CLEI electronic journal, Volume 24, Number 2, Paper 10, July 2021

# UCarpooling: decongesting traffic through carpooling using automatic pairings

**Alejandro Lugo, Nathalie Aquino, Magalí González, Luca Cernuzzi**
Universidad Católica "Nuestra Señora de la Asunción"
Asunción, Paraguay
{alejandro.lugo, nathalie.aquino, mgonzalez, lcernuzz}@uc.edu.py

and

**Ronald Chenú-Abente**
University of Trento
Trento, Italy
chenu@disi.unitn.it

**Abstract**

A low average number of people per private vehicle and inappropriate road infrastructure results in heavy traffic that wastes space, time and money for the people involved. To optimize these resources, it is intended to promote carpooling between people who share the same destination, for example, colleagues at work or students at a university. This paper presents UCarpooling, a matching system for commuting between people in the same institution. UCarpooling is aimed at optimizing the number of passengers in vehicles during routine trips to and from work or study. The difference with respect to other similar proposals is that UCarpooling takes into account logistical details (place of departure, time of entry, etc.) and personal traits (if you smoke, what genres of music you listen to, etc.) as variables to calculate the percentage compatibility that different people have to carry out a carpool. A simulation of the use of UCarpooling in a university in Asuncion, Paraguay, yields favorable data reaching the conclusion that its adoption is quite beneficial for the institution that adopts it, the people who use it, and the cities where it is adopted.

**Keywords:** carpool, commuting, matcher, simulation

## 1 Introduction

According to a poll made in the context of the SmartTraffic project [1], 73% of the people in Asuncion (Paraguay) use their own vehicle as a their main transportation method [1]. Some possible reasons are related to the deficiencies of the public transport system that include the low frequency, the transport routes that are not always close, the travel times, and the lack of comfort in the trips. In this situation, which is typical in developing countries, riding in each vehicle there is only the driver alone. Considering the volume of the vehicle and the unused space inside, it would be fair to say that the transit space used by such vehicle is used sub-optimally. Taking this into consideration, along with the different city indicators (e.g. population projections, traffic capacity of the streets and their scalability) and the causes leading to the current situation, it becomes clear that main streets during peak hours are becoming increasingly saturated. It also becomes evident that the sub-optimal use of resources leads to losses of both time and money, in addition to the increased damage to the environment caused by CO2 emissions [2].

As a potential solution to the previous problems, this work proposes the design of a carpooling [2] system. In particular, a system called UCarpooling is proposed, with the objective of improving the effective use of personal vehicles for transporting people. This paper will focus on the design and validation of a back-end

[1] Research project co-financed by CONACYT-Paraguay within the PROCIENCIA program - http://www.smarttraffic.com.py/

[2] Action of transporting two or more people in a same vehicle.

that may be used to implement solutions for different types of client application (both in the mobile and web environments). A key socially relevant aspect is that UCarpooling is specifically oriented to closed communities (i.e., persons that frequent the same work or study location). We expect that this contributes to several factors that facilitates finding and carrying out these shared trips; among them the increased familiarity (and thus reliability/security) that users that attend to the same institution have, specially when comparing with two strangers without this link. The main task that UCarpooling aims to solve is the matching of persons that are likely to make a shared trip to a common destination (work or study location), while harmonizing other differing factors like the starting point of the trip the target arrival time and other personal traits of the involved participants. It is worth considering, that the problem of a perfect carpooling match is a challenge that the computer sciences have yet to solve. As such a wide variety of proposals focused in different aspects of the problem are currently under discussion in the literature. Furthermore, the feasibility and potential impact of the application-based solution are analyzed through a simulated use of the system. This simulation is based on the application of the system among the students of the Asuncion's University Campus of the Universidad Católica "Nuestra Señora de la Asunción". It considers the different interactions that happen among potential poolers [3] once matched and analyzes the system's potential impact when comparing it with the current (system-less) situation to thus arrive to obtain the reduction in vehicles from students that attend to the campus and the saved total distances from saved from these trips.

This work is presented as an extended version of [3] and, as such, it contains more details about the implementation, validation, and findings and discussion with respect to its conference version. Specifically, in the proposal section we added two new sub-sections focusing on the functional specifications and technical considerations. Indeed, we included the modular architecture of UCarpooling to better understand not just the overall model of functioning of the system but also the specific constituent dimensions. The analysis and simulation section was also enriched with new sub-sections on the data sources which constitutes one of the important challenges for this type of application. In particular, both the survey data and those obtained by means of an application that automatically captures data from sensors and other devices on the mobile phone, play a fundamental role for the performance of the matching module. Moreover, more details are presented on the simulation process. Accordingly, we enrich the results discussion. Finally, we expanded the threats to validity discussion.

The rest of the paper is organized as follow: section 2 presents related work; in section 3 the functionalities and architecture of UCarpooling are presented, with special emphasis on the matching component; in section 4 the data acquisition, the simulation process and the results analysis are discussed; and finally, section 5 presents the discussion and future work.

## 2 Related Work

This section will analyze existing solutions found in the literature, proposing carpooling systems. Table 1 contains a summary of all found works with relevant elements related to the current problematic associated to this work.

From the table it can be noticed that many different approaches to matching people exist and are used for generating carpools. The most currently used are: heuristic formulas, lineal programming, machine learning, pathing similarity. This approaches attempt to provide an answer to the so called car-pooling problem (CPP), which is defined as "*find the subset of passengers that with share the vehicle and the routes that the drivers must follow so the sharing is maximized and the total cost of transport is minimized*" [10]. The CPP presents a significant challenge that is currently not considered as completely solved, because it belongs to a family of algorithms with a non-polynomial complexity (NP-hard). To perform their matching, the articles main take into account two groups of variables: carpool logistics (e.g. start/end of the trip, the time), and personal traits (e.g., is the person a smoker, gender); which evidences the importance of personal traits in the success of the carpool. From the selected literature, it is also possible to notice a tendency towards automatic matching. This means that it is the system (as opposed to the users) is tasked with finding the best match. Finally, Table 1 only contains two articles where security is taken into account, so this aspect was not significantly relevant within the current studies.

Nevertheless, for the correct functioning of a carpooling system, it is essential to have the trust of its users [5]. Users become much harder to engage they feel that their security is compromised by the use of the carpooling application, even if the use of the application demonstrate significant improvement in transportation costs and environmental impact. If the users do not feel safe they will not use the system and this is particularly troublesome in a carpooling application, where a healthy and active user base is needed to avoid falling into the cold start [11] problem.

[3] Members of a given carpool.

Table 1: Current carpooling comparison

| Article | Matching algorithm | Accounts for | Pool selection | Security |
|---|---|---|---|---|
| Hsieh [4] | Heuristic formula | Pathing similarities | Automatic | N/A |
| Bruck et. al [2] | Lineal programming | Origins, destinations and transport method | Automatic | N/A |
| Tsao et. al [5] | Data mining and recommendation algorithms | Interests: smoker, music, politics. Rating similarity | Automatic | Common affiliation. Central credit system |
| Huang et. al [6] | Lineal programming | Origins, destinations and transport method | N/A | N/A |
| Hong et. al [7] | Data mining, pathing clustering | Trip patterns and street preferences | Automatic | N/A |
| Li et. al [8] | Lineal programming | Pathing similarities | Automatic | N/A |
| Giglio et. al [9] | Heuristic formula | Interests: smoker, social network, gender. Ratings similarity | Automatic | Filtering for related friends only. Black listing |

Furthermore, the same act of carpooling is intrinsically linked to the social aspect of people (possibly previously strangers) making a pact to share their ride with others for a determined period of time. As such, factors like the personality of the people participating in the pool may be the difference between a pleasant or stressing ride. For example, if the driver is a smoker, it is likely that the inside of their vehicle has (what others non-smoking persons may consider) an intolerable tobacco smell.

The literature also shows as preferable the applications that free the end users of the burden of manually researching and deciding on the different aspects related to their ride. This automatic matching should furthermore be timely, from the point of view of the user, as taking too long to decide on a match will also hurt the desirability of the application.

Another study of carpooling solutions based on commercial applications that do not (necessarily) have publications attached to them was carried out. Taking as the main sample the applications currently available at the Google Play [4] application distribution platform. Special focus was given to those available and working in Paraguay, the main application context for this study. Many carpooling applications of different levels of complexity were found but most of them do not have a big enough user pool to guarantee a good service (also known as the cold start problem). The applications with a good user base include *Uber* [5] and *Muv* [6], but these two have a model based around monetarily compensating the driver for the service of transporting the rider; which contrasts with our objective towards finding and arranging semi-regular riding pools of people commuting to the same destinations. Another relevant carpooling solution was the web-based *Viadedo* [7], released in 2013 and available until 2017. The platform provided tools for users to post their periodic and casual routes, allowing other users to declare interest in joining them. Nevertheless, aspects related to matching and paying were not addressed and only a service to connect to Facebook to be able to filter among known users was provided to users to help them organize their trips.

Based on the previous study of articles and apps, we define the following as the desirable features for our carpooling system:

- The users must feel secure when using the system.
- The effort required from the user to make/find carpools will be minimized.
- A motivation or incentive for users to actually use the system must be present. This requirement is essential to guarantee an active self-sustaining user base for the system.

[4] `https://play.google.com/`
[5] `https://www.uber.com/py/es/`
[6] `https://muv-app.co`
[7] `https://www.viadedo.com/`

# 3 UCarpooling

This work's main objective is to facilitate the carpooling between colleagues/coworkers or persons that frequent the same institution. This focus on closed community was chosen to improve the perception of security and trust between users as, depending on the shared institution that they frequent, there is an increased likely-hood that the affinity between these people would be higher (at least with respect to making the participants strangers). This becomes a guiding principle for our design of system implementing person matching for carpooling. The design and modeling of such system includes back-end and front-end but, in this work, we are mainly focusing on the back-end component while leaving the front-end component for future work.

## 3.1 Solution description

Taking into account the requirements that we reach in the section 2, the back-end server of such carpooling system is designed to comprise of a single server that will provide RESTful web services for any client (be it mobile or web).

The system named UCarpooling, with the objective of reducing the amount of cars and traffic in the streets, allows to the members of a given institution to look for carpooling groups for the pendular migration towards or returning from the institution. Restricting the use of the platform between people of a closed community like this is aimed to boost the security of the carpool itself and also to facilitate the matching since either the origin or the destination for the trips (as well as the times) are assumed to be the same in this particular situation.

UCarpooling is in charge of recommending potential carpooling partners for their pendular migration related to the institution. As such, it will look of a driver for the users that do not have access to (or want to use less) a vehicle and riders (or other drivers) for users in possession of a vehicle. The matching offered by the system aims to reduce the friction related to interacting with and finding people that have compatible schedules and origin/destinations.

With the objective of matching the most compatible persons in a carpooling, after analyzing the strategies and methods used in related works, we determined that the key variables that we would consider when a person A is interested of generating carpool with person B are:

1. If B is the driver, the distance to the point where A needs to be picked up from the normal trajectory B would take.
2. The distance that both A and B will share until the destination.
3. If B smokes or not.
4. The extroversion level of B.
5. Musical preferences (alternatively, whether the radio will be off).
6. The gender of B.

Out of these variables we are able to identify two main groups: the carpool *logistics*, points 1 and 2; and the *personal traits*, points 3 to 6. It is worth mentioning that other personal traits (e.g., political ideology, sport club or religion) could also be considered but they were deemed less relevant to the act of sharing a ride and were as such excluded to avoid an overly complex profiling and matching of users.

From the point of view of this work we leave open the key motivation/incentives that a person may have for participating in a carpooling (e.g., cost saving, environmental worries, altruism, participation/collaboration in institution-led programs) [12]. As such the negotiation and details on how a carpool is remunerated are left to the users or alternatively to the common institution that they both belong to.

### *3.1.1 Interactions flow*

To better understand how the different persons involved in UCarpooling interact, this section will detail the flow of involved interactions between them. This interactions flow specifies the sequence of actions that the UCarpooling users must follow, from the registration to the system to the successful execution of a carpool.

To register, the persons first need to specify to which institution they belong to. Next they must fill up a profile with some personal information (photograph, name, gender, musical preferences, extroversion level and whether the person is a smoker or not).

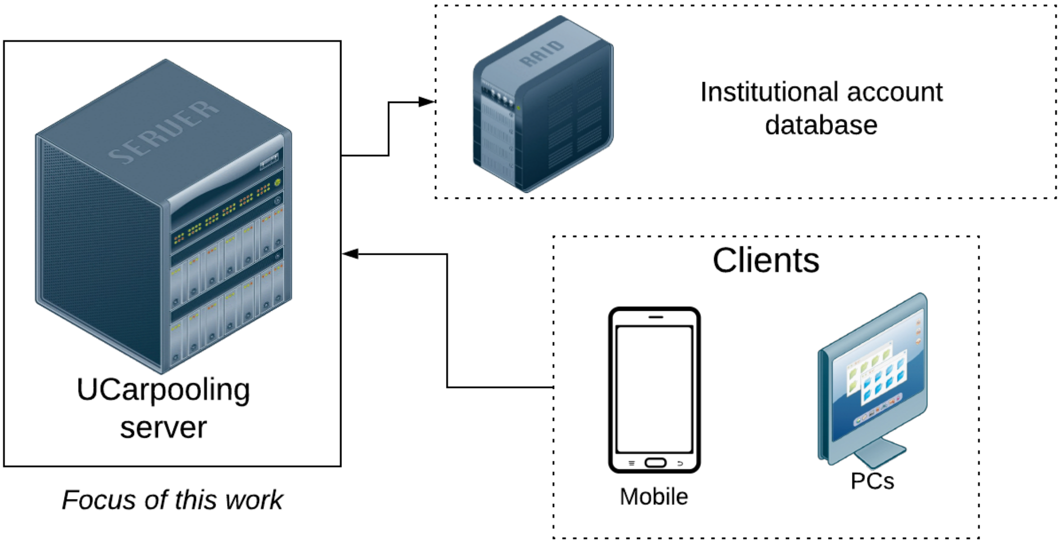


Figure 1: UCarpooling architecture

Once this data is provided, the UCarpooling system checks that the registering user in fact belongs to an institution using UCarpooling and may integrate the registration with information provided by the institution to facilitate the generation of carpools.

Registered users may use the system in two modalities: as a driver, specifying time and place for departure/arrival and the available spots in their vehicle; or alternatively as a rider, specifying only the time and place for departure/arrival.

The persons in rider modality may only be matched with persons that have vehicles. Nevertheless, the persons in driver modality may be matched with both persons with or without vehicles (thus allowing for the possibility that a driver may no use their vehicle, choosing instead to ride as a passenger).

The matching between users for carpooling are guided by the objective of matching the most compatible institution members; while taking into consideration the details of their trips, along with their personal traits. Users will be notified about the current state of their matching through the use of push notifications.

Once a match is suggested by the system, users are tasked of contacting the other users through the system to arrange the carpool; and likewise, contacted users are free to accept/confirm those invitations. This means that for the system to consider that a carpool has been created, all involved users need to confirm and accept it.

The UCarpooling system also provides ratings and reviews features for the users in both the driver and rider modalities. This way, the reputation of a user is part of the information in the system that may be consulted by other users in future interactions.

It is important to note that utmost care is given to protecting the user's personal data at all times. With a reduced level of information being released initially (username, rating/reviews and meeting place) and the rest of the relevant information being revealed only if both sides have consented to the carpool.

### 3.2 Architecture

UCarpooling's architecture, presented in Figure 1, is proposed following a classic client/server organization and it aims to support any client platform (be it mobile or web). Going back to Figure 1, the first step forward (which is covered in this work) is shown surrounded with a solid line; while future work is surrounded by a dotted line.

As the architecture figure shows, the UCapooling server connects to both the *Institutional account database* and the *client devices*. This connection to the institutional data-source is important for UCarpooling because this verifies the participants identity and that they actually belong to the target institution. On the other hand, the client devices will be the ones that will interact directly with the final user. Communications between client and server are done through conventional REST APIs that use the JSON format.

The carpooling server stores the system's global vision in which regards the pool of users that have access to the services offered by the platform, along with the group conformation of the different travel groups. Furthermore, the server is in charge of the calculations needed to match users and communicates these

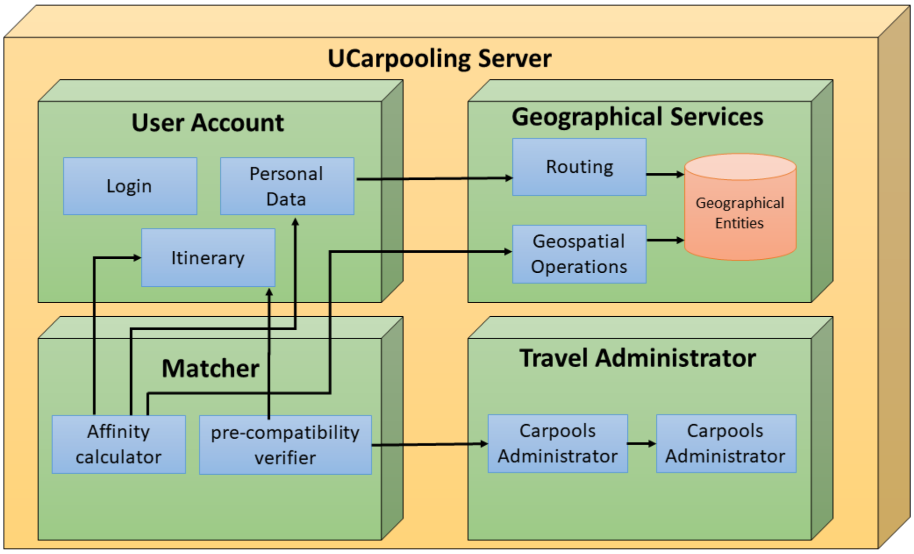


Figure 2: Ucarpooling modules diagram

matching to the end users through the client devices.

### 3.3 Functional Specifications

This section describes the different functionalities covered by the UCarpooling design, grouped in modules as shown in Figure 2. The matcher module is explained in section 3.4 due to its paramount role in this study. The other modules are detailed next.

The *user account* module keeps track of users and personal data required by the system. This module allows users to register, identify themselves and obtain the necessary authorizations to use the applications implemented in the platform. With the exception of sign up and log in, which are started in anonymous mode, all other requests require users to be properly authenticated. It is compose of the login, personal data and itinerary sub-modules.

The *login* sub-module is in charge of verifying the credentials of users each time they enter their username and password. This module is also in charge of keeping the session open once the user is logged in. When creating an account, the user must provide his/her e-mail and password to verify that he/she belongs to the institution. The option of using mail as a credential is configurable since it is available to the institution's account storage.

The *personal data* sub-module is in charge of handling the data and personal features in order to model the users' profile. The personal data that helps to identify people are: institution, name, last name, e-mail, profile photo. Since the matcher (section 3.4) uses personal features as variables, UCarpooling handles the following for each user: type of cigarette (regular or electronic) the person smokes, eloquence level (shy, outgoing or in the middle of the two), musical preferences, and gender. In case they are willing to be drivers, the user must complete with data that identifies their vehicle such as license plate number, vehicle model, color and number of seats available. This module is also in charge of handling the ratings and reviews that users perform on each other.

The *itinerary* sub-module records logistical data related to pendular migrations such as the time of entry or departure, where you are leaving from, and whether you will be driving in the case of a driver. Additionally, in the case of people who have a vehicle for that itinerary, the route is obtained using the routing module.

The *geographical services* are composed of the database, routing and geospatial operations.

The *database* contains geographical entities that represent the network of streets around the institution. Specifically, it contains the network of streets in the form of a graph, where the streets are the edges and the nodes are the intersections of the streets. Additionally, it contains some cost variable to be used when calculating the routing. In our case we use the street distances.

The *routing* has the function of finding the sequence of geopoints (latitude and longitude) given an origin and a destination using Dijkstra's shortest route first algorithm [13].

Table 2: Summary of variables in the affinity calculator

| Group | Id | Variable | Reason |
|---|---|---|---|
| Logistics | 1 | Distance from the passenger's origin to the driver's path | From the point of view of the passenger, the closer the meeting point to their origin the better [12] |
| | 2 | Shared trip distance | A carpool is considered to save more resources the longer the shared trip is [12] |
| Personal traits | 3 | Smoking taste difference | The smell of tobacco during the trip may be unpleasant for non-smokers [9] |
| | 4 | Levels of extroversion difference | The trip is more enjoyable to all parties when the involved persons have compatible extroversion levels |
| | 5 | Musical taste differences | The trip is more enjoyable to all parties when the involved persons have compatible musical tastes [9] |
| | 6 | Gender differences | Personal preferences, security perception [12] |

The *geospatial operations* module implements spatial operations on geometric entities. The operations that this module implements are: distance between geopoints and finding the minimum distance between a point and a trajectory.

The *travel administrator* module is in charge of handling everything related to the carpools group. It includes CRUD operations of carpools and the handling of requests between users. This module is composed of the request and carpool administrators.

The *request administrator* manages the requests that are made. For example, when a user A wants to make a carpool with another B, A makes a formal request through the system. The system notifies B of A's intention and awaits a positive or negative response. If B's response is positive, A is notified about his new carpool partner.

The *carpools administrator* keeps all instances of travel groups updated, adding members as both parties accept requests from other users.

The following section will explain in more detail the *matcher* module.

### 3.4 Matcher

The *matching* module is the one in charge of obtaining the recommendation list of possible carpool partners for each user. The matching is done according to each user's itinerary, allowing each user to have different recommendation lists depending on the day and time for its regular trips. In order for a user to obtain a list of pairings for any trip, an itinerary must first be created on the server specifying their destination, origin, what time he/she is to arrive, and whether he/she has a vehicle for such a trip.

To obtain the affinity percent between two potential carpooling partners, the matcher uses an heuristic formula that takes into account the most predominant variables that persons use when deciding to accept or reject a carpool. A heuristic formula was chosen because (often) they are conceptually simpler and (almost always) computationally cheaper than optimal algorithms [9]. Table 2 summarizes the variable list that the matcher takes into consideration, along with the justification of the choice of each of those variables. Additionally, the variables are assigned a numeric id that will be used to reference them for the rest of this work. The variables from Table 2 are categorized in two main groups: logistics for the carpool (id 1-2) and personal traits of the persons participating in it (id 3-6).

The heuristic formula takes two compatible itineraries as input and returns the compatibility percentage that exists between the two itineraries. The compatibility calculation between an itinerary $I_X$ and another itinerary $I_Y$ that belong to the persons $X$ and $Y$ respectively is obtained by using Formula 1, where $i$ represents the variable's identifier from Table 2, $w_i$ represent how important is the factor $i$ when deciding to accept or reject a carpool, and $f_i(m_{iX}, m_{iY})$ is the calculating function from the variable $i$ regarding the aspects to consider from $X$ and $Y$. It is worth mentioning that the relation of the matching formula is symmetric, that is the level of affinity between $I_X$ and $I_Y$ is the same as the one from $I_Y$ with $I_X$. The value range from Formula 1 is the set of real numbers $\mathbb{R}$ from 0 to 1.

$$AffinityPercentage(I_X, I_Y) = \sum_{i=1}^{6} w_i * f_i(m_{iX}, m_{iY}) \tag{1}$$

To obtain each $w_i$ averaged answers from an online survey from the institution's members (in this case the students from the campus Santa Librada of the Universidad Católica "Nuestra Señora de la Asunción" or UC for short), data presented in 4.1, were used. This online survey asked its participants to categorize how

Table 3: Weight for the variables considered in Table 2

| Variable | Percent |
|---|---|
| Distance from origin to trip | 12.75% |
| Shared trip distance | 72.25% |
| Smoking taste difference | 5.38% |
| Levels of extroversion difference | 3.98% |
| Musical taste differences | 3.3% |
| Gender differences | 2.36% |
| Total | 100% |

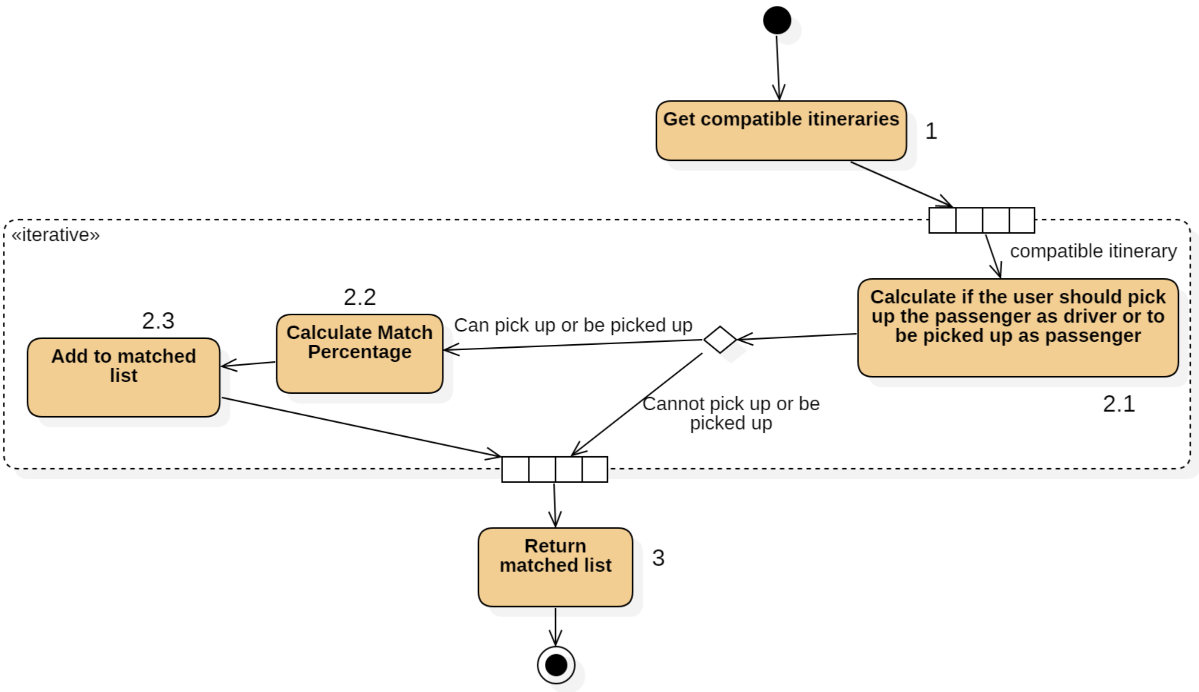


Figure 3: Activity diagram for the UCarpooling matcher

important each of the variables in Table 2 would be for them during carpooling by using a 1 to 10 rating (with 1 being "I don't care" and 10 being "It is very important to me"). Table 3 shows the result of the application of this procedure with the students from the UC. For each variable $i$ it shows the percentage that measures how important is this affinity-related aspect between two persons that want to make a carpool together. It is worth noting that each person may have different motives and priorities when choosing to participate in carpooling, and thus the variables $w_i$ are configurable in the matcher. This leaves the possibility open in future works of using more accurate variables and personalized values for each of the users as opposed to averaged values as we present here. Furthermore it is thinkable to include this preference measuring survey into the clients for UCarpooling through the use of client delivered questions that feed a personal profile encoding this information.

#### *3.4.1 Process for obtaining the matching list*

Figure 3 presents an activity diagram summarizing the process used for obtaining the matching list and enumerating the different steps that the matcher follows.

The diagram assumes that a user $Z$ with a $I_Z$ itinerary wants to obtain their matching list for the pendular migration represented by $I_Z$. The matcher obtains this matching lists by going through the following steps:

1. Filtering of the itineraries that are compatible with the time restrictions from itinerary $I_Z$. A configurable tolerance interval is set at 10 minutes (assuming this is a time that most people is able wait for), so this filtering will still keep in itineraries that share the same time considering this tolerance interval.

2. Calculation of the compatibility percent for each itinerary that could effectively become a carpooling.

This is done with all the compatible itineraries up to this point in the following way:

(a) Check first if $Z$ can be picked up by any drivers and, in the case $Z$ also has a vehicle, it is also checked whether they can pick up other riders. In any of the two previous cases, the algorithm computes the minimal distance from the passenger's origin to the driver's path with a tolerance of 400 meters. Similar to the previously explained time tolerance, this 400 meters tolerance is a configurable parameter aimed to represent how much people are willing to go out their way in order to form the carpool. If $Z$ cannot be picked up and cannot pick up, then this potentially compatible itinerary is discarded and the next one is considered.

(b) If $Z$ can be picked up and/or can pick up, then the algorithm calculates the compatibility percentage with respect to the variables we previously categorized as personal traits or logistics in Table 2 by using Formula 1.

(c) Finally the compatible itinerary, along with its compatibility percentage and $Z$'s role in the carpool (passenger or driver), is put in a list.

3. When all potentially compatible itineraries are iterated, $Z$ is informed about their matching list for the given trip $I_Z$.

### 3.5 Technical Considerations

The client-server communication is done through secure channels provided by the HTTPS protocol. With the exception of new user registration requests and login, which by their nature are initiated in anonymous mode, all other requests require users to be properly authenticated.

The server implements the application programming interface (API) using Python 3 and the Django [8] framework (2.2) for requests and responses. Django is chosen as the back-end framework because it offers web-based module management interfaces. The modules that implement web interfaces are the following:

1. User and profile administration. This interface allows the administrator to invite or block users and start password reset processes.

2. Events and indicators management. This interface allows the administrator to define different types of events that will be sent by the different applications and the indicators with which the analysis of these events is made.

3. Report management. Provides functions for creating, updating and viewing metrics and reports. The report visualization uses Grafana [9].

The database of geographic entities uses sources provided by OpenStreetMap [10] because it is a free and complete database for Ucarpooling requirements. The database engine will be PostgreSQL 10 [11] with PostGIS extensions [12] to implement routing and geospatial operations. Postgres 10 is chosen for its ability to integrate extensions and PostGIS is the necessary for UCarpooling. There are other routing engines such as the API Directions of Google Maps [13] but it has a disadvantage to PostGis since the Google API is paid. Since only the UCarpooling back-end is implemented, it was implemented in such a way that it can later be extended as future work, either by adding new features to the system or by implementing the text front-end. The source code for the UCarpooling system is available in GitHub [14]. The repository has an installation guide and documented code. The access link is public and can be found at `https://github.com/SmartTrafficPY/smartcity-asuncion`.

## 4 Analysis and simulation

In this section we give the details about a feasibility study carried out to test the impact that UCarpooling would have if deployed in a given institution. To do so the impact of implementing this solution in the institution will be analyzed by comparing different indicators when using or not using the proposed system. Special attention is given to quantifying the reduction in quantity of vehicles used and the reduction of total traveled distance thanks to the potential carpools that would be formed by using UCarpooling.

[8] `https://www.djangoproject.com/`
[9] `https://grafana.com/`
[10] `https://www.openstreetmap.org/about`
[11] `https://www.postgresql.org/`
[12] `https://postgis.net/`
[13] `https://developers.google.com/maps/documentation/directions`
[14] `https://github.com/`

This information will be obtained by performing a simulation of the use of the UCarpooling after being successfully adopted within an institution and thus it already has an established user base. In this work the UC (Universidad Católica “Nuestra Señora de la Asunción” - campus Santa Librada) is chosen because of the ease of access to institutional data and further availability of its members for participating in data collection activities that will help make this simulation much more accurate.

To obtain all the data required to run the UCarpooling matcher (logistic details about the pendular migration to and from the institution, personal traits of the participants, etc.) we have considered two main data sources: the institutional data sources, and data directly collected from the potential users through activities with voluntary participants.

### 4.1 Data

The generation and preparation of a significant dataset is essential in order to conduct a simulation that actually informs of the key aspects and potential results related to carpooling. More specifically according to the objectives of this work, the focus is in regular pendular migration of a population of commuters that belong to the same institution. Furthermore, modeling aspects related to personality of the riders/drivers will also be important for matching.

To comply with these requirements (and obtain the variables we need based on Table 2) three source datasets were obtained and prepared, as elaborated in the following subsections.

#### *4.1.1 Institutional account database*

For the UC the authority that guards the data of all students from the institution is the DGI (Dirección General de Informática). The DGI provided the authors of this study with anonymous data from all 6.146 students registered at the UC and that attend classes at the target campus (Santa Librada). From this dataset we were able to obtain (all other information was discarded for privacy related reasons):

1. Gender: M or F
2. Career: name of the career the student (useful for knowing the times in which the student needs to go to the university)
3. Home address: a string containing the home address (useful for knowing the origin of potential trips to the university)

#### *4.1.2 Online Survey*

To complement the information provided by the DGI, willing students participated in an online survey where information about sharing trips was asked. The survey was distributed through institutional social networks and lists, including a virtual classroom application [15]. The survey included the following questions:

1. What university campus do you attend to?
2. What is your current university career?
3. On average, at what time do you need to be at the campus?
4. What transportation method do you use for arriving to the UC?
5. Do you smoke?
6. How extroverted do you consider yourself? (Introverted, Extroverted or neutral)
7. What music genres do you listen to on your trip to the UC?

The survey was answered by 648 students belonging to our target campus, which represents about the 10.54% of the 6,146 registered students. Questions 2 to 4 from the survey are used to complement the missing information from the logistics related needed data, and questions 5 to 7 are used for the personal trait related data. Finally, the trust values the answers was deemed acceptable with a 95% trust level and a confidence interval of 3.64.

[15] `https://aulavirtual.uc.edu.py/`

#### *4.1.3 App-based data collection*

A third data collection activity was also performed for complementing the previous two data sources. The focus here was obtaining detailed and real information from people and their periodic trips to the UC. This data collection was based around a tool named *i-Log* [14] that is used to collect GPS, other sensors and ask questions through the user's mobile phones. More specifically the i-Log app offers the following functionalities:

- Automatic data collection: i-Log is designed to collect data from multiple mobile phone sensors simultaneously. This includes both hardware (e.g., GPS, accelerometer, giroscope) and software (e.g., applications running, wifi networks). i-Log also offers a back-end infrastructure that manages the tasks of sincronization and storage of the data streams generated in each of the user's mobile devices.
- User questions: i-Log is also able to send direct questions to the user's mobile devices. Time-based questions that are periodically sent to the users are called timediaries and these can be used to ask about what the users are currently doing at that moment.
- Registration Survey: a short set of information collection questions that are asked to the user only once when registering to a new data collection campaign.

The three main features of the i-Log app were configured towards capturing data and, eventually, generating useful datasets for the simulation of UCarpooling. The data collection activity using i-Log had a duration of three weeks over which 54 students that do regular trips to and from the UC downloaded the app and actively participated. Upon downloading the app, the users first filled the registration survey, which was configured to ask for the following information:

1. Extroverted
2. Musical tastes
3. Smoking preferences

After registration, users were asked to have the app open in the background of their phones while traveling to and from the UC. The automatic data collection from the app was configured to capture the GPS information of the participants in this situation with a frequency of 3 times per minute. This allowed not only to capture the concrete trajectory that each participant took from origin to destination but also the relative speed and fast/slow parts of the journey. Finally the user questions where configured to send questions at the start and end of the trip asking users questions about their trips (time of departure, time of arrival, means of transportation, people in vehicle, etc).

#### *4.1.4 Data integration and preparation*

After all the three mentioned data collection activities concluded, this generated in turn 3 datasets with different granularities and a variety of (sometimes overlapping) pieces of information. To obtain all the target information necessary for the simulation (as shown in Table 2), different data preparation and integration operations were performed.

As the starting point, the data source with the most users (i.e., the institutional account database), was taken as base and the following operations were performed:

- Each of the 6,146 users was assigned distinct ID that would be used (instead of their name) for the rest of the simulation.
- Registers that had information that was incomplete or with errors were cleaned from the dataset.
- The home addresses for each user were converted into geolocated latitude and longitude corresponding to that address by using the Geocoding API from Google [16]; the registers that produced inconsistent results were cleaned from the database.
- The departure trip for each user was filled based on the career they were registered to and a schedule with all the classes (provided by the UC). As such, we assumed that students in the same career would have similar departure times.

[16] `https://developers.google.com/maps/documentation/geocoding/intro`

After this process, we were left with 3928 users with most of the logistic related information in place (origin and time of the trip) and gender as the only personal trait in the dataset.

The next step was to merge this dataset with the one from the second datasource (i.e., the online survey) in order to enrich our dataset with information related to the personal traits of each of the users. Similarly to what was done before, an ID was generated for each of the users of this second dataset (matching names with the ones of the first dataset so persons with the same name get assigned the same ID). This allowed us to match IDs between the two datasets and append all of the personal trait information from the survey to our current dataset. However, as mentioned previously, the survey only covered about 10% of all the registered users so it was assumed that the rest of the student population had a similar distribution. This means that, supported by the high trust values of the survey, if 80% of the survey participants do not smoke then it would be a fair assumption to make that 80% of the registered students also do not smoke. Applying this procedure we are able to estimate all the data we need (trip origin, time of arrival, transport method, gender and smoking/extrovertedness/music preferences) for all of the student population in our combined dataset.

Finally, a similar procedure was applied for integrating the data from the third data source (app-based data collection). Given that this information contains actual real data of the paths and itinerary of the 54 participating studies, this dataset was used to inform the pathing simulation of all the formed carpools and be able to calculate with better accuracy the distances between drivers and passengers.

### 4.2 Simulation

The main objective of the performed analysis is to quantify the carpools that are made thanks to the presence of the UCarpooling system. Since UCarpooling is a carpooling recommendation system, we also need to account for the behaviour of people when accepting/rejecting their travel partners. For this reason we will also simulate the acceptance/rejection rates between people consolidating a carpool trip.

In fact, the UCarpooling system is based around only notifying users about the possible poolers that they could travel with but leaving the actual choice of consenting the creation of a carpool to the users. As such the users have the final choice when sending an invitation to a pooler and that pooler in turn has the final decision of accepting or rejecting that invitation. It may happen that many times a given recommended carpool is not made because of coordination or negotiation related reasons [15]. So to account for these factors in our simulation we define the *consent probability* variable. This variable represents the existing probability that a person actually does a carpool with other person that was recommended in their matching list. Without this variable a key aspect of carpooling, the complex interactions and dealings between the different users of the system [15], would not be accounted for in the simulation. Formula 2 shows how the consent probability is obtained by taking into account the compatibility percentage between two persons and their general inclination towards carpooling.

$$Cons.\ prob.(A, B) = comp.\ per.(A, B) * carp.\ incl.(A) \tag{2}$$

A simulation with values for the carpooling inclination (carp. incl.) informed from the results of the online survey that the UC students filled (as explained in subsection 4.1). According to the different categories of transportation methods they use, the students were asked about their general inclination towards carpooling. Students that have their own vehicle responded positively towards the idea of carpooling in 90.75% of the cases; meanwhile, this number is 95.15% for students without personal vehicles. Finally, it is worth noting that we will take drivers as the starting point for the matching with passengers in the simulation of whether a carpool is actually agreed or not. As previously mentioned, the matcher operates on a symmetric function where the compatibility percent between two persons $A$ and $B$ is the same as the compatibility percent between $B$ and $A$. Thus, the converse approach would also work (i.e., matching passengers to drivers) but it will not be further expanded in this work.

#### *4.2.1 Simulated carpool generation*

The activity diagram from Figure 4 explains the simulation process and each of its steps.

The simulator obtains first the set of all the people with vehicles to then simulate their interactions with others to form carpools. By assuming a value for the consent probability given $P_{conc}$ and a driver user $U_v$, the simulator obtains the formed carpools by execution the following steps:

1. Obtain a list of the matchings for $U_v$. This list will be ordered in a decreasing manner based on the compatibility percentage. This prioritizes the simulation of “negotiations” between those with the highest compatibility percentage.

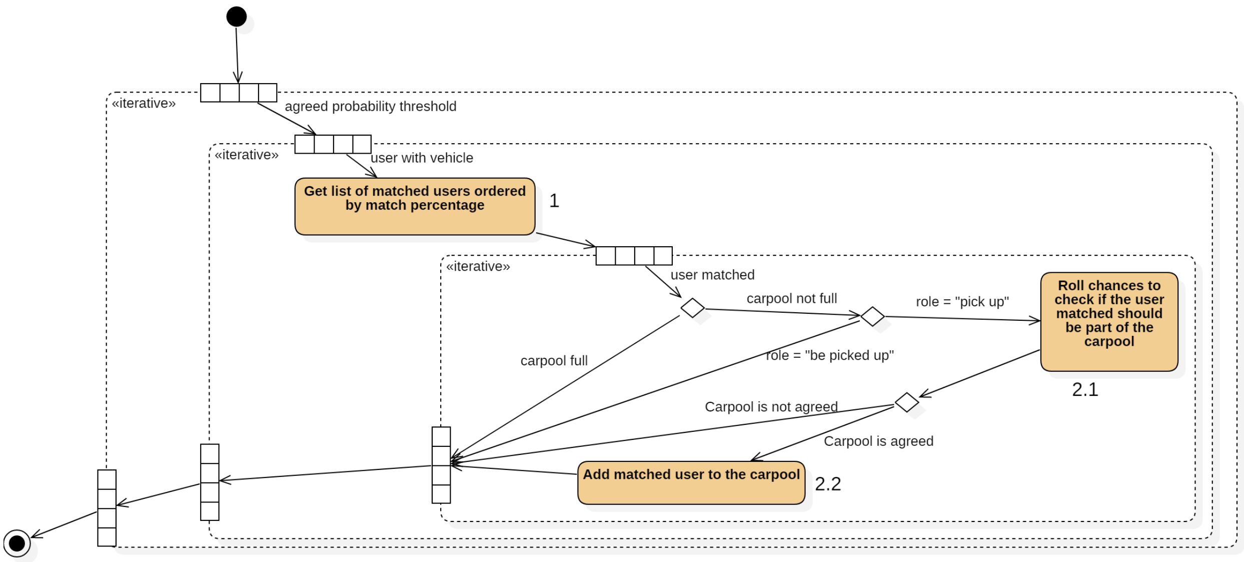


Figure 4: Activity diagram for the simulator

2. If the carpool that $U_v$ drives for is not full and it is compatible with the current passenger then the chances for a successful carpool formation through negotiations are rolled (i.e., similarly to how dice are rolled). Depending on this roll, the negotiation is considered successful or failed. This is done by considering:

   (a) If there is free space in the carpool and if $U_v$ is looking to be matched with passengers then Formula 2 is executed to obtain the real probability that the carpool materializes. The result from Formula 2 is then rolled to obtain the result of the negotiation between $U_v$ and its resulting match. If the roll fails, it means for the simulator that the negotiation failed and that the processes should continue with the next matched user.

   (b) On the other hand, if the negotiation is successful then a seat in $U_v$'s car for that matched user is saved and its assumed that they have successfully became carpooling partners.

### 4.3 Threats to Validity

Considering Wohlin et al. [16] threats to validity definitions, some aspects that may have attempted against the validity of this first evaluation and how they were mitigated, are discussed below.

- Internal validity: Several informed assumptions are made as part of this work to keep the simulation useful but also easy to manage and run. For example, the negotiations that occur between two users to form a carpool together are normally much more complex as different people may have different priorities when choosing a carpooling partner [15]. Some people have an altruistic motivation to help others but others have a monetary motivation to share expenses. However, in our simulation, this negotiation is modeled considering that people have the same predisposition to carpooling depending on their mode of transportation. In this study we have tried to approximate this value based on the results collected, but a certain difference with its real value is to be expected. It is left for future work to use a more sophisticated simulation such as the one proposed by Hussain et al., which is based on agents [15].

  Another assumption that we chose to make is that the home address data that we got from UC's DGI (as seen in section 4.1) are the ones being used as the origin address for the pendular migration towards the university. Nevertheless, students may go to the university from work or from another places that are not necessarily their home address. However, we do not anticipate that these types of differences, when used in large data sets, will significantly impact the results.

  Finally, the geolocation of home addresses to latitude and longitude based points are subject to the precision offered by the Google Geocoding API [17]. For example, some home addresses were encoded as the same point in the map (this is specially true for addresses that had the same street name but had different home number) according the Geocoding API. It was also the case that many addresses were

[17] https://developers.google.com/maps/documentation/geocoding/

Table 4: Simulation result table with varying carpooling willingness scenarios

| Scenarios | Variable | Result | Observation |
|---|---|---|---|
| Scenario 1: Carpooling willingness 0% | Vehicle volume | 2,444 vehicles | 0% reduction |
| | Traveled distance | 17,957 kilometers | 0% reduction |
| Scenario 2: Carpooling willingness 25% | Vehicle volume | 1,578 vehicles | 35.43% reduction |
| | Traveled distance | 13,054 kilometers | 27.3% reduction |
| Scenario 3: Carpooling willingness 50% | Vehicle volume | 1,356 vehicles | 44.52% reduction |
| | Traveled distance | 11,507 kilometers | 35.91% reduction |
| Scenario 4: Carpooling willingness 75% | Vehicle volume | 1,251 vehicles | 48.81% reduction |
| | Traveled distance | 10,733 kilometers | 40.23% reduction |
| Scenario 5: Carpooling willingness 100% | Vehicle volume | 1,169 vehicles | 52.17% reduction |
| | Traveled distance | 10,102 kilometers | 43.74% reduction |

not in the intersection format between 2 streets, but simply the name of the street. This was mitigated with the data collected through the i-Log app (since this app captured the actual exact trajectory that the user took from home to the university) but this information was not available for all users. In general this issue introduces only a moderate level of inaccuracy to the simulation because the general distances to cover remain more or less unaffected regardless.

- External validity: Considering that: 1) the DGI provided anonymous data from all 6,146 students registered at the UC and that attend classes at the target campus (Santa Librada); 2) the survey was answered by 648 students belonging to the our target campus, which represents about the 10.54% of the 6,146 registered students; and 3) a third data collection activity was also performed for collecting pathing and trip-related data from willing student participants; we believe that we have work with enough number of participants and data in order to validate this experience.

In general this work assumes that no significant issues that radically change the end result of the simulation have been introduced with these assumptions.

### 4.4 Result Analysis

By running the simulation results corresponding to the presence and use of the UCarpooling system have been obtained. Data for a second scenario was also produced, namely one where no UCarpooling system exists and no carpools were formed. For each of the scenarios the following were produced:

- Number of vehicles that reached the UC campus
- Total distance traveled by the vehicles that reached the UC campus

To better think about the application of the UCarpooling system in a variety of different situations, we introduce here the variable *carpooling willingness*. This variable captures the general willingness of the population towards carpooling and in a realistic scenario would be determined by the overall incentives and motivations that the target population has for participating in carpools. Table 4 shows in more detail the results for some proposed scenarios.

Scenario 1 from Table 4 represents a situation where the population is not willing to participate in carpools no matter how good is the resulting matching that the system provides (and as such, it is equivalent to the case where no system is deployed). Conversely, Scenario 5 shows a situation where the users accept all proposed carpools. According to our previously reported survey report, we have gauged this willingness to be around 95% for students of the UC. Nevertheless, even when this willingness is minimal (25% as seen in Scenario 2), considerable savings can be obtained in both the number of vehicles and travel distance.

By analyzing the presented results, it may be noted that the use of UCarpooling system causes a significant impact in the reduction of vehicles and in the combined distance these travel. In particular, a reduction close to halving these values may be noted for both indicators. Within this context, specially for institutions with a high volume of vehicles like the UC, this impact would translate in a significant reduction of the traffic around that area during peak hours. Besides of the saved time and money, it would also have a positive impact in the environment thanks to the reduction of $CO2$ emissions from vehicles. Although the gains may look relatively modest on the short term, the benefits of the continuous use of the UCarpooling system would accumulate into significant gains for the students throughout the semester and year.

## 5 Discussion and future work

In this work we introduced UCarpooling, a system for matching of people going regularly to the same institution/organization. Restricting this matching to people with the same routine and the same destination boosts the security and trust between users because a higher level of affinity normally exists between colleagues attending the same institution as opposed to strangers. UCarpooling obtains the level of compatibility that exists between two persons by using logistical data (origin, destination, transportation method). Furthermore, as a social component is very important for carpooling, UCarpooling also uses the personal traits of their participants (gender, musical tastes, extrovertedness, smoking preferences) to propose more compatible matchings.

The feasibility of proposed approach and its potential impact has been validated by the implementation of the back-end of the system and the simulation of its use and adoption by using several collected data sources provided by the university and its students. With the positive impact of the system verified, the next steps ahead are related mainly towards its implementation and adoption. This requires:

- A user study to better define which are the main factors that influence the matching for a carpool (e.g., relevance of political/sport affiliations, friends in common).
- Client application (web and/or mobile) implementation that will use the developed back-end for the implementation of a carpooling system for the end-user.
- Extend the use and adoption of the apps for its use in multiple institutions and zones of high traffic volume. This may lead to a more generalized city-wide impact and would also help to obtain more information to improve the algorithms and clients.

As such and given the results presented and analyzed in section 4, it is possible to conclude that the simulation points towards a positive result in the set objectives of reducing traffic and environmental impact. Furthermore, as special care to minimize and control the factors that would make this simulation different from a real application, it is possible conclude with good confidence that the real application/adoption of the UCarpooling system would also be positive with regards to the identified objectives. This is specially true in a city like Asunción where currently no similar system is available. As such, its possible to claim that if institutions (specially in those with a high volume of pendulum migrations) adopt UCarpooling this would produce a significant change in the traffic conditions in the city.

The development of the back-end and the simulation to validate its use, justify the continuation of the development of the UCarpooling system architecture (shown in Figure 1). The next big step forward would be the development of a client or clients to interact with the back-end. This work has already setup the framework for that step by defining some of the requirements that those clients would need to consider (e.g., always giving the choice to accept/reject to the users, keeping incentives and remunerations outside of the system, the involved information and data sources). One of such clients has already been designed by a group of students and its currently in prototyping stages.

Another interesting insight emerging from this work is that the diversity existing between riders and drivers is a key aspect that involves the matching and formation of carpools. In this sense, besides the saving of resources, future studies may focus on measuring and managing the diversity in order to lead to more comfortable/enjoyable rides and even potential friendships and collaborations that endure after the ride's end.

Future surveys could focus in identifying the key personal traits or diversity aspects that the system needs to model and account for the creation of successful carpools and could also be used to obtain feedback on current configurations and incentives schemes applied for the system.

Beyond the personal diversity in a given population, the diversity existing across different groups of persons and cultures is also an interesting future research path. Applying the UCarpooling system in different groups of people (i.e., different types of institutions, different countries and different cultures) would reveal a great variety of diverse configurations for the system; where some particular details become more or less important depending of the place where it was applied. And, in turn, by analyzing common and divergent patterns between those it would be possible to draw more general conclusions for the further improvement of the system and the achievement of its goals of reducing the costs and negative consequences of pendular migration.

Future work could also expand the data analysis carried out in this study. Improvements are certainly possible for collecting, preparing, combining and ultimately using data for simulating the adoption of the back-end. This would be further enhanced by the arrival of the datasets generated from the actual use of the system on a real population. Many of the tools that were used for the simulation, like surveys or the i-Log application would be key during this process to aid in quick identification of the most suitable configurations for the system and thus maximize the chance for a continued and successful operation.

Another option would be to integrate i-Log to generate highly granular trip data about each carpool and allow to further optimization of the itinerary/trips and other social aspects that are characteristic of carpools.

Finally, a work in progress is related to the validation of the matching algorithm with experiences that are being planned in other closed communities, located in another city (Villarrica). With this experience, we seek to identify strengths and weaknesses of the current proposal. The development of the front-end is also being carried out, as well as an experience to analyze the User Experience (UX) of the UCArpooling app based on pilot tests with university students.

## Acknowledgment

This work was co-financed by CONACYT (Consejo Nacional de Ciencia y Tecnología) with resources from the FEEI, within the framework of the project "SmartTraffic: sistemas colectivos adaptativos para una ciudad inteligente” (PINV15-166). Furthermore, this work has also received support from the Horizon 2020 FET Proactive project “WeNet - The Internet of us”, agreement nº 823783.

## References

[1] P. Fauvety, M. Fatecha, A. Lugo, J. Olivera, N. Aquino, R. Chenú-Abente, J. Paniagua, and J. Saldívar, “Percepción ciudadana acerca de los problemas de movilidad de Asunción,” Universidad Católica “Nuestra Señora de la Asunción”, Tech. Rep., June 2018, https://drive.google.com/file/d/1le_-Tt-N-RaizqCKOR3m6LyOPfDdLc0v/view.

[2] B. P. Bruck, V. Incerti, M. Iori, and M. Vignoli, “Minimizing $CO_2$ emissions in a practical daily carpooling problem,” *Comput. Oper. Res.*, vol. 81, pp. 40–50, 2017. [Online]. Available: https://doi.org/10.1016/j.cor.2016.12.003

[3] A. Lugo, N. Aquino, M. González, L. Cernuzzi, and R. Chenu-Abente, “UCarpooling: Decongesting Traffic through Carpooling using Automatic Pairings,” in *XLVI Latin American Computing Conference, CLEI 2020, Loja, Ecuador, October 19-23, 2020.* IEEE, 2020, pp. 358–366. [Online]. Available: https://doi.org/10.1109/CLEI52000.2020.00048

[4] F. Hsieh, “Car Pooling Based on Trajectories of Drivers and Requirements of Passengers,” in *31st IEEE International Conference on Advanced Information Networking and Applications, AINA 2017, Taipei, Taiwan, March 27-29, 2017*, L. Barolli, M. Takizawa, T. Enokido, H. Hsu, and C. Lin, Eds. IEEE Computer Society, 2017, pp. 972–978. [Online]. Available: https://doi.org/10.1109/AINA.2017.41

[5] H.-S. J. Tsao and M. Eirinaki, “A human-centered credit-banking system for convenient, fair and secure carpooling among members of an association,” *Procedia Manufacturing*, vol. 3, pp. 3599 – 3606, 2015, 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2351978915007301

[6] C. Huang, D. Zhang, Y.-W. Si, and S. C. H. Leung, “Tabu search for the real-world carpooling problem,” *Journal of Combinatorial Optimization*, vol. 32, no. 2, pp. 492–512, Aug 2016. [Online]. Available: https://doi.org/10.1007/s10878-016-0015-y

[7] Z. Hong, Y. Chen, H. S. Mahmassani, and S. Xu, “Commuter ride-sharing using topology-based vehicle trajectory clustering: Methodology, application and impact evaluation,” *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 573 – 590, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X17302930

[8] R. Li, Z. Liu, and R. Zhang, “Studying the benefits of carpooling in an urban area using automatic vehicle identification data,” *Transportation Research Part C: Emerging Technologies*, vol. 93, pp. 367 – 380, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X18308787

[9] C. Giglio and R. Palmieri, “An ICT solution for shared mobility in universities,” in *Proceedings of the Second International Afro-European Conference for Industrial Advancement, AECIA 2015, Villejuif (Paris-sud), France, 9-11 September 2015*, ser. Advances in Intelligent Systems and Computing, A. Abraham, K. Wegrzyn-Wolska, A. E. Hassanien, V. Snásel, and A. M. Alimi, Eds., vol. 427. Springer, 2015, pp. 205–215. [Online]. Available: https://doi.org/10.1007/978-3-319-29504-6_21

[10] H. Abdel-Latif, K. Elaraby, A. Alsobky, and A. Mohamed, "Car-pooling attractiveness modeling in greater cairo organizations – a case study," in *Intelligent Transport Systems – From Research and Development to the Market Uptake*, T. Kováčiková, Ľ. Buzna, G. Pourhashem, G. Lugano, Y. Cornet, and N. Lugano, Eds. Cham: Springer International Publishing, 2018, pp. 21–30. [Online]. Available: https://doi.org/10.1007/978-3-319-93710-6_3

[11] N. Silva, D. Carvalho, A. C. Pereira, F. Mourão, and L. Rocha, "The pure cold-start problem: A deep study about how to conquer first-time users in recommendations domains," *Information Systems*, vol. 80, pp. 1 – 12, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306437918303260

[12] F. Bachmann, A. Hanimann, J. Artho, and K. Jonas, "What drives people to carpool? explaining carpooling intention from the perspectives of carpooling passengers and drivers," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 59, pp. 260–268, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1369847817306459

[13] D. B. Johnson, "A note on dijkstra's shortest path algorithm," *J. ACM*, vol. 20, no. 3, p. 385–388, Jul. 1973. [Online]. Available: https://doi.org/10.1145/321765.321768

[14] M. Zeni, "Bridging sensor data streams and human knowledge," Ph.D. dissertation, Università degli Studi di Trento, 11 2017. [Online]. Available: http://eprints-phd.biblio.unitn.it/2724/1/Thesis.pdf

[15] I. Hussain, L. Knapen, S. Galland, A. Yasar, T. Bellemans, D. Janssens, and G. Wets, "Agent-based simulation model for long-term carpooling: Effect of activity planning constraints," in *Proceedings of the 6th International Conference on Ambient Systems, Networks and Technologies (ANT 2015), the 5th International Conference on Sustainable Energy Information Technology (SEIT-2015), London, UK, June 2-5, 2015*, ser. Procedia Computer Science, E. M. Shakshuki, Ed., vol. 52. Elsevier, 2015, pp. 412–419. [Online]. Available: https://doi.org/10.1016/j.procs.2015.05.006

[16] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, and B. Regnell, *Experimentation in Software Engineering.* Springer, 2012. [Online]. Available: https://doi.org/10.1007/978-3-642-29044-2