The matching frameworks has two main stages Preselection stage, where all feasible pairs are generated, namely the groups of users in which one of them has a car and driver's license, have an interceptable time of departure and arrival time windows and the maximum detour duration does not exceed a solo car trip. This step is followed by the Evaluation of the carpool group. In this stage each group is evaluated according to the product of the following scoring functions, the monetary cost of the journey, the excess travelling time and the value of the preferred departure/arrival time. Afterwards the best options are shown to the user.

C. *Carpooling Incentives*

Another project [5], conducted by researchers of the Texas Tech University, aims to find the environmental consequences of ride sharing in the United States of America. Carpooling can help reduce the number of cars on roads and consequently relieve traffic congestion. It also focuses on determining the variables that would increase the carpooling rate. For example, an increase of the gas price and a higher number of people per household causes the rate to be higher, on the other hand people more educated and with higher income have more intention to drive alone.

D. *Gap Analysis*

There are already several papers related to the impact of carpooling, quantifying the impact of an organization on the environment and modelling a carpooling multi-agent system. However, to the extent of our knowledge, there is not a paper that focus on characterizing a population, in this case FEUP's students, simulate a traffic model and verify which matching algorithm and incentives to carpooling would result in a lesser impact to the environment.

TABLE I
Gap analysis from related work

	Population characterization	Matching algorithms	Uses traffic Simulator	Carpooling Incentives and Impact
[2]	x			
[6]		x	x	x
[7]		x	x	
[5]	x			x
Our approach	x	x	x	x

III. Definition of Terms

- Rider, person that rides their own personal car to the destination, may or may not have passengers.
- Unmatched People, people that weren't selected during the matching process and don't have a private means of transportation. For that reason, they have to either walk or use a public means to get to their destination.
- Matching Possibility, a group of users that may be matched with a rider. The number of members of the possibility varies from 1 to the full capacity of the rider's car.
- Detour, approximately added distance to the rider's original path because of the passengers he must pick up.

IV. Methodological Approach

A. Data Collection and Analysis

We built a survey to characterize FEUP's population in areas related to ride sharing, with questions about how they get to the faculty, which means of transportation they use to get there, their willingness to practice ride sharing with their friends, colleagues and other students or members of the faculty that they didn't know, among other questions. Full survey in IX.

We sent out the survey to all the 8149 enrolled students of FEUP. During the 3 weeks the survey was open we collected 562 answers.

After collecting the answers, we then had to clean the Data set, by removing all answers that had blank values regarding the incentives and answers with strange values.

We then used Factor Analysis [15] to reduce the number of variables and to determine the influential underlying factors. The steps taken to factor analysis our Data Set were:

- Check if we could find factors in our Data Set;
- Choosing number of factors;
- Interpreting the relations of the underlying factors.

After the Factor Analysis we applied a Clustering algorithm (K-Means) to our Data Set and reached our final Data Set for most of the variables.

Some of the results of the deleted questions will be necessary to include into variables in the NetLogo model so, after the clustering, a new Data Set that only has the people that answered the required question is built and used to characterize the answers to that question.

B. Model

The base model used is the Traffic Grid Goal [1] in the NetLogo [2] Models Library, which is a good model of a real-world city and the interactions between real world cars. Our model comprises of two types of users - Riders and Passengers. Riders have a private vehicle with available seats (1-6) and Passengers are users that need a ride to get to the faculty.

The first step in our model is to match the Riders to the Passengers, according to the matching policy used, and then each Rider picks the Passengers they got matched with from their house and then goes to the destination- the faculty.

In each iteration we calculate the amount of harmful gases that are being released to the atmosphere, due to the transportation of the users to the faculty. Different matching algorithms are compared to conclude which is the best in terms of the emissions.

The possible matching algorithms are:

- None, no matching is used, all the people that have a car use it to get to their destination and all the other people have to go by walking or public transport;
- Random, each person with a car is matched with as many people as possible (the car's capacity), the people chosen to be matched are random;
- Closest starting point, people are matched to a rider by how close they are to the rider's starting point (similar to matching between neighbours);
- Real life, this algorithm tries to mimic the aspects that any of us would take into consideration when practicing ride sharing in our everyday life. This algorithm takes into consideration 3 factors:
  - The existing relationship between a rider and the potential passenger, such as friends or students enrolled in the same year and same degree. People usually prefer to travel with friends rather than strangers (this preference is taken from the data collected in the survey).
  - The detour the rider would need to make to pick the passenger up, in terms of distance the rider would need to make, in comparison with the usual distance he makes.
  - How full the vehicle is, for two different matching possibilities: if the other factors have the same score, the matching in which the cars capacity is filled the most should be picked.

C. Input Variables

- The Matching algorithm used in that run;
- Number of users of each type (types identified by the clustering phase on IV-A)
- Grid size in terms of number of roads there are;

D. Output Variables

- Total Carbon Monoxide emissions;
- Number of riders;
- Number of stopped cars;
- Average car speed;
- Average wait time of cars;
- Average score of riders;
- Average score of unmatched users;
- Number of unmatched users.

V. Implementation

A. In depth Factor Analysis

Before doing the Factor Analysis we checked if there were missing responses in our Data Set and then removed the entire response if there were any missing responses for any question. After cleaning up the data we then were able to use Factor Analysis, as mentioned in IV-A.

First, we did two adequacy tests to check if we could find factors in our Data Set. (Used FactorAnalyzer Python Package [3])

- Bartletts Test [10] - obtained 0 value, meaning that we have relationships between our variables.
- Kaiser-Meyer-Olkin Test [11] - Determines the adequacy for each observed variable and for the complete model. We obtained the value 0.7587, meaning that it is suitable for Factor Analysis

Next step was choosing the number of factors using Kaiser criterion, which is based on eigenvalues. An Eigenvalue is a measure of how much of the variance of the observed variables a factor explains. The number of factors used should be the

”elbow” of the plot (Figure 1), but in our case the ”elbow” wasn’t distinguishable enough.

To help with the selection of the number of factors, our next step was to plot the cumulative variance dependent on the number of factors (Figure 2)

Based on the plot, we chose 9 factors because after 9 factors the plot’s slope starts to decrease. The cumulative variance is approximately 60%.

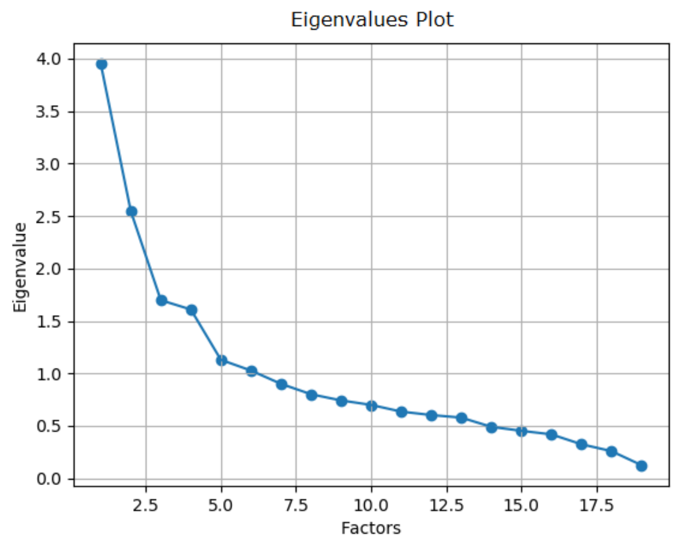


Fig. 1. Eigenvalues Plot

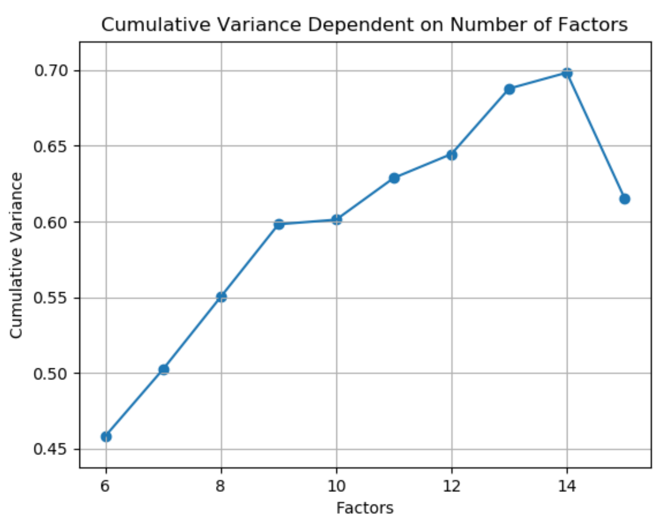


Fig. 2. Cumulative Variance Plot

Now knowing the number of factors, we performed the factor analysis and observed the factor loadings to check for meaningful relation between the variables. Factor loading is a matrix which shows the relationship of each variable to the underlying factor. With the Factor loadings we concluded the following:

- Factor 1 had high factor loadings for ”Willingness to use ride sharing”, and Willingness to ride share with the ”friends”, ”year colleagues”, ”course colleagues” and other students from FEUP (Friendliness), students who are willing to use ride sharing are also likely to be willing to ride share with their friends and year colleagues.
- Factor 2 had high factor loadings for ”How many years have you had your driver’s license for”, ”Having private transportation”, ”How much money do you spend on fuel” and ”How much money do you spend in public transportation”. It’s interesting to point out that there is a high relation between having a private transportation and money spent on fuel, so if the user has a private means of transport than they will probably spent more money on fuel. We can also see that there is an inverse relationship between money spent on fuel and on public transportation, which means that if they spent a lot of money on fuel they will probably spend less in public transportation (Transport)
- Factor 3 had high factor loadings for all the incentive options that we had (Credit for the printing system, credit in the coffee machines, credit in the cafeteria, credit in the FEUP store, discount in the tuition and priority parking for the users offering rides (Susceptible), meaning that people who value one incentive will probably value the others in a positive way.
- Factor 4 had high factor loading for ”How do you go to the faculty” and ”Money spent in public transportation”. We have four different values to represent the different ways one could use to go to the faculty, from 0 meaning walking, which is the closest and 4 meaning using both private and public transportation probably used by people further away from the destination. This means that there is a direct relation between the means of transportation used and how much money a user spends in public transportation (Public User), for example a user that walks to the destination (value 0) will spend less money than an user who picks a higher value.
- Factor 5 had high factor loadings for ”How many years have you had your driver’s license for” and ”Having private transportation” which means that the amount of years the user has a driver’s license is related to the ownership of a private means of transportation (Car Owner), more years with a driver’s licence more likely to own a private vehicle.
- Factor 6 had high factor loading for ”Willingness to ride share with friends” and ”Willingness to ride share with same year colleagues” which means that users who share their ride with friends have a direct relation with sharing ride with same year colleagues (Introvert).
- Factor 7 had high factor loadings for the incentive options regarding a discount in the tuition and priority parking for the users offering ride (Parking).
- Factor 8 had high factor loadings for ”Do you already use ride sharing” and ”Willingness to use ride sharing”, which means that if someone already uses ride sharing

they are also willing to use ride sharing (Willing).

- Factor 9 had high factor loadings for "How do you go to the faculty", "How far away do you live from the faculty (in km)", "Do you live inside or outside Porto", "How much money do you spend on fuel" and "How much money do you spend in public transportation", all these factors are related in a positive way (Extra Urban), for example a person that uses public and private means of transportation (high value) is also likely to live far away from the faculty, live outside Porto, spend more money on fuel and on public transportation fares.

B. *In depth Clustering*

To classify our users in different groups in our model, we used the K-Means [9] clustering algorithm.

To check how many clusters should be present in our model, we plotted the score of the mean Silhouette Coefficient [16] for each number of clusters(Figure 3), we ran the algorithm 30 times in order to obtain the average for each number of clusters because the score results would vary.

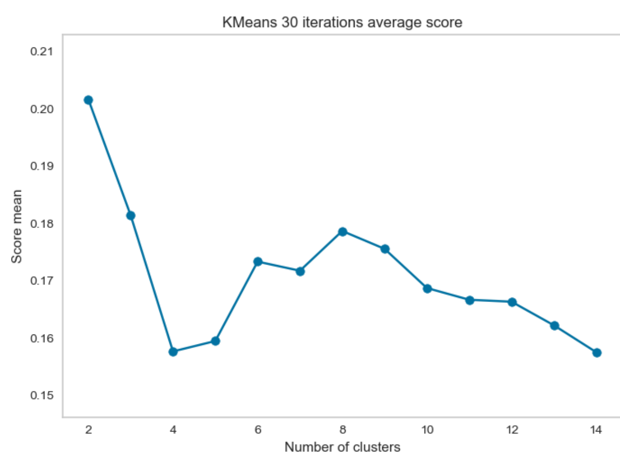


Fig. 3. Mean Silhouette Coefficient plot

Based on the plot we chose 8 clusters.

We also used the Silhouette Visualizer to check the silhouette coefficient for each sample and the density of each cluster (Figure 4).

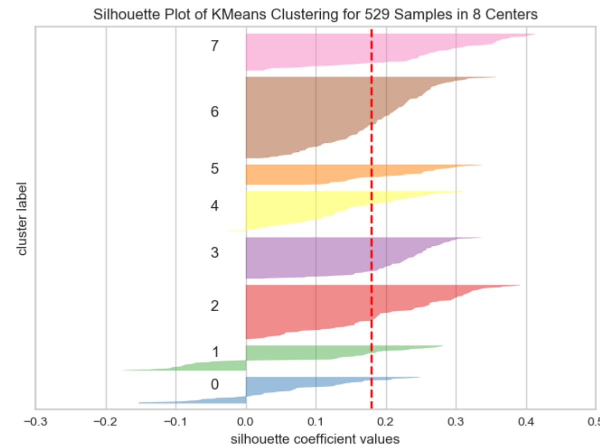


Fig. 4. Silhouette Visualizer

C. *Applying Data to Model*

Each of the clusters found in V-B were then used to classify the users of our Data Set into different clusters with different values for variables like the distance to the faculty and different behaviours regarding the willingness to ride share with friends or colleagues.

Using the distribution of each variable for each cluster we then have the necessary values to identify a user in our model. Every user has these variables and their distribution is different depending on which cluster the user belongs too:

- Willingness to ride share with:
  - Friends (discrete, 0 to 5)
  - Year colleagues, colleagues of FEUP in the same year and course as them (discrete, 0 to 5)
  - Degree colleagues, colleagues of FEUP in the same course as them (discrete, 0 to 5)
  - FEUP colleagues, any other person of FEUP (discrete, 0 to 5)

- Vehicle capacity (discrete, 0 to 5, from the survey data there was small chance for capacity 6, users with private vehicle belonging to cluster 3 had a 2.3% chance of having capacity 6 , but the matching for the users with this capacity would take too much time because of the complexity of the real life matching algorithm V-D)
- Distance to destination (continuous, detail in V-C1)
- Vehicle ownership (discrete, 0 or 1)

1) Distance to destination distribution: The first step was to look at the histograms representing the travel distance for each cluster. (Figure 5 of Cluster 0 to exemplify)

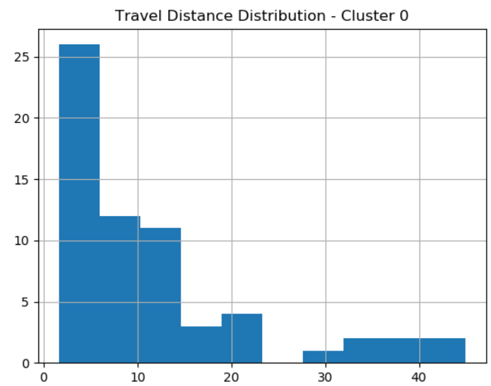


Fig. 5. Distribution of the distance variable of cluster 0

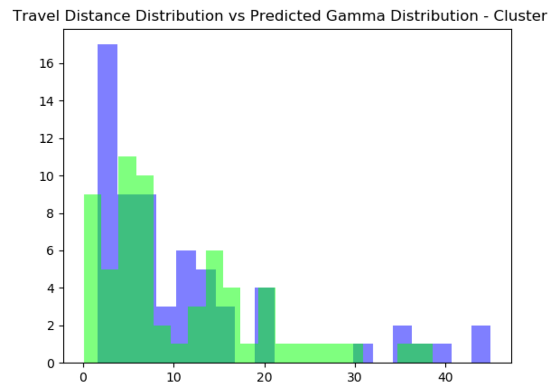


Fig. 6. Comparing original(blue) vs our Gamma distribution(green)

Analysing the histograms and comparing them to the random functions using distributions present in NetLogo (Exponential, Gamma, Normal and Poisson) we came to the conclusion that the distribution that fitted our data best was the Gamma distribution.

We obtained the following values for the mean of the distribution and the variance of the distribution of the distance variable for each cluster:

- Cluster 0 - 11.32, 115.33
- Cluster 1 - 12.05, 129.32
- Cluster 2 - 1.42, 0.69
- Cluster 3 - 15.67, 155.62
- Cluster 4 - 8.11, 98.15
- Cluster 5 - 14.46, 190.14
- Cluster 6 - 12.99736842105263 169.39657202216068
- Cluster 7 - 1.42, 0.46

With those values we were able to compute the inputs of the random gamma distribution function:

- $alpha = mean * mean/variance$
- $lambda = 1/(variance/mean)$

Using those values, we compared our gamma distribution with the original distribution (Figure 6 of Cluster 0 to exemplify). We concluded that the Gamma Distribution is a good approximation to our original distribution for each cluster.

These values are used to characterize the distance to destination variable for each user.

D. Matching

In our project we are going to simulate using multiple matching techniques.

For all the techniques: Passengers are not matched to a Rider with full occupation on their vehicle

- **None**
  No matching is really used.
  Every person that has a car, is a rider and goes in their car to the destination. Every person that doesn't have a car is discarded.
- **Random allocation**:
  This matching algorithm only takes into consideration the number of seats available in the rider's car.
  For each rider:
  - The algorithm selects capacity number of users, to fill all the available seats in the vehicle. Capacity is the number of available seats in the rider's vehicle.

  This selection doesn't take into consideration the relationship between the users or even the distance between the rider's and the user's pickup location. It is completely random as the name indicates and focuses only on giving, as many people as possible a ride to the destination.
- **Closest starting point**:
  For each rider:
  - A triple is created with the rider, the potential passenger and the distance that that user pickup location is from the rider's starting point;
  - A list of all the potential triples is constructed.

  With the final list we then order the elements of the list in ascending order and finally we start to process the list, assigning the passengers to the riders and assigning at most capacity passengers to each rider. This algorithm is different from the others in this list in the sense that we try to assign people according to the "best" ride they could take in terms of minimum distance to the rider's starting point.

This is also the only algorithm we implemented that doesn't try to fill a rider's car capacity at once, instead it chooses, in each iteration of an internal for loop, the matching possibility with the shortest distance between the two people.

- **Real Life**
  This algorithm is the closest out of all 3 that mimics the aspects that any of us would take into consideration when practicing ride sharing in our everyday life.
  This algorithm takes into consideration 3 factors:
  - The existing relationship between a rider and the potential passenger, such as friends or students that enrolled in the same year and same degree.
  - The detour the rider would need to make to pick the passenger up, in terms of distance the rider would need to make, in comparison with the usual distance they make;
  - How many seats are available/ how full the vehicle is.

  The steps are as follow:
  - For each rider we calculate the set of all potential passengers that are within the maximum distance that the rider is willing to go - the rider has a maximum distance with which he feels comfortable picking people up from;
  - The next step is to reduce the number of possible passengers. For that we score each potential passenger by multiplying the relationship and the detour distance score and then pick until 18 possible passengers with the highest scores;
  - After we have reduced the number of user's to be taken into consideration the score for each possible combination of those 18 people (combinations from 1 until the capacity of the car) are calculated and the combination/group of users with the highest score is chosen to be matched with the rider.

  Each possible matching is scored using the multiplication of the 3 factors mentioned previously. Regarding the relationship aspect, the **relationship score** starts at 1 and then, for each passenger:
  - If the passenger and the rider are friends: the relationship score is multiplied with the willingness to ride share with friends of the rider;
  - If they are same year and degree colleagues: the relationship score is multiplied with the willingness to ride share with year colleagues of the rider;
  - If they are students of the same degree: the relationship score is multiplied with the willingness to ride share with same degree colleagues of the rider;
  - If they are just colleagues at FEUP the relationship score is multiplied with the willingness to ride share with FEUP colleagues.

  Regarding the **detour distance score**: For each possible matching of each rider: the rider has a distance within which he does not mind to travel and pick users up, called minimum distance. Each rider using the algorithm explained at V-E calculates the full path distance to pick up all the people in that matching possibility.
  - If the path distance is less than the minimum distance stated by the rider, than the score is 1;
  - If the path distance is between the minimum distance and twice the value of the minimum distance, then the $score$ is

    $$e^{((sum\ of\ distances - minimum\_distance) * \alpha)}$$

    In which $alpha$ is

    $$ln(r\_min)/minimum\_distance$$

    and r_min is 0.1.
    r_min is the minimum score a rider can have in this factor. This formula is from [14].
  - If the path distance is greater than or equal to twice the value of the minimum distance, then the score is r_min.

  The **number of available seats** is also taken into consideration to the score.
  - If originally the rider didn't have any available space in his vehicle than the value 1 is multiplied to the score;
  - Otherwise the division of the number of elements of the group by the capacity of the vehicle is multiplied to the score;

In all these algorithms, one of two situations can happen to the user's which aren't selected to be matched:

1) If the user has their own private vehicle, then the user simply uses their vehicle;
2) If the user doesn't own a vehicle, then the user will either walk or use a public means of transportation to go to their desired destination. Being discarded in the model.

E. Shortest Path

In the matching algorithms there was a need for a shortest path algorithm.

This algorithm runs at the end of every type of matching and is used to choose the order that the rider should pick up the people matched to him so that the distance travelled by the rider is the shortest possible. This algorithm is also used in the real-life algorithm to find the detour score

A starting point, ending point and some required nodes that it needs to pass through are given. The objective of the algorithm is to find the order that the required nodes should be travelled through that gives the shortest path.

In our context, we decided to use a brute force approach because in our worst case only has 5! = 120 possible orders. For all those possible orders the approximate path distance is returned: the summation of the distance in a straight line between the consecutive points of the path.

- $numberNodes$ is the number of nodes in the path, including the start and end nodes;

- $path$ is an ordered array of the path nodes, start node is the first and end node is the last;
- $distance(p1, p2)$ is the distance in a straight line between the points p1 and p2.

$$pathDistance = \sum_{n=0}^{numberNodes-1} distance(path_n, path_{n+1})$$

The approximate path distance doesn't consider the road layout.

The shortest path is the order of nodes with the lowest path distance.

F. Improvements to the model

Initially the Traffic Grid model wasn't using a variable number of stops, each user had just two stops, their "house" which was their initial stop and their "work" which was their destination. Since we have matching in our model, which implies the user to have multiple stops, we had to incorporate that in the original model.

We achieved that by introducing a list representing all the stops, including the destination, and a variable which represents the index in the list of the stop the user is headed to. The user would start in their initial position, their house, and then sequentially go to the next stop, iterating through the list of stops. The user advances though to the next stop after arriving at the current stop and when it arrives to the destination the user disappears from the grid.

In our implementation of the variable number of stops we faced an issue, in some occasions the riders would travel back and forth in a road instead of going to the next goal. This happened, for example, in the occasion presented in the figure 7 (rider is the car in blue and the brown patch is the rider's next stop), where using the algorithm used by the model to give the rider the next patch that he should travel to, the rider would be stuck between two patches.

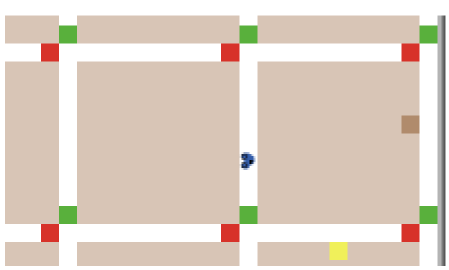


Fig. 7. Situation where a rider would be stuck going to the next stop

This happens because the next-patch function in the Traffic Grid Goal model is greedy and it chooses the next patch based only on its distance in a straight line to the goal, trying to minimize it. This implementation wouldn't choose the patches on the intersections above or below the rider because their distance in a straight line to the goal is greater than the patches at the middle of the road that he is in.

Solving this problem in the next-patch algorithm was too complex for our application, but we had to resolve this issue to be able to run our model multiple times, automatically.

Our solution was to have a list of the last 5 patches that the rider was at and if for 50 ticks those patches don't change, meaning that the rider didn't move to a new patch during the 50 ticks, the rider skips that goal.

This implementation has false positives but the consequences to the results are minimal.

G. Emissions Function

In [8], the emissions level of car, $Footprint$ expressed in g of CO per kilometre, is characterized as:

$$Footprint = \frac{a + cV + eV^2}{1 + bV + dV^2}$$

where:

a-e are parameters factors for a car with a capacity of 1.4 L-2.0 L;

V is the average speed;

TABLE II
PARAMETER OF EMISSION FACTORS FOR CARBON MONOXIDE (CO)

Pollutant	a	b	c	d	e
CO	71.7	35.4	11.4	-0.248	0

VI. RESULTS & DISCUSSION

We decided to have 40 runs of each of the 3 matching algorithms and the none algorithm, with each of the following numbers of users: 10,50,75,100,125,150,175,200,250,300,350. The ratio of each cluster was kept the same as the initial survey population, for every run.

Ratios:

- Cluster 0 - 0.122873346
- Cluster 1 - 0.105860113
- Cluster 2 - 0.258979206
- Cluster 3 - 0.079395085
- Cluster 4 - 0.18147448
- Cluster 5 - 0.071833648
- Cluster 6 - 0.060491493
- Cluster 7 - 0.119092628

In each run we collected the results for the total number of Carbon Monoxide emissions, the number of riders, the average score of the riders, the number of unmatched users and the average score of unmatched users.

We decided to collect the number of Carbon Monoxide (CO) emissions since although it's not one of the main components of combustion gas it still has very important indirect effects on global warming.

The **average score of a rider** is obtained using the scoring method explained in V-D which takes into consideration the relationship between rider and passengers, the ratio of available seats after matching and the distance the rider has to travel with passengers in comparison if he were to travel alone.

The **average score of an unmatched person** is obtained using only their distance to the destination in a straight line, it uses the same general formula as the detour distance score but with different parameters.

- If the path distance is less than 3 km, than the score is 1;
- If the path distance is between the minimum distance and 40 km, then the $score$ is

  $$e^{((sum\ of\ distances-minimum\_distance)*\alpha)}$$

  In which $alpha$ is

  $$ln(r\_min)/minimum\_distance$$

  and r_min is 0.1
- If the path distance is greater than 40 km, then the score is r_min

After the runs we have the following charts in which:

- Red Line - None, worst case scenario. Every user with a private vehicle uses it go to their destination, alone.
- Orange line - Random Matching is the first matching algorithm in V-D
- Green line - Minimum Distance is the second matching algorithm in V-D
- Blue line - Real Life is the third and final matching algorithm in V-D. It's named Real Life since we feel this algorithm is the closest out of all 3 that mimics more how people act in real life and the aspects they take into consideration.

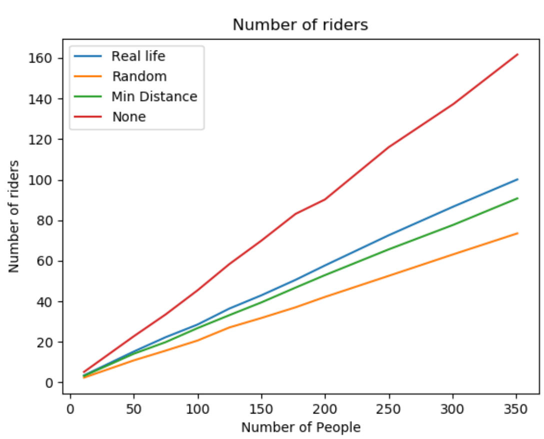


Fig. 8. Number of riders in each matching alternative and worst case

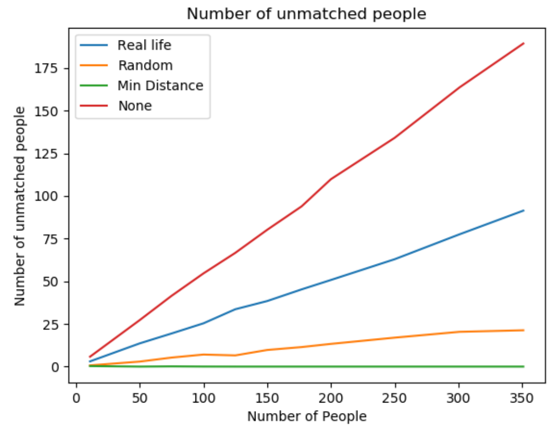


Fig. 9. Number of unmatched people in each matching alternative and worst case

With charts 8 and 9 we see that the number of unmatched users is much higher in the Real Life algorithm in comparison with the other matching algorithms since Real Life takes into consideration the factors we previously discussed, which makes it harder to match users, instead of Min Distance for example which doesn't take in to account the rider's willingness to share rides.

We can also notice that almost all the matching algorithms with a higher number of riders also have a higher number of unmatched people.

This happens because the chart 8 is a good indicator of how well the matching algorithm works. Of the people that are matched to a rider, some of them own a private vehicle, but since they are matched that means that they aren't going to use their vehicle. If those people didn't get a match, then they would use their vehicle to travel to the destination. This means that less riders imply more people matched.

The exception to this rule is the Min Distance algorithm because it works in a different way, it iterates the possible people to match to a rider while the other algorithms iterate through the people with a vehicle and match as many people as possible to them. This means that in the Min Distance algorithm the number of people per car is lower.

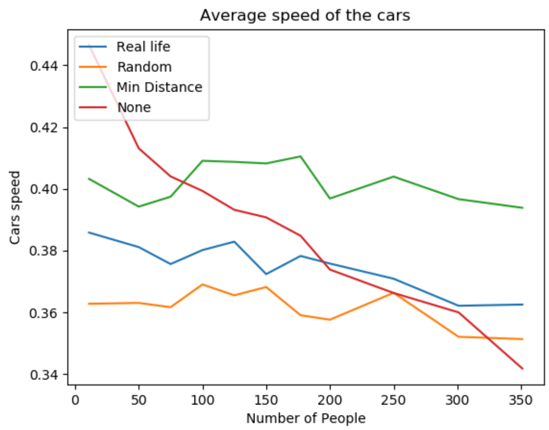


Fig. 10. Cars speed in each matching alternative and worst case

We can also relate the chart 10 with the number of riders. The average car speed chart relates to the amount of traffic on the road, but it has some nuances. With a low number of people, the matching algorithm with the highest average speed is also the one with the highest number of cars on the road (figure 8). We think this happens because with the None algorithm the cars pass through less traffic lights since they can just go to the destination without any detours.

With the growing number of riders that property of the None algorithm is quickly outweighed by the traffic caused by the amount of cars on the road.

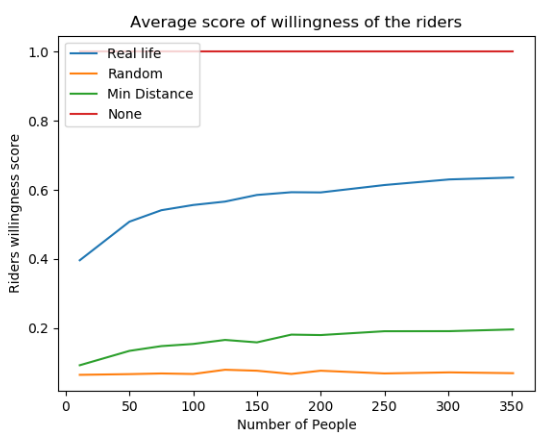


Fig. 11. Rider's Score in each matching alternative and worst case

This chart 11 can be used to measure the "happiness" that the riders have in each of the matching strategies used, using the 3 factors mentioned in V-D. In the chart we can see that there is a significant difference between the Real life strategy and the other algorithms since the Real Life is the only strategy that takes into consideration some personal aspects about the rider such as his relationships with others and the distance between the rider and his passengers, while Random is, as the name indicates, completely random and Min Distance solely focuses on the distance between the rider's starting point and the pick-up location of his passengers.

In terms of the average score the highest matching strategy is the Real Life algorithm, as explained before and in second place we have the Min Distance strategy, since it takes into consideration the distance between rider and passenger, followed closely by the Random strategy.

The worst case scenario, the situation in which no matching algorithm is used, has the highest score since the rider doesn't need to make any deviation from the original path he always takes to the destination. He doesn't need to make any additional stops or share his space with anyone, so that's the ideal situation for the rider. In this paper we didn't consider incentives for the riders.

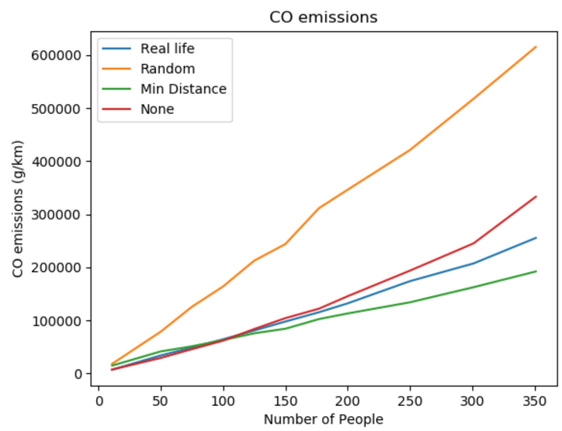


Fig. 12. Carbon Monoxide emissions in each matching alternative and worst case

In this chart 12 we can see that the two matching strategies that are more reasonable and take into consideration more aspects of the user - Real Life and Min Distance - produce less harmful emissions than the Worst case scenario, in which no matching algorithm is used, and the Random strategy, which has significantly higher emissions than the rest of the algorithms. We assume that the CO emitted by the unmatched people is insignificant.

1) The Random matching algorithm is the one with the highest amount of CO emissions, much higher than the rest of the matching algorithms since in this matching strategy the riders many times have their car with full capacity and they can go pick up any user in need, no matter how far away from each other they may be, spending a lot of fuel in the process.
2) None, this is a decent alternative because the riders go to their destination without detours. The "comfort" of all the people that don't have a vehicle is lower than in the other algorithms because not having a vehicle means that they all need to walk or use public transports.
3) Real-Life scenario, this alternative prioritizes not only the distance between users but also the relationship between them. It's better than no matching in terms

of CO emissions, not compromising the willingness of the rider, and increases the "comfort" of the passengers because then they don't need to walk or use public transports.

4) Min Distance, it is the best in terms of CO emissions because it doesn't take into account the riders willingness to make that ride.

VII. Conclusion

In this paper, we presented the theme as well as the objectives of the simulation. The current state of art was taken in consideration to visualize what has been done in this field and what this project needs to do.

The best of the alternatives we have presented in this work, is the Minimum Distance, if we only take into consideration the results of the emissions.

However this alternative doesn't consider the willingness of the driver to make that ride, for example in real life everyone takes into consideration the person or people with who they would share their ride with and how great is the detour.

Taking into consideration those factors the Real Life alternative is good, since it reduces the number of riders on the road compared to the scenario representing the population right now, which leads to a higher average speed of the cars with a lower traffic volume. Although it has the drawback of higher percentage of unmatched people, in comparison with the Random or Min Distance algorithms, the "happiness" of those that are matched is higher than the other matching algorithms. Finally, for a higher volume of users, the reduction in CO emissions is also noticeable.

VIII. Future Work

In the future this work could potentially be improved if the algorithms took even more concepts in consideration such as the schedule of the users and the journey back home.

It would also be interesting to explore the distribution and know with the clusters partitioning, the clusters that are more willing to ride share with different levels of relationship with other users.

Another aspect that would be interesting to take into consideration would be the incentives to ride share, such as credit for the coffee machines or the printing system, and see how it would influence the willingness the riders have and the matching solutions.

Acknowledgment

We thank our colleagues from FEUP who answered our survey that greatly assisted the research.

We wish to acknowledge Informatics Engineering Department (DEI) from FEUP for granting the prizes that greatly increased the number of responses.

We would also like to thank Professor Rosaldo J. F. Rossetti who was involved in the validation of the survey for this research project and for convincing the department of the importance of this project and thus providing the rewards for the survey. Without his participation and input, the survey could not have been successfully conducted.

Finally, we must express our gratitude to the Researcher Zafeiris Kokkinogenis for sharing insightful comments and encouragement with us during the course of this research and steered us in the right the direction whenever he thought we needed it.

References

[1] Rand, W., Wilensky, U. (2008). NetLogo Traffic Grid Goal model. http://ccl.northwestern.edu/netlogo/models/TrafficGridGoal. Center for Connected Learning and Computer-Based Modeling, Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL.

[2] Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

[3] Jeremy Biggs, FactorAnalyzer package for Python, https://factor-analyzer.readthedocs.io/en/latest/index.html# [Online; accessed 2019-05-10].

[4] Janis, Jaclyn, Quantifying the Ecological Footprint of the Ohio State University, 2007.

[5] Javid, Roxana and Nejat, Ali and Salari, Mahmoud, The Environmental Impacts of Carpooling in the United States, 2016.

[6] Marcelo Armendáriz and Juan C. Burguillo and Ana Peleteiro-Ramallo and Gérald Arnould and Djamel Khadraoui, Carpooling: A Multi-Agent Simulation In Netlogo, 2011.

[7] Iftikhar Hussain, Luk Knapen, Tom Bellemans, Davy Janssens and Geert Wets, The Matching Framework for Employees to Support Carpooling in the Context of Large Companies, 2018.

[8] Kan, Zihan et al. Estimating Vehicle Fuel Consumption and Emissions Using GPS Big Data. International journal of environmental research and public health vol. 15,4 566. 21 Mar. 2018, doi:10.3390/ijerph15040566

[9] Lloyd, Stuart P. "Least squares quantization in PCM." Information Theory, IEEE Transactions on 28.2 (1982): 129-137.

[10] Snedecor, George W. and Cochran, William G. (1989), Statistical Methods, Eighth Edition, Iowa State University Press.

[11] Kaiser, H. 1974. An index of factor simplicity. Psychometrika 39: 3136.

[12] Paulo Salteiro Rodrigues Direo Geral de Energia e Geologia.Vendas de Produtos de Petrleo em Portugal 1970-2016. 2018.

[13] I.P Instituto Nacional de Estatstica. Mobilidade e funcionalidade do territrio nas reas Metropolitanas do Porto e de Lisboa 2017. 2018.

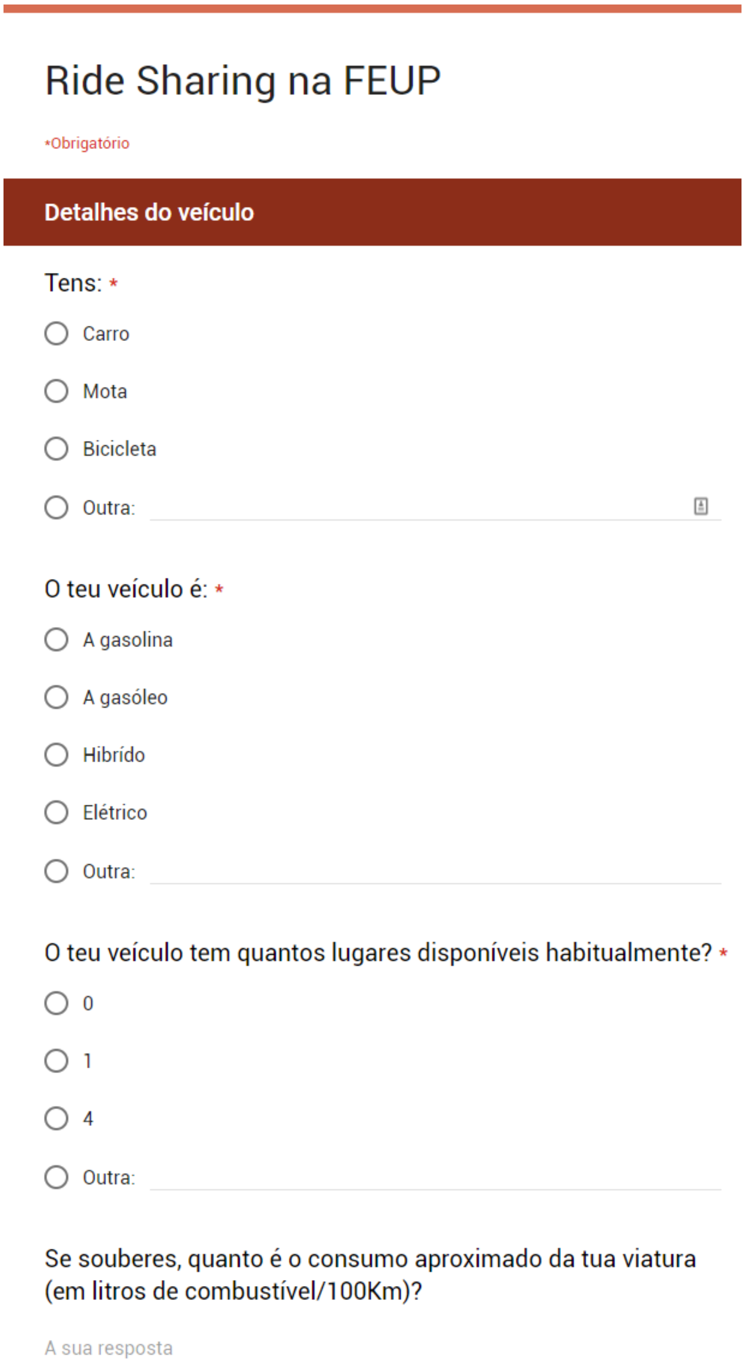
[14] Hussain, Iftikhar & Knapen, Luk& Bellemans, Tom& Janssens, Davy & Wets, Geert. (2017). Employees Matching to Support Carpooling in Context of Large Companies.

[15] Jolliffe I.T. (1986) Principal Component Analysis and Factor Analysis. In: Principal Component Analysis. Springer Series in Statistics. Springer, New York, NY

[16] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

IX. Appendix

A. Survey



Ride Sharing na FEUP

*Obrigatório

Detalhes do veículo

Tens: *

- Carro
- Mota
- Bicicleta
- Outra:

O teu veículo é: *

- A gasolina
- A gasóleo
- Hibrído
- Elétrico
- Outra:

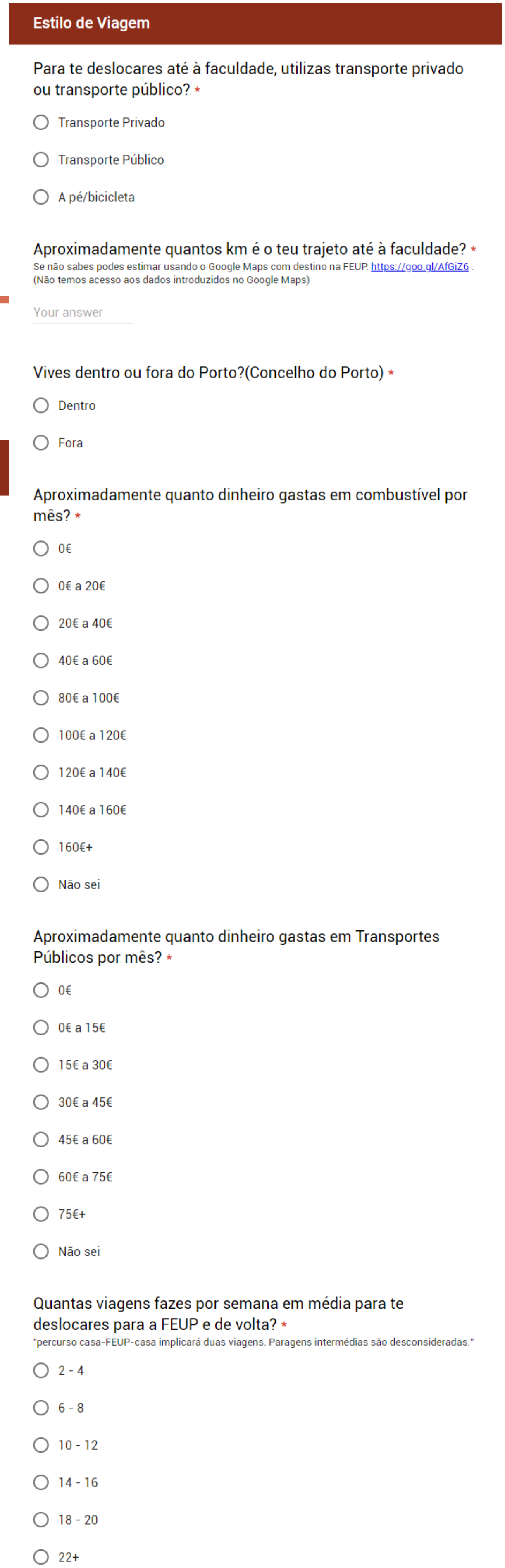
O teu veículo tem quantos lugares disponíveis habitualmente? *

- 0
- 1
- 4
- Outra:

Se souberes, quanto é o consumo aproximado da tua viatura (em litros de combustível/100Km)?

A sua resposta

Fig. 13. Vehicle Characteristics



Estilo de Viagem

Para te deslocares até à faculdade, utilizas transporte privado ou transporte público? *

- Transporte Privado
- Transporte Público
- A pé/bicicleta

Aproximadamente quantos km é o teu trajeto até à faculdade? *

Se não sabes podes estimar usando o Google Maps com destino na FEUP. https://goo.gl/AfGiZ6 . (Não temos acesso aos dados introduzidos no Google Maps)

Your answer

Vives dentro ou fora do Porto?(Concelho do Porto) *

- Dentro
- Fora

Aproximadamente quanto dinheiro gastas em combustível por mês? *

- 0€
- 0€ a 20€
- 20€ a 40€
- 40€ a 60€
- 80€ a 100€
- 100€ a 120€
- 120€ a 140€
- 140€ a 160€
- 160€+
- Não sei

Aproximadamente quanto dinheiro gastas em Transportes Públicos por mês? *

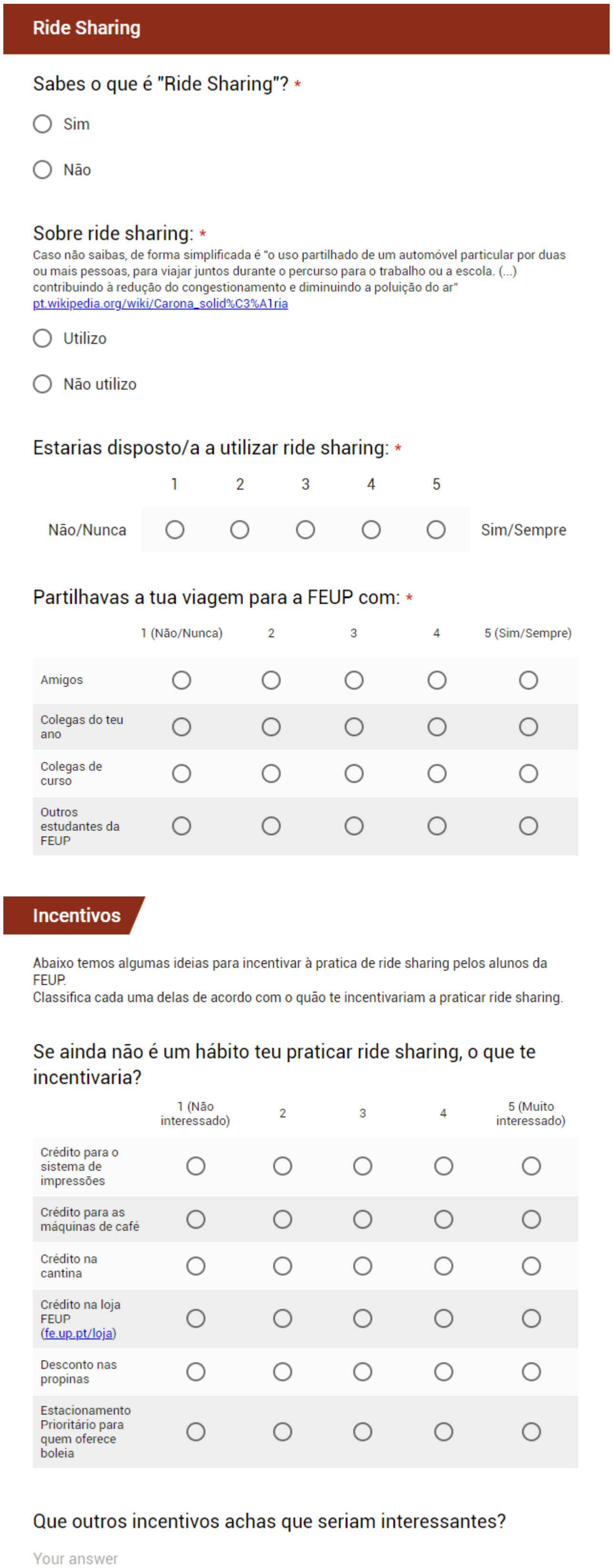
- 0€
- 0€ a 15€
- 15€ a 30€
- 30€ a 45€
- 45€ a 60€
- 60€ a 75€
- 75€+
- Não sei

Quantas viagens fazes por semana em média para te deslocares para a FEUP e de volta? *

"percurso casa-FEUP-casa implicará duas viagens. Paragens intermédias são desconsideradas."

- 2 - 4
- 6 - 8
- 10 - 12
- 14 - 16
- 18 - 20
- 22+

Fig. 14. Commuting Characteristics



Ride Sharing

Sabes o que é "Ride Sharing"? *

- ◯ Sim
- ◯ Não

Sobre ride sharing: *

Caso não saibas, de forma simplificada é "o uso partilhado de um automóvel particular por duas ou mais pessoas, para viajar juntos durante o percurso para o trabalho ou a escola. (...) contribuindo à redução do congestionamento e diminuindo a poluição do ar"
pt.wikipedia.org/wiki/Carona_solid%C3%A1ria

- ◯ Utilizo
- ◯ Não utilizo

Estarias disposto/a a utilizar ride sharing: *

	1	2	3	4	5	
Não/Nunca	◯	◯	◯	◯	◯	Sim/Sempre

Partilhavas a tua viagem para a FEUP com: *

	1 (Não/Nunca)	2	3	4	5 (Sim/Sempre)
Amigos	◯	◯	◯	◯	◯
Colegas do teu ano	◯	◯	◯	◯	◯
Colegas de curso	◯	◯	◯	◯	◯
Outros estudantes da FEUP	◯	◯	◯	◯	◯

Incentivos

Abaixo temos algumas ideias para incentivar à pratica de ride sharing pelos alunos da FEUP.
Classifica cada uma delas de acordo com o quão te incentivariam a praticar ride sharing.

Se ainda não é um hábito teu praticar ride sharing, o que te incentivaria?

	1 (Não interessado)	2	3	4	5 (Muito interessado)
Crédito para o sistema de impressões	◯	◯	◯	◯	◯
Crédito para as máquinas de café	◯	◯	◯	◯	◯
Crédito na cantina	◯	◯	◯	◯	◯
Crédito na loja FEUP (fe.up.pt/loja)	◯	◯	◯	◯	◯
Desconto nas propinas	◯	◯	◯	◯	◯
Estacionamento Prioritário para quem oferece boleia	◯	◯	◯	◯	◯

Que outros incentivos achas que seriam interessantes?

Your answer

Fig. 15. Ride Sharing Questions