POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Control-enabled policy design for fostering sustainable and just mobility habits

Tesi di Laurea Magistrale in
Automation and Control Engineering - Ingegneria dell'Automazione

Author: **Oriana Guagliardi**

Student ID: 968871
Advisor: Prof. Mara Tanelli
Co-advisors: Valentina Breschi, Eugenia Villa
Academic Year: 2021-22

Abstract

Starting from the issue of the ecological transition in the field of mobility and the challenge of widespread adoption of environmentally friendly transportation options, the thesis aims at developing optimal schemes for the design of fair mobility policies. The work here presented firstly analyzes the importance of proposing personalized incentives to people to promote the use of green technologies among them by accounting for individual socio-economic characterization. After that, it is discussed the relevance to account for social justice, for a fair and inclusive transition to green mobility solutions.

The thesis proposes a data-driven network-based human-centered approach to model the adoption of Car-sharing services, referencing a European Commission survey on transport and mobility. The data of this survey is used to both characterize in a simple way each person's propensity towards shared mobility and build a multi-agent network used in this work to study the diffusion of sharing mobility services.

Moreover, we exploit a Linear Quadratic Regulator to obtain control schemes for the design of optimal policies and try to innovatively introduce the concept of fairness for social justice directly within the design process, to accomplish not only an ecological, but also a just transition.

Keywords: Green Mobility, Technology Adoption Models, Opinion Dynamics Control, Policy Design, Ecological Transition, Just Transition, Fairness, LQR

Abstract in lingua italiana

Partendo dal tema della transizione ecologica nel campo della mobilità e dalla sfida della diffusione dell'adozione di opzioni di trasporto ecologiche, la tesi mira a sviluppare schemi ottimali per la progettazione di politiche di mobilità eque. Dopo aver analizzato l'importanza di proporre incentivi alle persone per promuovere l'uso di tecnologie verdi, viene discussa la necessità di tenere conto della giustizia sociale, per una transizione equa e inclusiva verso soluzioni di mobilità sostenibile.

Facendo riferimento a uno studio portato avanti della Commissione Europea sui trasporti e la mobilità, la tesi, dunque, propone un approccio data-driven basato sull'utilizzo di reti e human-centered per riuscire a modellare l'adozione dei servizi di Car-sharing. I dati di questa indagine sono utilizzati sia per caratterizzare in modo semplice ed efficace la propensione di ogni persona verso la mobilità condivisa e sia per costruire la rete multiagente che verrà usata in questo lavoro.

Inoltre, facendo riferimento a un regolatore lineare quadratico, la tesi propone schemi di controllo per la progettazione di politiche ottimali e cerca di introdurre in modo innovativo il concetto di equità per la giustizia sociale nel processo di progettazione, per realizzare una transizione che non sia solo ecologica, ma giusta.

Parole chiave: Mobilità sostenibile, Modelli di adozione, Dinamica delle opinioni, Policy Design, Transizione Ecologica, Transizione Giusta, Giustizia sociale, LQR

Contents

1 | Introduction

When talking about new technologies, especially those related to the ecological and environmental transition, large-scale adoption is a fundamental aspect. Especially concerning the current climate emergency, even though new technologies may be valuable to the cause, they result ineffective if not exploited by a large number of users.

The acceptance of new technologies into the everyday lives of consumers is heavily influenced by the traits of the people who will be asked to embrace them. The motivations that push an individual to start using a certain type of innovation, as well as the reasons for a user to be reluctant to do so, can be manifold.

An important aspect of environmental transition involves the mobility sector being indeed responsible for nearly a quarter of Europe's greenhouse gas emissions [8]. In this context, the transition to green mobility can be difficult and slow since it requires overcoming cultural and mindset barriers. Habits in general, and specifically transportation habits, are deeply ingrained and can be difficult to change. Mobility is a very important aspect of a person's life: it enables people to access goods and services, employment, education, and social activities. It impacts different aspects of individuals' conditions, such as finances and time, and the decisions made on the subject are guided by a complex set of factors, including personal preferences, social norms, infrastructure, and economic status. Given the complexity of the acceptance process, public incentives that foster the widespread adoption of innovative and sustainable mobility solutions are crucial. European are indeed putting more of an emphasis on public policies promoting new mobility patterns. Hence, considering the importance of personal traits and tendencies, it is evident that wide and complete knowledge of the individual characteristic affecting their adoption propensity must be taken into consideration to design successful and effective government policies. Taking the example of the strategies considered to foster the adoption of electric vehicles, although they are often similar they have different effects based on the receiving population so that the diffusion of electric or hybrid vehicles is not the same across Europe [12]. Such an outcome can be explained by saying that the final effects of policy programmes are heavily dependent on the socio-economic context in which they are implemented. In general, for what concerns green mobility, the impact of actual socio-economic status has been the subject of numerous research [2, 25]. Various variables that slow down the widespread adoption of environmentally friendly transportation options have been identified. The current demographic and economic factors are among the key determinants that influence individual mobility preferences and personal needs. Hence, it is crucial to address these barriers so that sustainable transportation solutions become appealing to all members of society. In addition, considering that people are not static beings, a dynamic approach is important as well.

Moreover, since new technologies and their subsequent introduction strongly impact individuals' lives and thus have social consequences, it is really important that these fostering policies are fair and equitable for all individuals. If a vision of justice is not taken into consideration, opinion dynamics models can include biases, i.e. tendencies or inclinations towards a particular point of view or belief based on prior knowledge, and the results can reflect the preferences of a specific group or individual. If one wanted to mitigate the unfair factors, the purpose is to integrate a concept of justice into the design process. In order to do so, one should first define fairness, a non-trivial task: no universal definition of fairness exists as it depends heavily on the context taken under consideration [18, 27]. Making reference to philosophical studies, fairness is perceived and valued differently across cultures: what is considered fair in one culture may not be perceived as fair in another [14]. Generally, fairness is the absence of any prejudice based on a person's inborn or learned traits that are irrelevant in the specific context of decision-making [24]. That being said, once a formalization of this concept is obtained, instead of using it in retrospect to analyze the outcomes as it is usually done, it is crucial in the designing of policies not only to consider a human-centered dynamic approach, but also a fair one.

1.1. Problem statement

In this work, considering the case study of the adoption of car-sharing, we aim at designing a control policy that fosters the adoption of sharing mobility services accounting for socio-economic factors while promoting social justice. Starting from the data-set resulting from the 2014 *EU survey on issues related to transport and mobility* [11] we analyze it and try to understand which are the main factors driving potential adopters. By extracting the most relevant features we can define a quantitative depiction of the individual tendency towards sharing services, the Sharing-DNA. After that, focusing on a certain region of Europe we build a data-driven network connecting individuals living in that region based on a specific proximity measure. Once the network is characterized, in the first place the goal is to study the network evolution, i.e. how the adoption increases through the network, starting by operating in open-loop and subsequently in closed-loop with the aim of designing an optimal policy. To do so we work with a *deterministic irreversible cascade model* [4] and a *Linear Quadratic Regulator*. Finally, the concept of fairness is added to the formulation of the Linear Quadratic Regulator to aim to find a fair optimal policy. The results will highlight the elements that influence the evolution of the network and how fairness influences it.

1.2. Thesis Contribution

This thesis aims to develop policy schemes for fostering the adoption of Car-sharing through a network. To do so first, the factors that lead to the evolution of networks have to be analyzed. In doing so, it is highlighted that individual characteristics, different approaches used in the design process and different networks affect the outcomes. After analyzing the European dataset regarding the mobility habits of the population, the work presents the tool of the Sharing-DNA, a data-driven tool that represents the personal propensity to use a sharing service. On the basis of the latter, the two networks used in this work are then retrieved. Then, considering already existing opinion dynamic models, in this thesis, we propose different formulations for the closed-loop control problem for the design of optimal policies. After testing the optimal schemes obtained on the two networks the work aims to define the effects that affect the different ways of evolution of the networks. Lastly, the thesis tries to integrate the concept of fairness into the policy design process by innovatively incorporating it directly in the control problem formulation so that a just approach is not only used in retrospect to analyze the situation after services deployment, as it is canonically done in the literature, but also during the very process of policy formation.

1.3. Structure of the Thesis

The thesis is organized as follows.

- **Chapter 2 - State of the art**. In this initial chapter, we will report from the literature the different tools used during this work, namely the Deterministic Irreversible Cascade Model and the Linear Quadratic Regulator. Moreover we will analyze the actual situation of what affects the demand for car-sharing services and we will try to summarize the knowledge of fairness from the literature;
- **Chapter 3 - Dataset and network building**. In this chapter, we will present the

dataset used during this work and the tool of the Sharing-DNA will be discussed. Using this data the two networks will be consequently retrieved and a process of clusterization will be carried out;

- **Chapter 4 - Adoption model**. In this chapter, the different formulations of the control problem will be presented and discussed;
- **Chapter 5 - Analysis and comparison of the results**. This chapter covers the analyses of the results obtained testing the formulations introduced in the previous chapter and the most important aspects of the evolution of the networks will be highlighted;
- **Chapter 6 - Conclusion and future developments**. In this final chapter, we will try to summarize the most important outcomes of the work and highlight the importance of future work concerning the use of the concept of fairness in the process of the control design of policies.

2 | State of the art

In this chapter, the theoretical background of the fundamental tools and concepts on which this work is based is going to be presented. Firstly, an overview of the literature on the adoption and attitudes towards car-sharing services is presented. Secondly, an introduction of the fundamental aspects and objectives of the subject area of opinion dynamics is made, focusing on the outline of the deterministic irreversible cascade model upon which this thesis will heavily rely on. Thirdly, the control theory tool of the Linear Quadratic Regulator will be revised, and in particular, we present how, in discrete time availing the feedback of a system, an optimal control action can be obtained. Lastly, starting from the notion of bias we try to briefly outline the concept of fairness and review the most used definitions making reference to important research papers.

2.1. What affects the demand for car-sharing services

In this thesis, we primarily focused on the diffusion of car-sharing services, an important ingredient of sustainable mobility as it provides an alternative to individual car ownership and encourages people to use vehicles more efficiently. Car-sharing programmes have been shown to have advantages like lowering new car sales, reducing the number of cars per household and having positive effects on the personal Carbon footprint of each user [21]. Hence it is of interest to understand how to encourage the adoption of this technology by exploiting human-centred policies. To this extent, we first analyze previous work investigating factors influencing individual willingness to subscribe to sharing mobile services.

In [21] it is presented a discussion about multiple research projects related to the acceptance of car-sharing schemes, especially to the improvement with respect to electric vehicles (EV) penetration and charging infrastructure across Europe, among which ELVITEN. By furnishing 225 light motor electric vehicles in six European cities for short-term and long-term renting, equipped with ICT tools and services, the project identified the main factors inhibiting this type of mobility from entering the market: low awareness of their use, concerns about the high cost and the lack of consistency by the local authorities in building a charging infrastructure essential to the use of these vehicles.

As discussed in [9], another important aspect that emerges is the economic situation of the user. The paper presents and analyzes the results of an online survey conducted in Greece, focusing on data from citizens aged 8 to 35, where car-sharing is non-existent and bike-sharing is just emerging. The study finds that respondents with an annual income between 15K and 25K Euros are more likely to join these programmes. Income is thus a significant determinant, with those in the low-mid income class being more likely to join when talking about the youngest.

Another notable article found in the literature is [22]. This paper analyzes the influencing factors based on multiple aspects: the type of services, geographic area, specific travel demand, family status, gender, scholastic level and occupation. It finds that car-sharing is more attractive in places where the proportion of single-parent households is high and the mean number of cars per family for members of a shared-mobility service is lower than for non-members. Moreover, these services result to be more attractive for potential female members and people with higher education levels, especially post-graduate or advanced degrees.

To get a wider understanding of influence factors, in this work we rely on and analyze the 2014 *EU survey on issues related to transport and mobility* [11], a large-scale survey carried out by the European Commission to learn more about how people in Europe feel about mobility and transportation.

2.2. Opinion dynamics

As stated in *Chapter 1* an important part in this work is played by opinion dynamics.

Opinion dynamics is the field of study that focuses on how beliefs and views evolve and spread over time within a population. As individuals in a social structure, the way of thinking and the propensities over different aspects of lives variate over time [10]. This research area is inherently interdisciplinary and draws from various fields in the social sciences, including social psychology and sociology. To investigate the spreading of opinions in a set of human beings, it utilizes mathematical, physical models and agent-based computational modelling techniques [26].

When talking about opinion dynamics two approaches can be outlined: data-driven and model-driven. In a data-driven framework, the focus is on the acquisition and analysis of large-scale social data to identify patterns and trends in opinion formation. In this case, there is an application of machine learning and statistical methods to extract information from large data-sets that can be of different natures. On the other hand, model-driven research is concerned with developing theoretical models that can capture the complex dynamics of shaping and shifting beliefs. This approach often involves the use of models to make simulations in different contexts and study the impact of various aspects, such as group dynamics and social networks. Although these two strategies might seem different, they are used as a combination and they complement each other providing data on various ranges [10].

Since formalising the complexity of beliefs formation and dynamics implies outlining a collection of attributes that best describe people's ideas in a particular social context and a set of rules according to how people interact with and influence one another, the considered models can be denoted as agent-based models [16]. Each individual or "agent" is depicted as a distinct entity with a distinctive collection of characteristics, including its current state, i.e. its opinion. According to the nature of the opinion state variable, models can be classified as discrete or continuous. In the first case, the state is a finite number of possibilities such as a predetermined response in a survey or a selection of a product; in the second case, it is a real-valued variable to describe the level of agreement-disagreement with respect to a subject of interest [10].

In this work, we will refer to a model-based approach and in particular to the model of the Deterministic Irreversible Cascade.

2.2.1. Deterministic Irreversible Cascade Model

The deterministic irreversible cascade model is an agent-based discrete model that will be widely used throughout this work.

The mathematical framework in which this model operates is the following. Given a time t, N agents are represented in a network outlined by an undirected graph

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}) \tag{2.1}$$

where $\mathcal{V}$ are nodes of the graph that represent the N users and $\mathcal{E}$ are the edges that established the mutual influence among them and so denotes when two nodes are considered neighbours.

Each individual is characterized by a state varable evolving over time. The state variables can assume binary values $x_v(t) \in \{0, 1\}$ that indicates whether at time $t \in \mathbb{N}_0$ the v_{th} agent is an adopter ($x_v(t) = 1$) or not ($x_v(t) = 0$). Since we have only two possible states, we can have a simple rule that describes the way the adoption is spread through the network, the so-called *transition rates* [13]. Specifically, in the case of the irreversible cascade model, the v_{th} user can only evolve from state $x_v = 0$ to $x_v = 1$ therefore, when an agent becomes an adopter remains such until the end of the horizon time. The transition rate of changing state is here considered as the mean of its neighbours at time t in state 1 compared to the number of agents belonging to its neighbourhood $\mathcal{N}_v$: $\frac{1}{|\mathcal{N}_v|}\sum_{w\in\mathcal{N}_v} x_w(t)$.

Wanting to test the model in a finite horizon time T, the model dynamics can now be considered. If we consider at time t=0 the starting point, a set of users with positive state is given

$$S(0) = \{v \in \mathcal{V} : x_v(0) = 1\} \tag{2.2}$$

and each agent $v \in \mathcal{V}$ evolves as

$$x_v(t+1) = \begin{cases} 1, & \text{if } x_v(t) = 1 \text{ or } \frac{1}{|\mathcal{N}_v|}\sum_{w\in\mathcal{N}_v} x_w(t) \geq \alpha_v \\ 0, & \text{otherwise} \end{cases} \tag{2.3}$$

Therefore, given the set of new positive state user at time t, $S(t) = \{v \in \mathcal{V} : x_v(t-1) = 0$ and $x_v(t) = 1\}$, the following holds

$$S(t) = \{v \in \mathcal{V}\ (\cup_{\tau=0}^{t-1} S(\tau)) : \frac{|S^*(t) \cap \mathcal{N}_v|}{|\mathcal{N}_v|} \geq \alpha_v\} \tag{2.4}$$

for t>1, where $S^*(t) := \cup_{\tau=0}^{t} S(\tau)$ indicate the total amount of positive state user until time t.

2.3. Linear Quadratic Regulator

In order to design the optimal policy exploiting the feedback of the closed-loop system, the tool used in this work is the *Linear Quadratic Regulator*.

The *Linear Quadratic Regulator* (LQR) is a control algorithm that combines a linear control law with a quadratic cost function. It is used in control theory to optimize the performance of a given dynamic system and can be used for both linear and nonlinear systems, in continuous and in discrete time. The basic idea of LQR is to compute a feedback control law at each time step t that minimizes a quadratic cost function [7]. The latter is composed of the sum of a quadratic term for the error between the current state of the system $x(t)$ and the desired state $\bar{x}$, and a quadratic term for the control input

$u(t)$. Given an time horizon of T, in continuous time the following holds

$$J = \int_{t=0}^{T-1} \left(Q(t)(x(t) - \bar{x})^2 + Ru(t)^2\right)\, dt + Q(T)(x(T) - \bar{x})^2 \tag{2.5}$$

$$\text{s.t. } \dot{x}(t) = Ax(t) + Bu(t) \tag{2.6}$$

while in discrete time it can be written as

$$J = \sum_{t=0}^{T-1} \left(Q(t)(x(t) - \bar{x})^2 + Ru(t)^2\right)\, dt + Q(T)(x(T) - \bar{x})^2 \tag{2.7}$$

$$\text{s.t. } x(t+1) = Ax(t) + Bu(t) \tag{2.8}$$

where

- $\{Q(t)\}_{t=0}^{T} \geq 0$ are time-varying parameters which penalize the mismatches between the desired and actual state;
- R > 0 is a parameter which penalizes the control action.

Therefore, the values of Q(t) and R can be tuned to change the relative importance of the control objectives. For example, if the values of Q(t) are greater than the value of R by a factor of 10, a more aggressive control action will result. Vice versa, if we want to preserve more of the inputs, the value of R needs to increase with respect to Q(t) [15].

Considering from now on the LQR in discrete-time, the optimal control action, that minimizes the *Cost function 2.7*, can be derived. If the system in *Eq. 2.8* is controllable, then it is possible to design a proportional controller such as

$$u(t) = -K(x(t) - \bar{x})^2 \tag{2.9}$$

First of all, we define the terminal cost as follows:

$$J(T) = P(T)(x(t) - \bar{x})^2 \tag{2.10}$$

with P(T)=Q(T). In addiction, the cost-to-go at time T-1 will be:

$$\begin{aligned} J(T-1) &= \min_{u(T-1)} [(Q(T-1)+P(T))(x(T-1)-\bar{x})^2 + Ru(T-1)^2] \\ &= (Q(T-1)+P(T))(x(T-1)-\bar{x})^2 \\ &+ \min_{u(T-1)} [(R+P(T)B^2)u(T-1)^2 + 2P(T)B(x(T-1)-\bar{x})u(T-1)] \end{aligned} \tag{2.11}$$

By solving the minimization problem of the *Cost function 2.11*, the optimal policy is given by

$$u^*(T-1) = -\frac{P(T)B}{R+P(T)B^2}(x(T-1)-\bar{x}) \tag{2.12}$$

When we replace $u^*(T-1)$ into the *Cost-to-go 2.11*, the equivalent expression follows:

$$J(T-1) = P(T-1)(x(T-1)-\bar{x})^2 \tag{2.13}$$

where P(T-1) can be recursively updated as

$$P(T-1) = Q(T-1) + P(T) - \frac{(P(T)B)^2}{R+P(T)B^2} \tag{2.14}$$

In general, at time $\sqcup$, for all $t \in [0, T]$, we have

$$P(t) = Q(t) + P(t+1) - \frac{(P(t+1)B)^2}{R+P(t+1)B^2} \tag{2.15}$$

$$K^*(t) = -\frac{(P(t+1)B)}{R+P(t+1)B^2} \tag{2.16}$$

and the resulting optimal policy will be

$$u^*(t) = K^*(t)(x(T-1)-\bar{x}) \tag{2.17}$$

In conclusion, the optimal policy will be computed forward in time, given the optimal gain and the current state thanks to the feedback.

2.4. Fairness

In designing policies to boost the adoption of new technology, if a concept of justice is not included in the process, the outcome may be unfair.

To understand how to consider fairness, the source of inequity first needs to be identified.

In [18], talking about machine learning, the cause is detected in biases. It is stated that if one works in a data-driven framework, existing biases in the data might affect the algorithms on which they are based, producing biased outcomes as a consequence. In this way, already existing biases might be amplified and maintained throughout the entire process. In this paper, the different types of biases presented in the data are outlined as follows:

- ***Measurement Bias***. Biases in measurement or reporting result from the way we select, apply, and evaluate specific features;
- ***Omitted Variable Bias***. Omitted variable biases happen when a model contains one or more significant variables but not others;
- ***Representation Bias***. The way we sample from a population when gathering data contributes to representation bias;
- ***Aggregation Bias***. Also known as the ecological fallacy, they occur when incorrect inferences about specific individuals are made after studying the entire community.

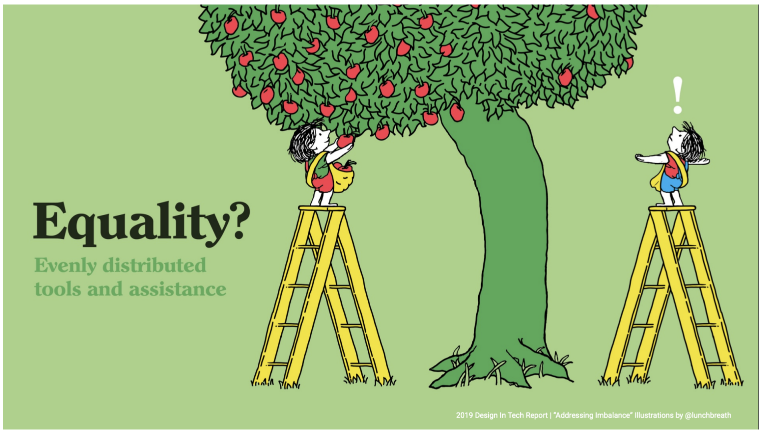
Trying to overcome bias and prejudice has been a subject of study in many different fields. Attempts were made to outline a concept of fairness, and what has been agreed upon in the literature is that no universal definition exists. As asserted in [27], fairness is deeply contextual. As it is not a fixed or predetermined concept, what is considered fair in one situation may not be fair in another. Decisions related to fairness often require judgement calls and the balancing of different interests and values, which can vary depending on the specific circumstances of each case. While the legally mandated principles of non-discrimination are founded on conventionally understood factors including ethnicity, race, language, gender, religion, or political opinions, automatic discrimination is more complex, intuitive, and conceptual as it is based on less distinct categories. The European Union has a well-established general framework of legislation for non-discrimination and the protection and promotion of equality, but this framework can fail in its coverage of potential new grounds in automated decision-making. In [18], some of the most widely used definitions are provided. Within the background of machine learning, considering two classes of members, positive and negative, and protected attributes such as gender, race, or age, fairness can be considered as:

- ***Equalized Odds***. For both members of protected and unprotected groups, the probability of a person in the positive class being successfully assigned a positive outcome and those of a person in the negative class being wrongfully assigned a positive outcome should be equal;

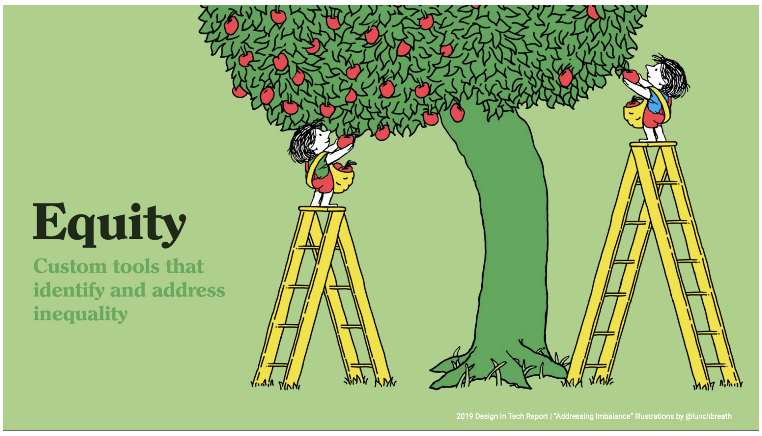
- ***Equal Opportunity***. Both protected and unprotected group members should have an equal chance of being allocated to a positive result if they are in a positive class;
- ***Demographic Parity***. Whether a person belongs to the protected group or not, the chances of a favourable outcome should be the same;
- ***Fairness Through Awareness***. Two individuals who are similar according to a specific similarity metric should receive similar outcomes.
- ***Fairness Through Unawareness***. In the decision-making process, one shouldn't explicitly use any protected attributes.
- ***Treatment Equality***. The ratio of false negatives and false positives should be the same for both protected group categories.
- ***Test Fairness***. The estimated probability that a particular individual belongs to a certain category or class should be such that both protected and unprotected groups have an equal probability of correctly belonging to the positive class.
- ***Counterfactual Fairness***. A decision should be the same in both the actual world and a hypothetical world where the individual belonged to a different demographic group.
- ***Fairness in Relational Domain***. One shouldn't consider the individual attributes of people, and instead take into account the relational structure in a domain, including the social, organizational, and other connections between individuals.
- ***Conditional Statistical Parity***. The distribution of outcomes should be the same for both protected and unprotected groups, even after controlling for legitimate factors that may be relevant to the decision.

In [23] Ralws asserts that justice is achieved when everyone in society has equal basic rights and when social and economic inequalities are arranged so that they benefit the least advantaged members of society. Thus, the principle of justice can be thought of as a fair distribution of goods among people. Within the context of this thesis, the given definitions of fairness can be extended to how resources can be fairly allocated. For example, Equalized Odds can be applied to ensure that both protected and unprotected groups have an equal chance of receiving goods based on their performance or achievements; meanwhile, Demographic Parity and Treatment Equality can be used to ensure that the chances of getting a favourable outcome are the same for all individuals, regardless of their demographic group or protected attribute. In other words, these definitions aim to promote equal treatment of all individuals or groups by ensuring that the outcomes are not affected by protected characteristics such as race, gender, or ethnicity. Hence, the definitions aim to achieve an equal allocation of goods.

Nevertheless, an equality approach is not always adequate to reach justice, and for this reason, this work considers fairness rather than an equal distribution of resources an equitable one. Equality is treating everyone equally, ensuring that identity does not influence opportunities through prevailing attitudes, practises, and regulations. As can be seen in Fig. 2.1, the difference between equity and equality is small but significant: whereas equality presupposes that all persons should be treated equally, equity takes into account a person's particular circumstances and adjusts the treatment accordingly so that the outcome is equal [5].



(a)



(b)

Figure 2.1: Equality vs. equity. From Design in Tech Report 2019 [17]

.

3 | Dataset and network building

In this chapter, the data used in this thesis will be presented firstly, by analyzing the *EU survey on issues related to transport and mobility* [11] and secondly, by describing the information extracted we introduce a data-driven tool to compactly represent individual inclination towards sharing mobility services. We call it, the Sharing-DNA. Lastly, we will describe how we build a social network based on the extracted information.

3.1. EU survey on issues related to transport and mobility

The *EU survey on issues related to transport and mobility* [11] is a survey of 2014. Data are collected to determine the capacity of the transportation policy to meet prescribed goals and consequently to assess changes in the mobility sector. In fact, to identify appropriate parameters to analyze the effects of policies across all EU Member States, plenty of information is needed. For these reasons, the Joint Research Centre – Institute for Perspective and Technological Studies (JRC-IPTS) of the European Commission assigned to TRT Trasporti e Territorio and IPSOS the task to carry out an EU-wide transport survey.

Involving 28 European countries, the primary goal of the survey was to gather facts on the usage of cars, the use of transport options for travelling long distances, as well as other policy-relevant topics.

The questionnaire was proposed in each country to a sample of 1000 individuals, with the exception of Cyprus, Luxembourg and Malta which presented a sample of 500. The specimens were segmented in line with the socio-economic characteristics of each area. The weighting procedure was applied so that they reflect the composition of the EU adult population (from the age of 16) in terms of gender, age class, employment status and living region.

The survey was divided into four sections with respect to the following four main subjects:

- ***General information***: the availability of vehicles and public transportation services, as well as demographic information about the responder (such as age, gender, and place of residence);

- ***Everyday mobility***: mode utilised, frequency of excursions, length, distance, intermodality, and also opinions on the primary issues encountered;
- ***Long distance trips***: longer journeys (between 1000 and 300 km and more than 1000 km) taken within the previous 12 months as well as the number of trips made with their purposes and the primary modes of transportation;
- ***Opinions on European transport policy***: thoughts regarding many facets of European transport policy, including the opportunity of road charging.

Moreover, the questionnaire was maintained the same by means of translations into the local languages of each country. In this way, the responses obtained are fully comparable across all of Europe.

3.1.1. Data description

The dataset utilised in this study is the outcome of the resultant responses to the 39 questions of the survey. In order to provide an overview of the data, Table 3.1 summarizes the available information already been categorised into seven main groups, including biological information (Bio), family status (Fam), geographic information (Geo), education (Edu), profession (Prof), environmental sensitivity (Env), and mobility habits (Mob).

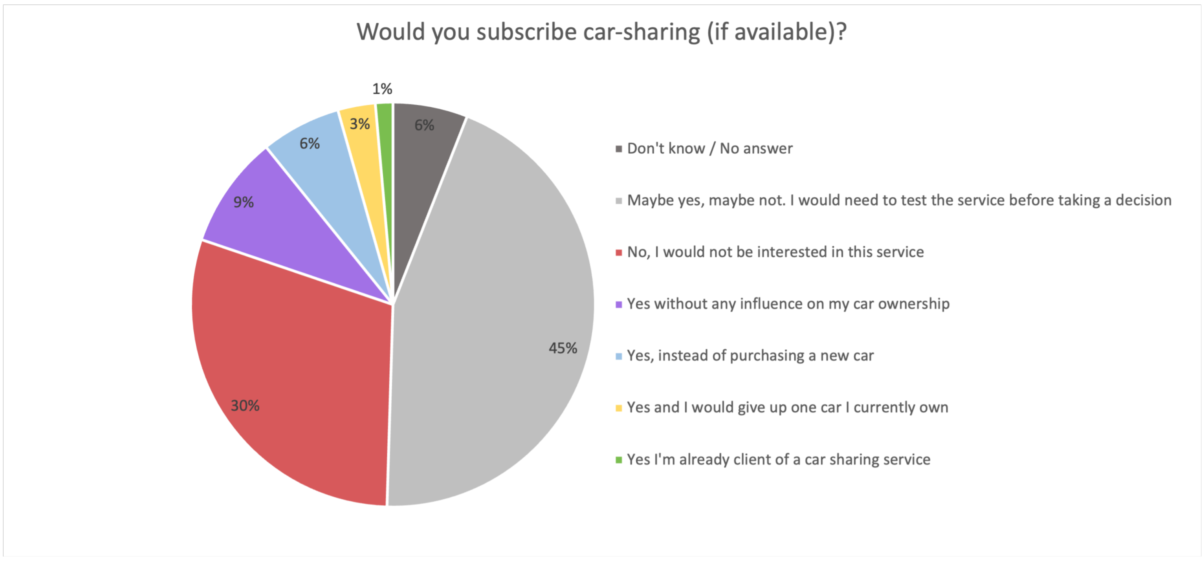
Attribute	Type	Encoding	Category
Country	Cat	label	Geo
Gender	Cat	binary	Bio
Age	Num	ordinal	Bio
Education	Cat	ordinal	Edu
Region	Cat	label	Geo
Profession	Cat	label	Prof
Work status	Cat	label	Prof
Household members	Cat	ordinal	Fam
Income level	Cat	ordinal	Fam
Location of residence	Cat	ordinal	Geo
Centre or suburbs	Cat	binary	Geo
Public transport service	Cat	ordinal	Mob
Car driving license	Cat	binary	Mob
Number vehicles in household	Int	numerical	Fam
Consider buying EV or HEV	Cat	label	Mob
Know what car sharing is	Cat	binary	Mob
Would subscribe car sharing	Cat	binary	Mob
Most freq trip Walk	Cat	binary	Mob
Most freq trip Bicycle	Cat	binary	Mob
Most freq trip Car as Driver	Cat	binary	Mob
Most freq trip Car as Passenger	Cat	binary	Mob
Most freq trip Train	Cat	binary	Mob
Most freq trip Underground	Cat	binary	Mob
Most freq trip Tram	Cat	binary	Mob
Most freq trip Bus	Cat	binary	Mob
Most freq trip Motorcycle	Cat	binary	Mob
Destination most freq trip	Cat	ordinal	Mob
Frequency most freq trip	Cat	ordinal	Mob
Problem Congestion	Cat	binary	Mob
Problem Parking	Cat	binary	Mob
Problem Lack of bike lanes	Cat	binary	Mob
Problem Infrequent pub trans	Cat	binary	Mob
Problem Lack pub trans coverage	Cat	binary	Mob
Problem most freq trip none	Cat	binary	Mob
Transfers between modes	Cat	ordinal	Mob
Freq trip duration [min]	Num	ordinal	Mob
Freq trip distance	Cat	ordinal	Mob
Concern environmental impacts	Cat	ordinal	Env
Preference tolls or traffic	Cat	ordinal	Mob

Table 3.1: Attributes available from the EU survey on issues related to transport and mobility.

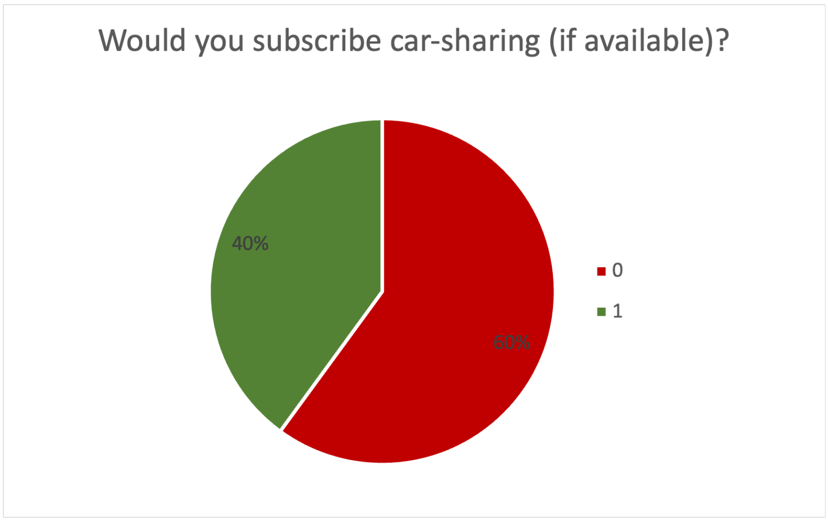
Among all of the 39 queries, it is important to highlight the one that says *Would subscribe car sharing (if available)*. In fact, the responses to this question will be considered during this work as the actual individual inclinations towards sharing-car services. The actual answers that were given to this query are seven different types. They were encoded as shown in Table 3.2 and their distribution is shown in Fig. 3.1

Answer	**Encoding**
Don't know / No answer	-
Maybe yes, maybe not	-
No, I would not be interested in this service	0
Yes, without any influence on my car ownership	1
Yes, instead of purchasing a new car	1
Yes, and I would give up one car I currently own	1
Yes, I'm already a client of a car-sharing service	1

Table 3.2: Encoding the target question.



(a) Distribution of answers to *Would subscribe car-sharing (if available)* before encoding.



(b) Distribution of answers to *Would subscribe car-sharing (if available)* after encoding.

Figure 3.1: Distribution of answers to *Would subscribe car-sharing (if available)* before (3.11a) and after (3.11b) encoding

By evaluating Fig. 3.1a, it is evident that over half of the survey participants were unsure about their opinions on car-sharing. In fact, 45% of the users gave the answer *Maybe yes, maybe not. I would need to test the service before taking a decision* and the 6% the answer *Don't know / No answer*. Given that the main goal of this work is to find the characteristics of people who have a clear attitude towards sharing service, the respondents providing unclear answer to the target question were discarded, reducing in this way the considered sample. Considering all the positive answers encoded to 1 and all the negative ones encoded to 0 (see Table 3.2), in the reduced dataset we have two classes: the positive attitude class and the negative attitude class (see Fig. 3.1b). In addition, some other data were discarded: the data resulting from the questions *Problem Congestion*, *Problem Parking*, *Problem Lack of bike lanes*, *Problem Infrequent pub trans*, *Problem Lack pub trans coverage* and *Problem most freq trip none* as they were given only to individuals who explicitly responded to questions on the method of transportation used for frequent trips and thus empty for many respondents. Furthermore, all the information pertaining to people who didn't clearly respond to the queries *Know what car sharing is* and *Household members* were dropped. In the first case, the reason was to prevent inconsistent outcomes given that it is essential to consider respondents that have a clear knowledge about sharing services. While in the second case, it resulted crucial to have a wide insight of the socio-economic status and so to have a clear picture of the family situation.

3.1.2. Sharing-DNA

The need of a clear and effective representation of an individual's attitude towards sharing mobility services comes from the fact that some features have a greater influence on individual opinion than other features. The tool presented in this section, the Sharing-DNA, fulfills this need, as it aim to be a novel data-driven tool that represent quantitatively and compactly the personal propensity to use a sharing service.

Classifier

The first thing to do is to identify which socio-economic factors are most strongly associated with people's likelihood of using shared mobility services. To this end, considering the target variable encoding (see Table 3.2), machine learning algorithms were used. Firstly, the reduced dataset presented in the previous section were divided in a balanced way divided into two subset: a training (the 75% of the reduced dataset) and a test set (the 25% of the reduced dataset). Considering the training dataset, a supervised classification problem was solved by using the machine learning algorithm of the Gradient Boosting [19] through which the classifier was built. This tool was applied by considering a technique that combines multiple decision trees into a single, more accurate prediction model. Decision trees were added sequentially to the model to progressively correct the errors of the previous trees. The resulting set of weak predictors is transformed with a sigmoid function in order to convert the output of the model into a probability estimate for each of the two target classes.

Features selection

Based on the results obtained, it was possible to analyze the importance of the available features in determining the final classification outcome. This was done using the Permutation Importance algorithm [1]. The letter is a feature selection method that measures the importance of each feature in a machine learning model. It quantifies how much a model's performance decreases when one feature's values are arbitrarily permuted while the values of the other features remain the same. A feature turns out to be significant if it is shuffled and the performance of the model drops. Thanks to this tool it was possible to select a small subset of attributes that more than others influence one's attitude towards sharing services:

- ***Considering buying an EV or HEV (Mob)***: it indicates the propensity of the user to buy an electric or hybrid electric vehicle;
- ***Country (Geo)***: it indicates the individual's country of origin;
- ***Concern environmental impacts (Env)***: it tells us how much the environmental issues worry the user;
- ***Education (Edu)***: it indicates level of education of the individual;
- ***Profession (Prof)***: it indicates the user's employment;
- ***Age (Bio)***: it indicates the age of the individual;
- ***Income level (Fam)***: it refers to the user's level of income.

Sharing-DNA definition

These seven chosen features made it possible to derive the compact, yet complete, representation of the individual attitude towards Car-sharing that we were looking for, the Sharing-DNA. Its extraction was made by determining the relationship between the answer to the question *Would subscribe car sharing (if available)* and the selected characteristics values depending on the responses to the associated questions. The chosen correlation is the likelihood l of being in the positive attitude class and of presenting a specific value of each of the seven features.

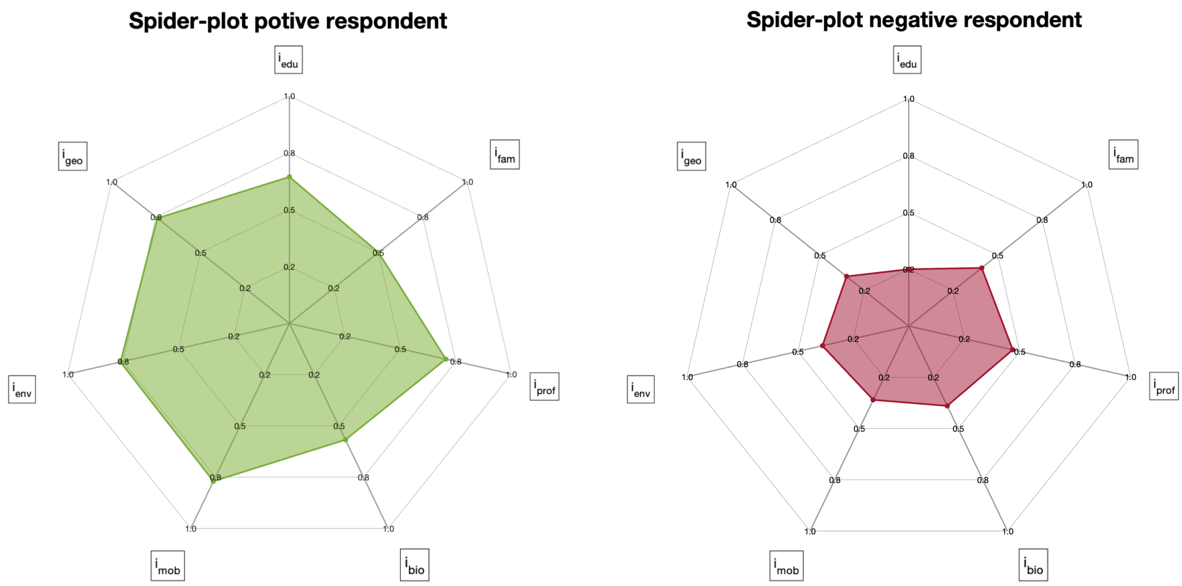
$$l_{i,j} = \frac{\#(P \cap A_{i,j})}{\#A_{i,j}} \tag{3.1}$$

with

- $\#(\cdot)$ is the cardinality of a set;

- P is the set of the respondents belonging to the positive attitude class;
- $A_{i,j}$ is the set of respondents for which the $i-th$ feature takes the $j-th$ value with i=1,...,7 and j = 1,...,n where n depends on the nature of the specific attribute.

In this way, it is possible to construct the Sharing-DNA of an individual by creating the vector of likelihoods based on their responses to specific questions. Hence, the Sharing-DNA will be represented as a vector that contains elements that are dependent on how the individual responds to questions associated with specific features. From now on, these seven likelihoods will be identified with the seven indices of i_{mob}, i_{geo}, i_{env}, i_{edu}, i_{prof}, i_{bio} and i_{fam}, one for each feature. Furthermore, for its nature, its best representation is made with spider-plots, as shown in Fig. 3.2. With the help of this representation, we can quickly see how individuals differ from one another both in terms of each of the seven features and the Sharing index. Particularly, it is important to note that the region within the spider plots is related to the inclination towards using shared mobility options.



(a) (b)

Figure 3.2: Spider plots for two reference individuals: a positive respondent 3.2a and negative respondent 3.2b.

Another important factor that will be widely used throughout this work is the Sharing index (i_{sha}) calculated by computing the Sharing-DNA normalized average as follows:

$$\begin{aligned} i_{norm,n} &= \frac{i_n - \mu_n}{\sigma_n} \\ i_{sha} &= \frac{1}{7}\sum_{n=1}^{7} i_{norm,n} \end{aligned} \tag{3.2}$$

where i_n represents the $n-th$ feature and $i_{norm,n}$ the normalized value of i_n with μ_n in the mean value of the $n-th$ feature and σ_n the standard deviation. Henceforth, this numerical value belonging to [0, 1] will be considered as the individual's inclination towards shared mobility.

3.2. Network analysis

The model that we are going to use in this thesis is the Deterministic Irreversible Cascade Model which is typically used to study the dynamic of opinions diffusion in a network. The letter, as viewed in Section 2.2.1, figures a set $\mathcal{V}$ of N nodes that represent the considered users and a set $\mathcal{E}$ of edges that represent their connection. In the following section, the procedure employed to build this social network is described.

3.2.1. Adjacency matrix

The connections among agents and thus the set $\mathcal{E}$ are defined through the adjacency matrix. Starting for the graph in Eq. 2.1, we consider each node characterized by two properties $\{b_v, r_v\}_{v\in\mathcal{V}}$ where $b_v \in \mathbb{R}^2$ is the position of the user's base and $r_v \in \mathbb{R}^2$ is the user's average daily displacement. The positions of the base are randomly chosen taking into account the individuals' property of living in a big city, in a small city or in a rural area. Consequently, individuals are randomly placed in either the center, suburbs, or rural area of their respective region according to their characteristic. Furthermore, for each node v the circumference $\mathcal{C}_v$ of radius r_v and centre in b_v can be obtained, and the proximity network can be derived as follows:

$$(v, w) \in \mathcal{E} \Leftrightarrow \mathcal{C}_v \cap \mathcal{C}_w \neq \varnothing \tag{3.3}$$

Thereby, two individuals are considered neighbours if the circumferences that describe their average daily displacement intersect, which means that they may come across.

Accordingly, the adjacency matrix $A \in \{0,1\}^{\mathcal{V}\times\mathcal{V}}$ can be obtained and defined as follows:

$$A_{v,w} = \begin{cases} 1, & \text{if } (v,w) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases} \tag{3.4}$$

and the neighbourhood of each agent v_{th} is consequently derived in this way:

$$\mathcal{N}_v = \{w \in \mathcal{V} : (v,w) \in \mathcal{E}\} \tag{3.5}$$

3.2.2. Network construction

As stated in section 2.2.1, each individual is characterized by a state binary variable ($x_v(t) \in \{0,1\}$) that evolves over time. It indicates whether at time $t \in \mathbb{N}_0$ the v_{th} agent is an user of a Car-sharing service ($x_v(t) = 1$) or not ($x_v(t) = 0$). At the starting point, t=0, we have a seeds set of early adopters (see Eq. 2.2) that is here considered composed by the individuals that are not only in the positive attitude class, namely the users that have responded positively to the question *Would subscribe car sharing (if available)*, but specifically those users that are already using a Car-sharing service (see Table 3.3).

v_{th} **user answer**	$x_v(0)$
No, I would not be interested in this service	0
Yes, without any influence on my car ownership	0
Yes, instead of purchasing a new car	0
Yes, and I would give up one car I currently own	0
Yes, I'm already a client of a car-sharing service	1

Table 3.3: Initial state variables.

Considering the evolution described by the Eq. 2.3, the transition rate is here set for each user v as the average of the adopters between the neighbours $\mathcal{N}_v$. On the other hand, the threshold $\alpha_v \in [0,1]$, i.e. the aversion towards the adoption, depends on the individual Sharing index $i_{sha,v}$ in such a way:

$$\alpha_v = 1 - i_{sha,v} \tag{3.6}$$

Therefore, being $i_{sha,v} \in [0,1]$, the thresholds are complementary to the Sharing indices.

Now that we have all of the necessary data, we can build the network. The software chosen in this work is *MATLAB*, which provides simple tools for the task required. The first thing we have to do is to collect all of the retrieved information in a *Excel* file. This file will include for each of the participants of the EU survey the values shown in Table 3.4.

Name	Type	Range	Description
$home_{lon,v}$	Real	$\mathbb{R}$	Longitude of user's base
$home_{lat,v}$	Real	$\mathbb{R}$	Latitude of user's base
$i_{mob,v}$	Real	$[0, 1]$	Index of the *Considering buying an EV or HEV* feature
$i_{geo,v}$	Real	$[0, 1]$	Index of the *Country* feature
$i_{env,v}$	Real	$[0, 1]$	Index of the *Concern environmental impacts* feature
$i_{edu,v}$	Real	$[0, 1]$	Index of the *Education* feature
$i_{prof,v}$	Real	$[0, 1]$	Index of the *Profession* feature
$i_{bio,v}$	Real	$[0, 1]$	Index of the *Age* feature
$i_{fam,v}$	Real	$[0, 1]$	Index of the *Income level* feature
$i_{sha,v}$	Real	$[0, 1]$	Sharing index
$seeds_v$	Binary	0 or 1	State variable

Table 3.4: Description of the nodes' characteristics.

where the seven indices i_{mob}, i_{geo}, i_{env}, i_{edu}, i_{prof}, i_{bio} and i_{fam} constitute the Sharing-DNA.

Lastly, a specific region has to be chosen and a wider number of people have to be extracted from the original sample to have a bigger network and consequently a better representation of the population. Throughout this thesis, we will work with two networks representing the population of two big European cities: Milan and Warsaw. The choice of these specific cities comes from the willingness of comparing the behaviour of two networks with different characteristics that will be presented in the following sectors.

3.2.3. Milan network

The first network taken under consideration is that of Milan. The original sample used is composed of the 158 participants of the survey that come from the region of Lombardy, Italy. From this initial specimen 1005 users were extracted.

Taking under consideration the coordinates of base positions (retrieved as explained in section 3.2.1) and the initial set of adopters, the result is a group of nodes located in a area whose center is in the city of Milan.

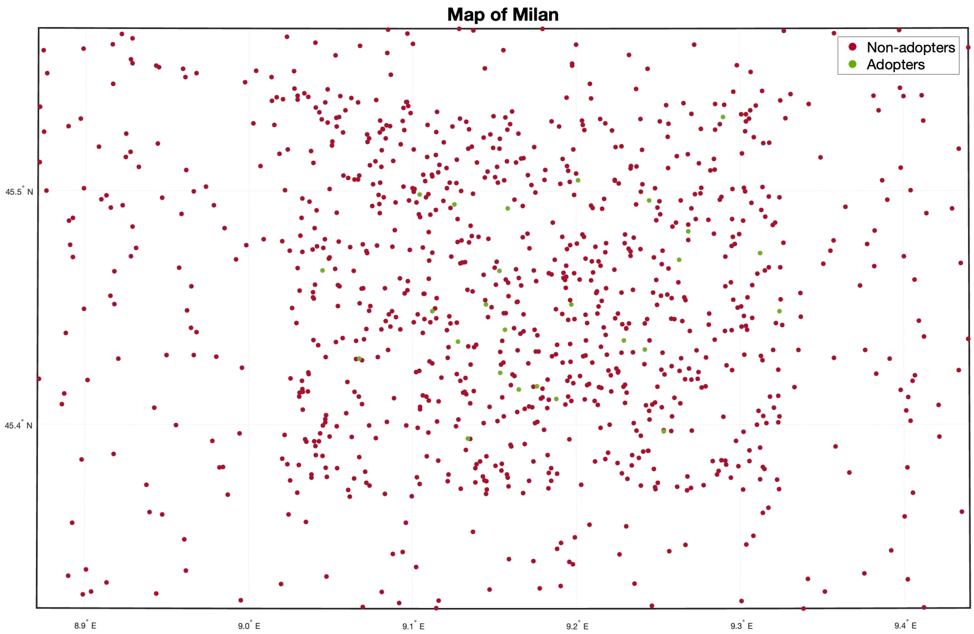


Figure 3.3: Map of Milan.

As we can see in Fig. 3.3, there is a higher density of users in the core of the map, which corresponds to the urban area. In the other hand, in the rural area, which is the extreme part of the network, there are far fewer nodes. Since the position of the base is retrieved according to the characteristic of the individual of living in a big city, in a small city or in a rural area, the node distribution is a data that can be extracted from the EU survey dataset. Moreover, all of the early adopters, the green nodes, live in the center of the city, and they represent 2.59% of the whole population.

Given the adjacency matrix, whose heatmap is been proposed in Fig. 3.4, the outcome is an almost fully connected network. And indeed, as we can see in Fig. 3.5 and especially in Fig. 3.5b, all of the degree of centrality values are bigger than 300 and consequently the largest connected component of a network contains all 1005 nodes.

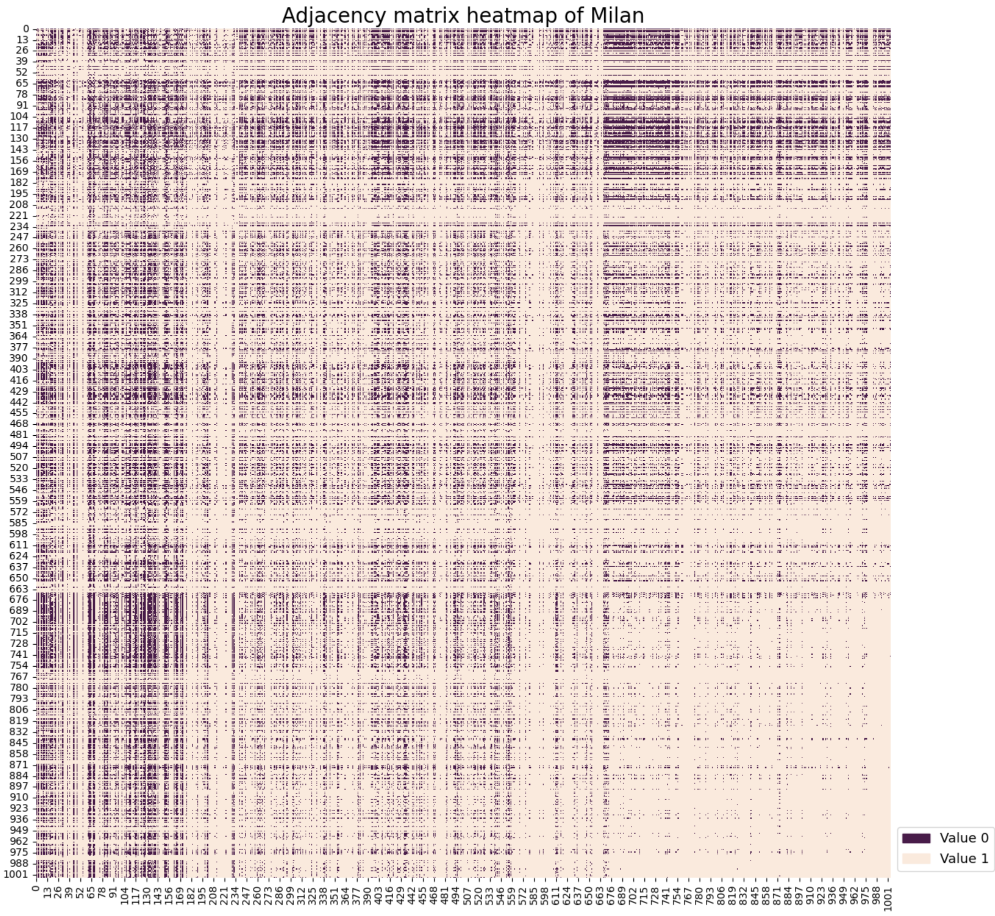
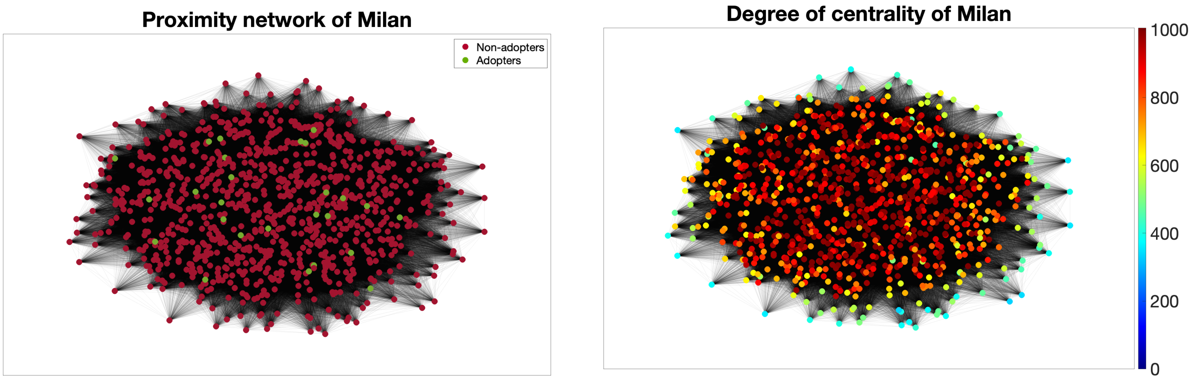


Figure 3.4: Heatmap of the adjacency matrix of Milan network.



(a) Proximity-based network of Milan. (b) Degree of centrality of Milan network nodes.

Figure 3.5: Topology of the Milan network.

3.2.4. Warsaw network

The second network that we will consider in this work is the Warsaw one. It is retrieved from the original specimen of the 148 participants of the survey from the Mazowieckie province, Poland. In this instance, 1002 users were extracted from the initial sample.

Also in this case, from the information on the base positions and the initial seeds, the nodes are situated in an area that has its center in Warsaw's heart.

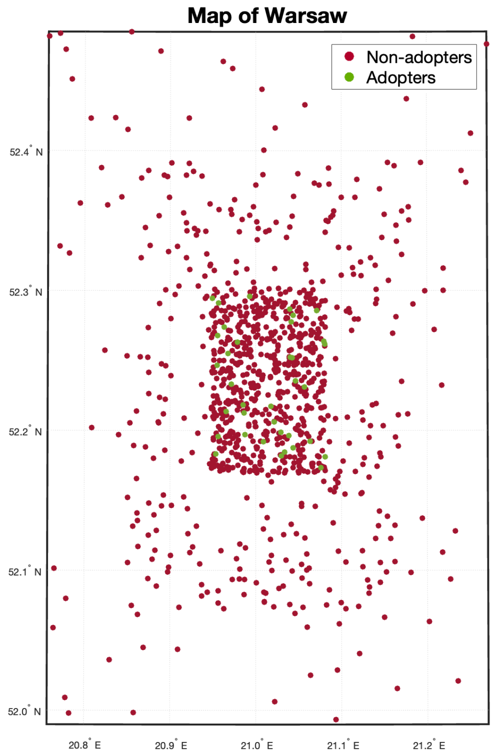


Figure 3.6: Map of Warsaw.

Also with regard to the Warsas network, as can be noted in Fig. 3.6, the majority of the users are located in the center of the map, the urban area, and the rural part is less populated also in this case. Again, the entire set of early adopters is located in the center of the city and here it represents the 3.59% of the whole population, a slightly greater percentage.

Similar to Milan, from the heatmap of Fig. 3.7 and the graphs of Fig. 3.8, it can be noticed that the network is almost fully connected, in fact, all of the nodes have a degree of centrality higher than 300 (see Fig. 3.8b). Hence its largest connected component contains all 1002 users.

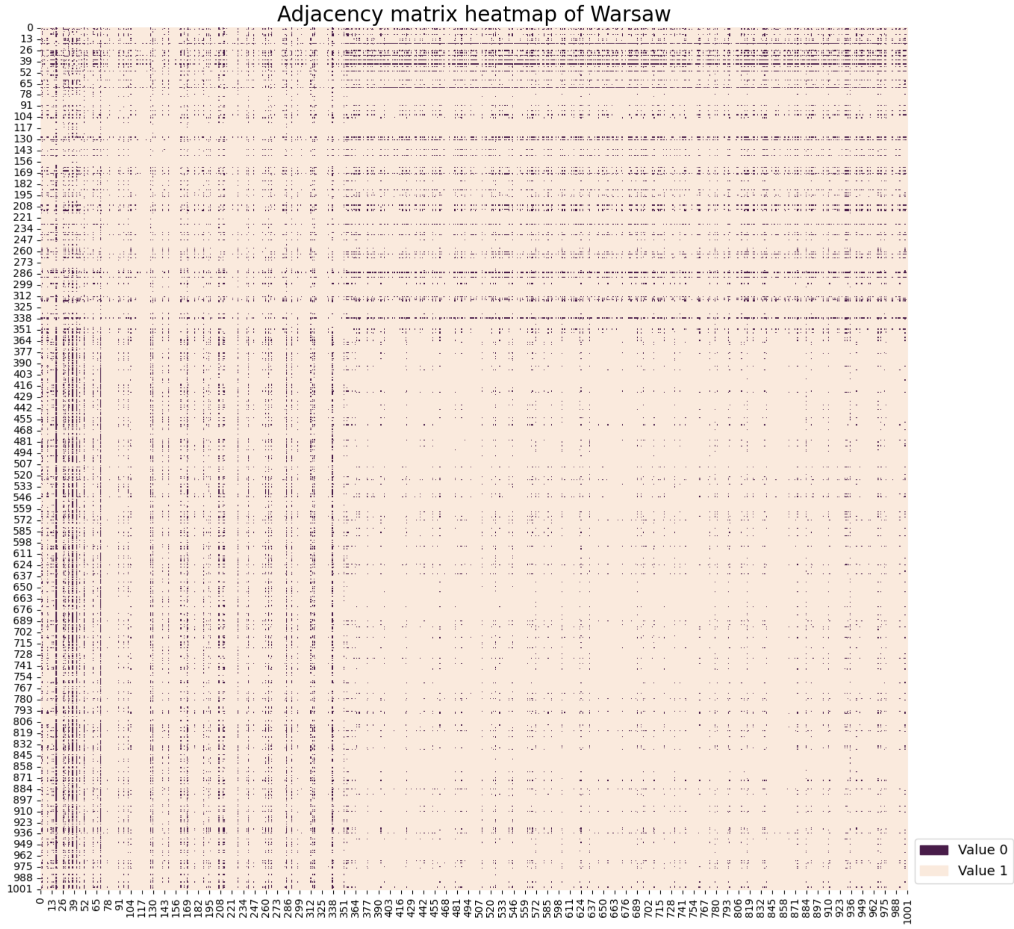
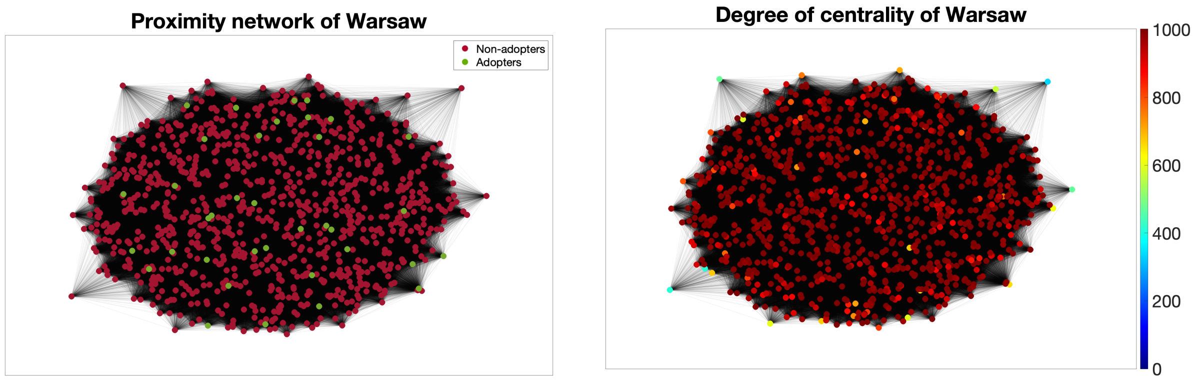


Figure 3.7: Heatmap of the adjacency matrix of Warsaw network.



(a) Proximity-based network of Warsaw. (b) Degree of centrality of Warsaw network nodes.

Figure 3.8: Topology of the Warsaw network.

These first analyses of the two networks don't reveal big differences. In fact, apart from the slightly different early adopters percentage, the only distinction between Milan and

Warsaw is that the first has lower values for what concern the degree of centrality (see Fig. 3.5b and Fig. 3.8b) and also it is less well-connected as is clear by comparing the Fig. 3.4 and Fig. 3.7.

3.3. Network clustering

In this section, we will discuss the clustering process and its outcomes for each of the two networks presented above. This procedure allows us to divide the population into groups of users with similar inclinations and therefore it is important during the analysis of the evolution of the networks. In fact, it helps us understand how different types of people react to policy and how the control action affects them.

To have an effective and an easy to implement clustering method, we chose k-means clustering. K-means clustering is an algorithm that aims to divide the dataset into K unique, non-overlapping subgroups (clusters), where every data point of the information set is part of only one group. The distribution of points is done in order to minimise the sum of the squared distances between the data points and the cluster centroid, i.e. the average value of all members of the cluster [6].

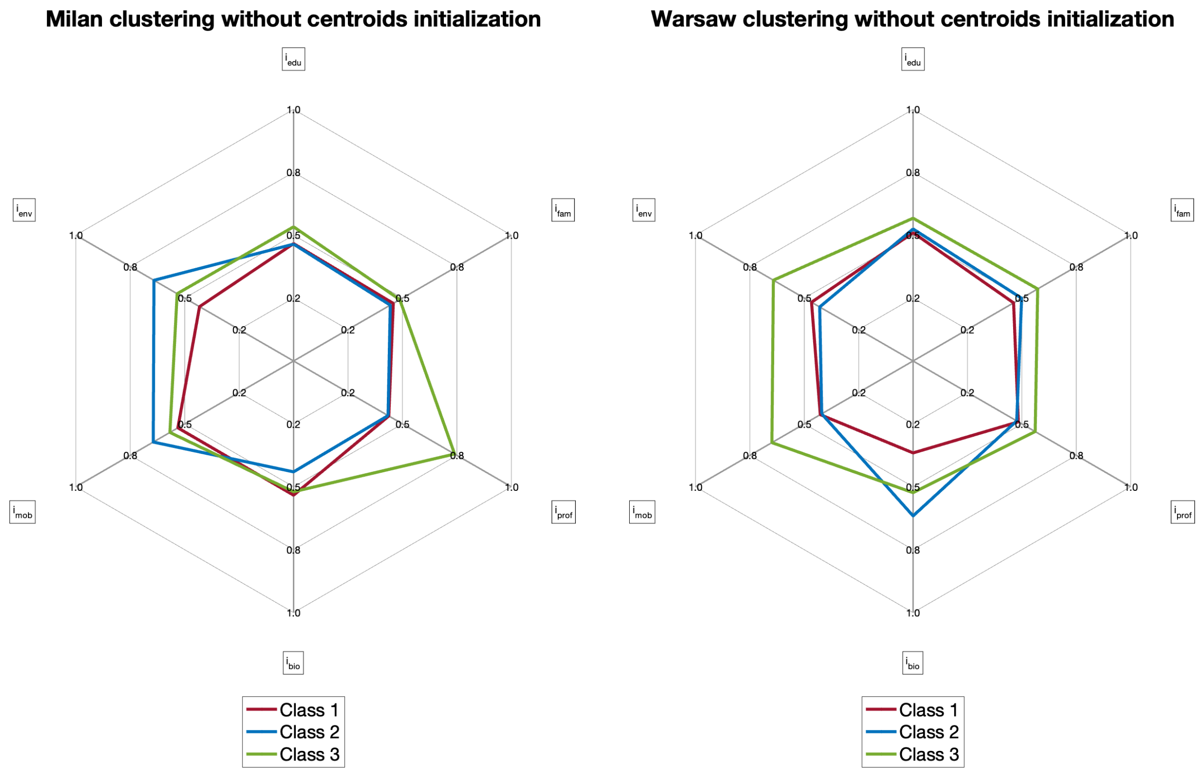
Clustering analysis may be done based on users' characteristics therefore considering sample subgroups based on chosen features. In this case, the chosen attribute is the Sharing-DNA, i.e. the seven indices i_{mob}, i_{geo}, i_{env}, i_{edu}, i_{prof}, i_{bio} and i_{fam}. Moreover, the choice of K is here K=3 so that the outcomes will present three clusters that embody high, medium, and low attitudes towards shared mobility services.

Finally, rather than making the centroids be calculated by the algorithm, here we decide to initialize them in order to have a better shaping of the clusters. In fact, as we will see in the next section, without this additional specification the clusters of both of the networks turn out to be not so well separated.

3.3.1. Results and comparison

The tool that will be used to achieve the goal of this part of the work is the function *kmeans* in *MATLAB*.

At first, for each of the two networks, the only things that we specified are the set of the six above-mentioned features and the wanted number of clusters. As stated above, to analyze easily and intuitively the outcomes, we decided to plot the results using two spider-plots, where vertices represent the average value of a certain index with respect to all users belonging to the same class. Here we exclude the country index since, being the networks built with respect to a specific region, all of the users share the same i_{geo}.

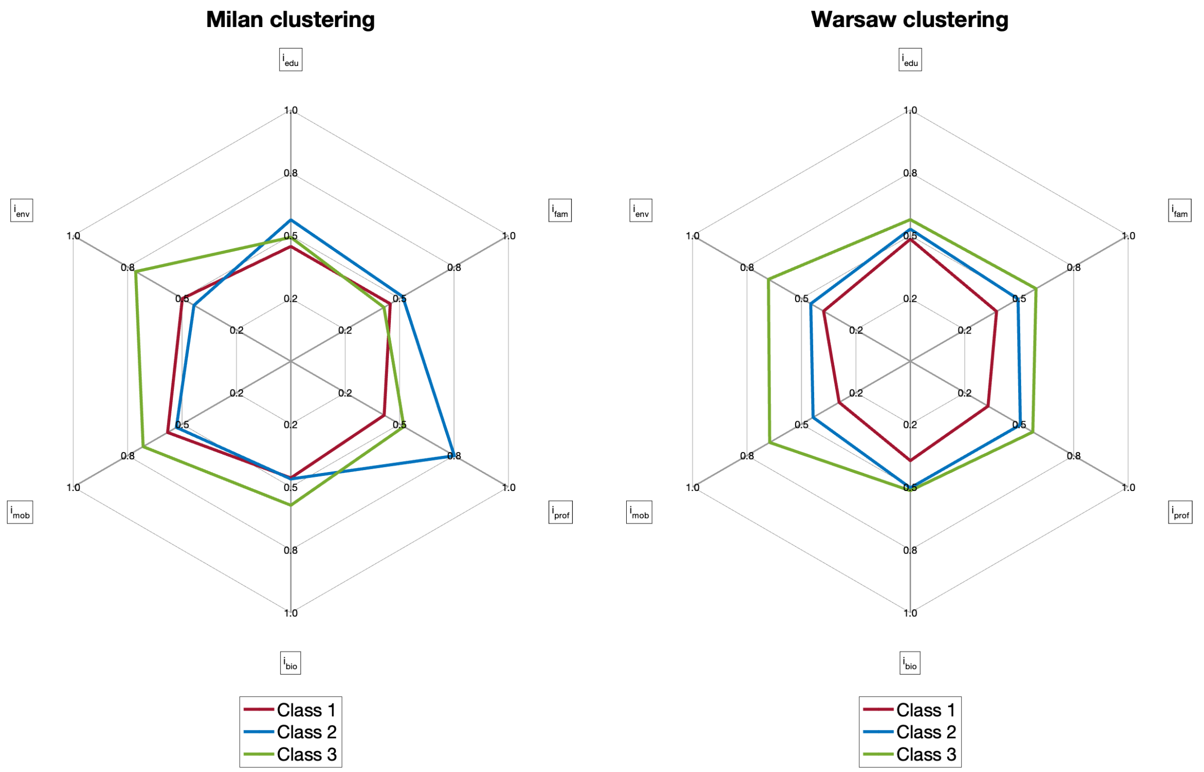


(a) (b)

Figure 3.9: The clustering process without centroids initialization leads to a result, both for 3.9a and for 3.9b, of not well-separated clusters.

From the outcomes presented in Fig. 3.9, it's evident that the obtained clustering is not so clear. The resulting is far from our given goal, as classes are not representative of high, medium, and low propensity. This can be seen from Milan clustering (see Fig. 3.8a), where the Class 2 has the higher values for the i_{env} and i_{mob}, but the lowers for the other indices. A similar thing happens for what concerns Warsaw (see Fig. 3.8b), where the Class 2 has the higher values for i_{edu}, i_{fam} and i_{prof} and lower values for i_{env} and i_{mob}. Furthermore, in both cases the area described by the plots are not clearly different from one another.

That being said, we decide to proceed with the centroids initialization and, after some tuning, the chosen centroids values are 0.3, 0.5 and 0.7, the same for all the seven indices.



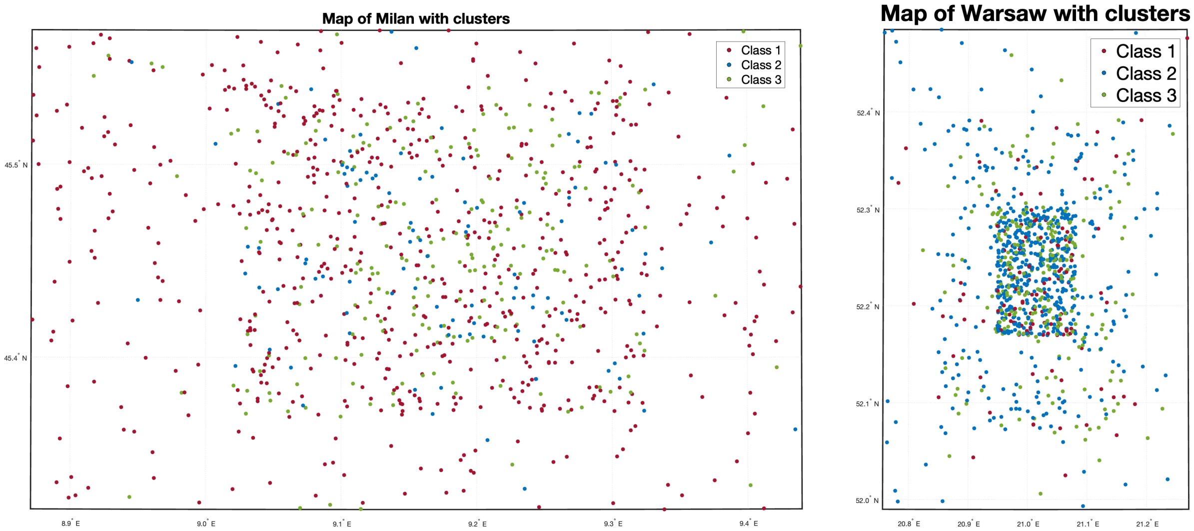
(a) Clustering of Milan network with centroids initialization.

(b) Clustering of Warsaw network with centroids initialization.

Figure 3.10: The clustering process with centroids initialization leads to the result of not well-separated clusters for 3.10a and well-separated ones for 3.10b.

These results enable us to see the big difference between the Milan and the Warsaw network. In fact, for Milan the outcome is not so different from the one with no centroids initialization, while for Warsaw now we have three well-separated clusters (see Fig. 3.10). This is probably due to the fact that in Milan the users are all equally too inclined to the technology. Now focusing on Warsaw network, the areas distinguish themselves for their size and the values of the hexagons' vertices are consistent with one another (see Fig. 3.10b). This allows us to finally identify the low, the medium, and the high propensity groups which are respectively the Class 1, the Class 2 and the Class 3.

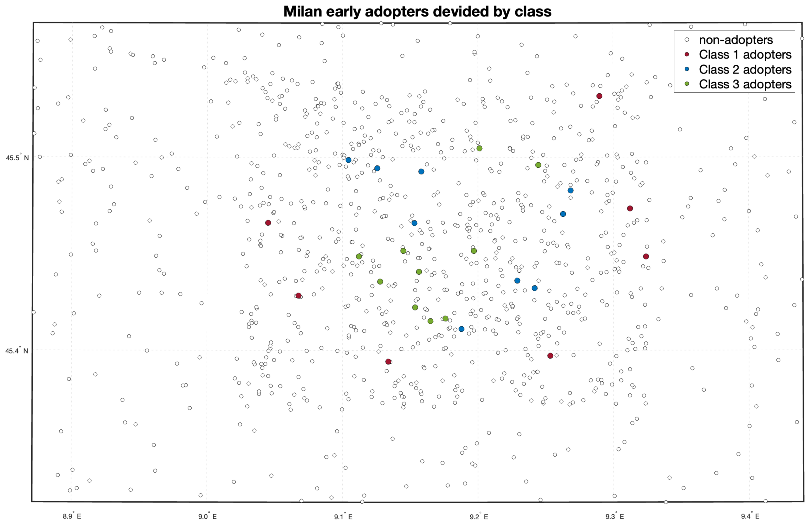
Using this clustering, some other considerations can be made. Even though there is no particular disposition of different type of users within the network (see Fig 3.11), it is interesting to see the class of the early adopters. Besides the fact that, as we already seen, they are all in the center of the cities, for what concern Warsaw network (see Fig 3.12b), all of the early adopters belong to the most inclined class, the Class 3. On the other hand, by the very nature of its clustering, Milan's early adopters belongs to all three classes.



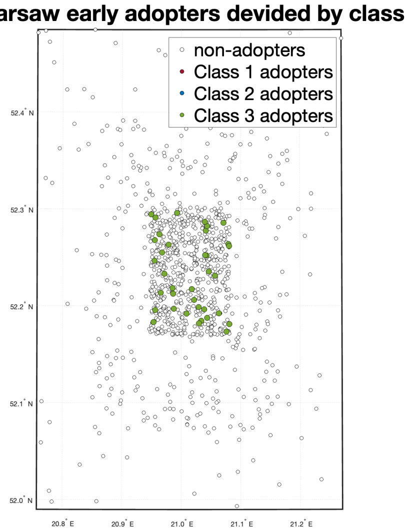
(a) Map of Milan with nodes dived by classes.

(b) Map of Warsaw with nodes dived by classes.

Figure 3.11: Maps of Milan and Warsaw with nodes dived by classes don't show a particular pattern.



(a) Milan's early users per classes.



(b) Warsaw's early users per classes.

Figure 3.12: Milan's early users belongs to all three classes (3.12a) while Warsaw's early users only belong to Class 3 (3.12b).

4 | Adoption model

This chapter will discuss how the two networks presented evolved over time and how adoption spread. We will see that to foster the adoption of Car-sharing services policies can be designed. That said, three different formulations will be shown. Firstly, a free evolution scenario will be proposed, where no use of incentives is made. Secondly, we will deal with a closed-loop environment where incentive schemes are optimized and applied. Lastly, a method to add the concept of fairness directly into the design process will be proposed and an index to assess its impact will be introduced.

4.1. Open-loop cascade model

4.1.1. Formulation

To be able to simulate the free evolution of the adoption model, we start by considering the model in Eq. 2.3 with an initial set (see Eq. 2.2) considered as in Table 3.3. At each time step t, the set of agents that have switched to Car-sharing services satisfies the relationship at Eq. 2.4. Hence, what guides the adoption through the network is only the relative popularity of the technology.

4.1.2. Simulation

Considering now the network of Milan and Warsaw and the model presented, in order to test the evolution of the adoption, we first have to choose the simulation parameters.

As the time span of the simulation, T=10 steps of 3 months each is considered. The total time period of two and a half years and chosen time intervals enable us to analyse a very realistic situation in which a person's opinion of one mobility solution does not vary often over time.

4.2. Closed-loop model with Linear Quadratic Regulator

In the original formulation of the model proposed in Section 4.1.1, the adoption is only driven by the relative popularity of the technology among the neighbours of each person without considering the time-varying nature of the individuals' propensity. Now, driven by the same logic, we consider the model described by Eq. 2.3, but with time-varying thresholds:

$$x_v(t+1) = \begin{cases} 1, & \text{if } x_v(t) = 1 \text{ or } \frac{1}{|\mathcal{N}_v|}\sum_{w\in\mathcal{N}_v} x_w(t) \geq \alpha_v(t) \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

Since we assume that the individual state is not directly controllable, this new formulation allows us to regard the thresholds as directly modifiable and therefore they can now directly be changed by external incentives. This assumption implies that the influence that the adoption behaviour of others has on an individual's decision to use a Car-sharing service, i.e. the social contagion, is favoured by the policy and consequently the adoption behaviour can be now triggered. Accordingly, the impact of a policy is taken into account by considering threshold dynamics which is driven by the following difference equation:

$$\alpha_v(t+1) = \alpha_v(t) + B_v u_v(t),\ \ t = 0, ..., T,\ \ v \in \mathcal{V} \tag{4.2}$$

where $B_v \in \mathbb{R}$ quantifies how much the $v-th$ agent is willing to accept the policy, thus indicating the impact level of the policy $u_v \in \mathbb{R}$ directed to the $v-th$ user.

4.2.1. Formulation

The properties of the irreversible cascade model in Eq. 4.1 are combined with the threshold dynamics in Eq. 4.5 in order to build feedback strategies that will increase adoption throughout the network. By doing so, we can actively use the latest information on the agents' opinions to adjust the incentive approach to shifts in those agents' attitudes regarding sharing mobility services over time.

Under the assumption that we can access the values of the threshold $\alpha_v(t)$ for all users and all time instant t=0,...,T, our aim is to find a policy that will enhance the number of adopters without requiring too many investments. To do so, the optimal solution that takes into account both objectives can be found using a finite horizon Linear Quadratic Control. Keeping in mind the generic form of the cost function in discrete time of Eq. 2.7, the problem that here we want to solve is the following minimization problem:

$$\min_{\{K_v\}_{v\in V}} \sum_{v\in V}\left[\sum_{t=0}^{T-1}\left(Q(t)\xi_v(t)^2 + Ru(t)^2\right) + Q(T)\xi_v(T)^2\right] \tag{4.3a}$$

$$s.t.\ \xi_v(t+1) = \xi_v(t) + B_v u_v(t),\ t = 0, ..., T-1,\ v \in \mathcal{V} \tag{4.3b}$$

$$u_v(t) = K_v(t)\xi_v(t),\ t = 0, ..., T-1,\ v \in \mathcal{V} \tag{4.3c}$$

where

- $\xi_v(t) = (\alpha_v(t) - \bar{\alpha}_v)$ is the tracking error, i.e the difference between the current threshold value and the corresponding target value;
- $\bar{\alpha}_v \in [0, 1]$ is the target values for the threshold of the $v-th$ agent;
- $\{Q(t)_v\}_{t=0}^{T} \geq 0$ are time-varying parameters which penalize the mismatches between the desired and actual state;
- R > 0 is a parameter which penalizes the control action over the time horizon T;
- $\{K_v(t)\}_{t=0}^{T}$ are the set of gains of the incentives given to the $v-th$ user over all time instants.

The target thresholds $\bar{\alpha}_v$ implicitly indicate how much we want the strategy to overcome agents' early reluctance to the adoption. Theoretically, we could want the resistance to adopting to be null and so the thresholds to be forced to drop to zero over time by setting the constant target $\bar{\alpha}_v = 0$. However, such a decision is likely to result in a waste of resources. In fact, for the $v-th$ agent to adopt, it is sufficient for $\alpha_v(t)$ to be equal to the percentage of neighbour adopters at time t. In order to consider this value constant over time, we impose this relationship:

$$\bar{\alpha}_v = \frac{1}{|N|_v}\sum_{w\in|N|_v} x_w(0),\ \forall v \in \mathcal{V}. \tag{4.4}$$

Therefore, for each user, the target value is the average of early adopters among the neighbours at the initial time t=0.

Other important terms of the formulation are the values $Q_v(t)$, the weights of the tracking objective. In order to design an efficient control, we can set them equal to the constant value Q when the $v-th$ agent is not an adopter and set them to 0 when the $v-th$ agent is an adopter. Consequently, the following holds:

$$Q_v(t) = Q(1 - x_v(t)),\ \forall v \in \mathcal{V}. \tag{4.5}$$

The minimization problem in 4.3 can now be solved by solving $\mathcal{V}$ different LQR problems relying on Bellman's equation [3]. The process will return the optimal gains $\{K_v(t)\}_{t=0}^{T}$ and through the Eq. 4.12c the optimal feedback policies will be derived.

The formulation just outlined is not the only possible method to design the policy schemes. In fact, throughout this work, we will consider the possibility to act directly not only on the thresholds α_v, but also on some of the features that characterize the Sharing-DNA.

Scenario A: policy schemes acting on the sharing index

With the purpose of making the formulation in 4.3 more intuitive, we can consider the tracking error rather than the difference between the thresholds, the reluctance to adopt, α_v and the target $\bar{\alpha_v}$, the difference between the Sharing index $i_{sha,v}$, the inclination to adopt, and the target $\bar{i}_{sha,v}$. The letter is found from the definition of the threshold (see Eq. 3.6) as follows:

$$\bar{i}_{sha,v} = 1 - \bar{\alpha}_v \tag{4.6}$$

hence, That being said, the first formulation that we will use for the simulation will be outlined in this way:

$$\min_{\{K_v\}_{v\in V}} \sum_{v\in V}\left[\sum_{t=0}^{T-1}\left(Q(t)\eta_v(t)^2 + Ru(t)^2\right) + Q(T)\eta_v(T)^2\right] \tag{4.7a}$$

$$s.t.\ \eta_v(t+1) = \eta_v(t) + B_v u_v(t),\ t = 0, ..., T-1,\ v \in \mathcal{V} \tag{4.7b}$$

$$u_v(t) = K_v(t)\eta_v(t),\ t = 0, ..., T-1,\ v \in \mathcal{V} \tag{4.7c}$$

where $\eta_v(t) = i_{sha,v}(t) - \bar{i}_{sha,v}$.

Scenario B: policy schemes acting on the alterable DNA features while tracking the sharing index

Another way to boost the adoption is to directly act on some of the features of the Sharing-DNA, the alterable ones. In fact, it is reasonable to think that not all of the indices can be modified by incentives, namely those regarding education, profession, age and country of origin. Accordingly, the indices on which we will directly act with policy schemes are i_{mob}, i_{env} and i_{fam}, i.e. those related to the mobility habits, the interest toward the environmental issue and the level of income. That being said, the formulation will look like this:

$$\min_{\{K_v\}_{v\in V}} \sum_{v\in V} \left[\sum_{t=0}^{T-1} \left(Q(t)\eta_v(t)^2 + Ru(t)^2\right) + Q(T)\eta_v(T)^2 \right] \tag{4.8a}$$

$$s.t.\ i_{DNA,v}(t+1) = i_{DNA,v}(t) + B_v u_v(t),\ t = 0, ..., T-1,\ v \in \mathcal{V} \tag{4.8b}$$

$$u_v(t) = K_v(t)\eta_v(t),\ t = 0, ..., T-1,\ v \in \mathcal{V} \tag{4.8c}$$

$$i_{sha,v}(t) = C i_{DNA,v}(t),\ t = 0, ..., T-1,\ v \in \mathcal{V} \tag{4.8d}$$

where $i_{DNA,v}$ is the vector composed by the features of the Sahring-DNA and the C in 4.8d represents the mean operator vector. In this case B_v rather than a scalar it is a vector with the element corresponding to the uneditable features set to 0.

It is worth noting that the tracking error and the dynamics on which we act have not the same object. In fact, in this case, the incentives influence the features of the DNA, but the error that LQR problem aims at tracking is related to the Sharing index.

Scenario C: policy schemes acting on the alterable DNA features while tracking the DNA features

Let us consider now the policy schemes acting on the changeable DNA features as it was in the above formulation, but unlike before, this time we want to consider the tracking error as the difference between the Sharing-DNA $i_{DNA,v}$ and the target $\bar{i}_{DNA,v}$. The letter is found considering the definition of the threshold (see Eq. 3.6) and the relationship between the Sharing-DNA and the Sharing index (see Eq. 3.2). In addition, if we set the targets of all of the features to be equal, we have the following outcome for all of the seven indices:

$$\bar{i}_{DNA,v}(i) = 1 - \bar{\alpha}_v,\ \forall i = 1, ..., 7. \tag{4.9}$$

Therefore, this third formulation is outlined as follows:

$$\min_{\{K_v\}_{v\in V}} \sum_{v\in V} \left[\sum_{t=0}^{T-1} \left(Q(t)\zeta_v(t)^2 + Ru(t)^2\right) + Q(T)\zeta_v(T)^2 \right] \tag{4.10a}$$

$$s.t.\ i_{DNA,v}(t+1) = i_{DNA,v}(t) + B_v u_v(t),\ t = 0, ..., T-1,\ v \in \mathcal{V} \tag{4.10b}$$

$$u_v(t) = K_v(t)\zeta_v(t),\ t = 0, ..., T-1,\ v \in \mathcal{V} \tag{4.10c}$$

where $\zeta_v(t) = i_{DNA,v}(t) - \bar{i}_{DNA,v}$.

4.2.2. Simulation

Also in the closed-loop scenario, the chosen finite time horizon has length T = 10 steps, each corresponding to 3 months. In order to investigate the optimal policies obtained as in the previous section, multiple simulations over different choices of B and for a selected tuning of the weights $Q = 10$ and $R = 1$ will be carried out. As such, this choice of weights attribute a priority to tracking objective with respect to the use of resources. In fact, in this work, we will not consider any limitation on the input that we can utilize.

Scenario A with random values of B

The first policy is designed with $Q = 10$ and $R = 1$ and the values B_v, scalar values, are selected as uniformly distributed random numbers in the interval [0,1]. With the random choice of the values of B_v, hence we do not make any difference between the level of acceptance of the users with respect to their characteristics.

Scenario A with selected values of B

The second policy scheme is designed with $Q = 10$ and $R = 1$ and the values B_v, scalar values, are selected with respect to the cluster of belonging of each user. In fact, as the choice to start using Car-sharing services depends on the features of the individual, it is reasonable to consider the level of acceptance of the policy, i.e. how much the policy can actually act and change a user's characteristics, conditioned by the Sharing-DNA as well. That being said, in order to pick the values of B_v, the interval [0,1] is divided into three so that each range corresponds to a cluster. The values are then randomly selected from the range corresponding to the users' cluster. In this way, the level of acceptance of the users depends on the cluster of belonging and hence on their characteristics.

Scenario B with selected values of B

The second policy scheme is designed with $Q = 10$ and $R = 1$ and the B_v, vectors in this scenario, are again selected with respect to the cluster of belonging of each user. In this case, as for each agent' the B_v elements correspond to each of the three features on which the policy act, the choice of these elements will depend on the values of the three features. Therefore, for each feature, an interval [0,1] is divided into three so that each range corresponds to a cluster. The values of B are then randomly selected from the ranges of corresponding to the users' features' cluster. In this way, the level of acceptance of the users depends again on their characteristics.

As we have discussed in this Scenario we will act on the users' adoption propensity by acting on only editable features. In order for the simulation to be meaningful, we have to make sure that the Sharing indices that we can obtain are high enough to overcome the adoption thresholds. Since this doesn't happen, in order to bypass the problem, the considered solution is to change the weights for the computation of the Sharing index i_{sha} of the seven features. Let us consider the outcomes obtained by the Permutation Importance algorithm discussed in Section 3.1.2. In addition to the list of the seven more relevant features that compose the Sharing-DNA, thanks to this algorithm we know that the three most influential features are those related to the indices i_{mob}, i_{env} and i_{geo}. Accordingly, to make the editable features (i_{mob}, i_{env},i_{fam}) weigh more on the final output of the Sharing index and at the same time be consistent with the results of Permutation Importance algorithm, from now on we will use the weights listed in Table 4.1.

DNA index	Weight
i_{mob}	0.0067
i_{geo}	0.03
i_{env}	0.46
i_{edu}	0.0067
i_{prof}	0.0067
i_{bio}	0.0067
i_{fam}	0.03

Table 4.1: Features weights.

Scenario C with selected values of B

The third policy scheme is designed with $Q = 10$ and $R = 1$ and the B_v, vectors, are selected with respect to the cluster of belonging of each user. To do that, the values are chosen as in the *Scenario B* and therefore, the level of acceptance of the users depends on their changeable features. Moreover, also for this simulation, we will consider the weights of the features listed in Table 4.1.

4.3. Adding Fairness to LQR

We now want to attempt to include a concept of fairness in the formulation of the LQR problem. To do so, as already discussed in Section 2.4, for the formalization of fairness we rely on the concept of equity. The key to the latter lies in how the resources are used. In fact, if the idea of equality implies that everyone or every group of individuals receive the same chances or resources, on the other hand, equity takes into account the fact that each person's circumstances are unique and provides the precise resources and opportunities required to achieve equality [20]. Based on this, we can translate this concept with the minimization of the distance from the target for every person. Accordingly, the equity can be described as follows:

$$\min \sum_{p \in P} \left[(x_p - \bar{x}_p) - \frac{1}{N} \sum_{b \in P,\, with\ b \neq p} (x_b - \bar{x}_b) \right] \tag{4.11}$$

where x_p is the state of the $p-th$ person of a group P of N people and $\bar{x}_p$ is its target. In this way, the objective of equity is that of minimizing the difference between each person's distance from the respective target and the mean of the others' distances from their target.

4.3.1. Formulation

By relying on the formulation of the Scenario C discussed in Section 4.2.1, we can now add the equity element as outlined in 4.11. We want to consider this additional element in such a way that it has the form of a tracking error objective where the actual state is the tracking error $\zeta_v(t)$ and the target is the mean of the other tracking errors at time t. Consequently, the fair LQR will be as follows:

$$\min_{\{K_v\}_{v \in V}} \sum_{v \in V} \left[\sum_{t=0}^{T-1} \left(Q(t)\zeta_v(t)^2 + W(t)\left(\zeta_v(t) - \frac{1}{N}\sum_{j=1}^{N} \zeta_j(t)\right)^2 + Ru(t)^2 \right) + Q(T)\zeta_v(T)^2 \right] \tag{4.12a}$$

$$s.t.\ i_{DNA,v}(t+1) = i_{DNA,v}(t) + B_v u_v(t),\ t = 0, ..., T-1,\ v \in \mathcal{V} \tag{4.12b}$$

$$u_v(t) = K_v(t)\zeta_v(t),\ t = 0, ..., T-1,\ v \in \mathcal{V} \tag{4.12c}$$

$$\tag{4.12d}$$

where the new weight $\{W(t)\}_{t=0}^{T} \geq 0$ penalize the equity factor.

4.3.2. Networks Resizing

To fully analyze how the concept of fairness influences the adoption process it is best to work with clearly separated clusters. That said, in this scenario, we will work only with the Warsaw network.

Considering the fair LQR formulation we encountered some problems as its implementation on MATLAB needed to work with much bigger matrices with respect to the other cases. As a result, because of its big number of nodes, it was impossible to work with the network used until now. For this reason, in order to do some simulations, we have decided to use another Warsaw's network that we built by extracting fewer nodes, i.e. 202.

Since this new network was retrieved in the way as the previous one, it has almost the same characteristics. In fact, also this network is almost fully connected (see Fig. 4.2) and the early adopters, which correspond to the 1.98% of the total nodes, are in the center of the map (see Fig. 4.1).

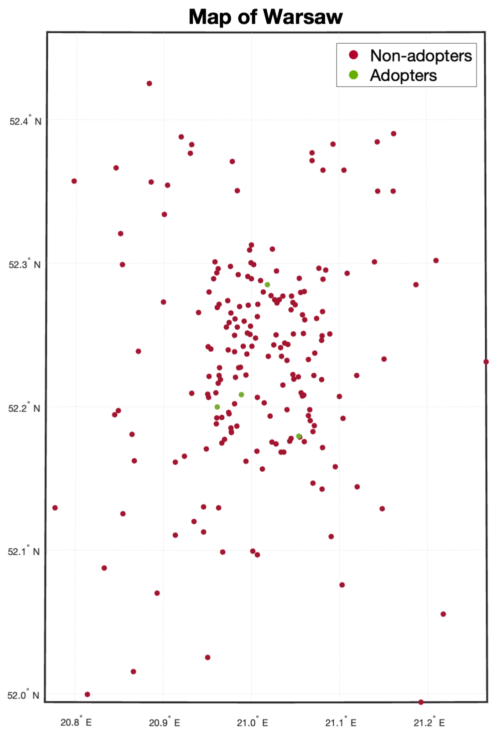


Figure 4.1: Proximity-based smaller network of Warsaw.

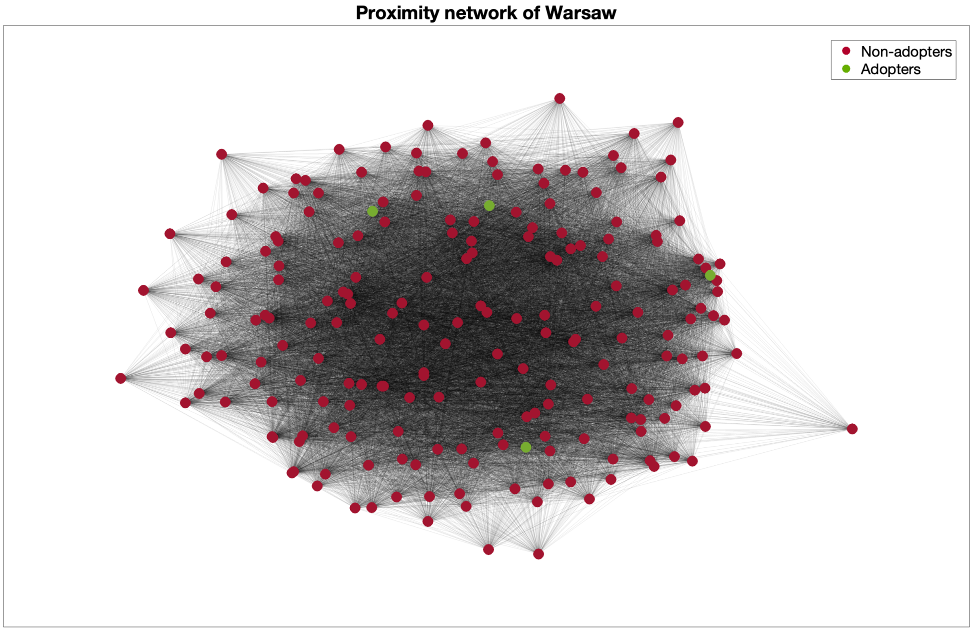


Figure 4.2: Map of Warsaw with the smaller network.

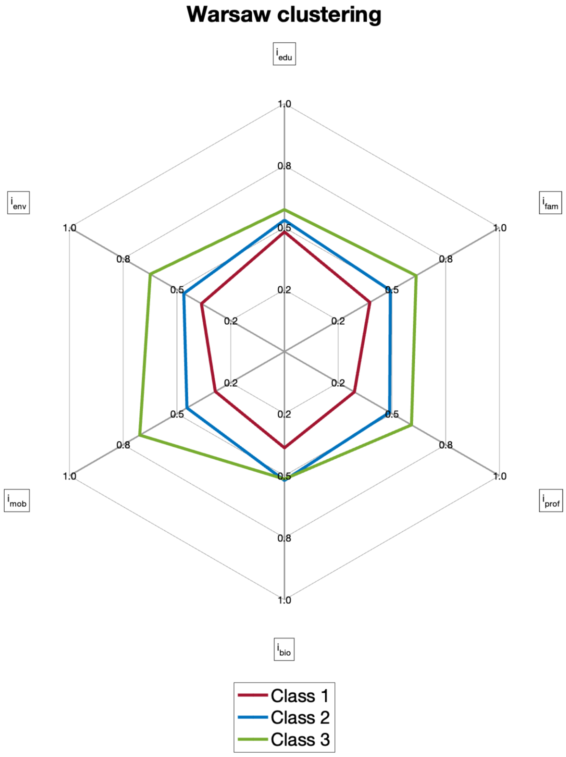


Figure 4.3: Clustering of the smaller network of Warsaw

The clustering outcome is similar likewise. As in the original networks (see Fig. 3.10b) it has clearly separated clusters (see Fig. 4.3). Moreover, also in this case all the early adopters belong to Class 3 (see Fig. 5.1).

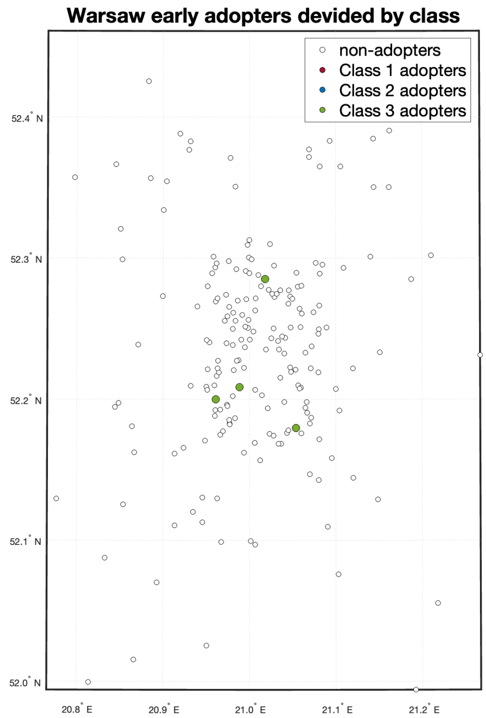


Figure 4.4: Warsaw's early users per classes using the smaller network

4.3.3. Simulation

The simulations carried out to find the fair policy schemes are designed with Q = 10, R = 1 and different tuning of W, namely W=1 and W=10. The B_v, that in this scenario are vecotors, are selected as in Scenario C discussed in Section 4.2.2 thus with respect to the clusters of belonging of each user so that their level of acceptance depends on their editable features.

Another important element of the simulations is the Index of equity which will allow us to quantify the actual effect that fairness has on the policy schemes. In order to estimate this effect we compare the individual tracking errors e_v for $v \in \mathcal{V}$ with respect to the average over the other $w-th$ agents for w $\in \mathcal{V}$. In particular, let $E(t)$ indicate the average deviation of the tracking errors with respect to their mean. It can be written for every time t in the following way:

$$E(t) = \frac{1}{N} \sum_{v \in V} \left\| e_v(t) - \frac{1}{N} \sum_{w \in V} e_w(t) \right\|_2 \tag{4.13}$$

Therefore we express the Index of equity, $\mathcal{I}_e$, as follows:

$$\mathcal{I}_e(t) = e^{-E(t)} \tag{4.14}$$

5 | Analyses and comparison of the results

In this Chapter, the results and analysis will be carried out. We start by proposing the results of the first scenario, i.e. the free evolution of the network. Afterwards the three scenarios in the closed loop described above we will analyze how the two networks evolve. To do so we will examine different factors that cause such evolution including the shape of Sharing-DNA over time and the optimal allocation of the inputs from the policy schemes. In addition, a further comparison of the two networks will be made and it will be investigated how the shape and the characteristics of the network influence the outcomes.

Lastly, the differences between the policy schemes obtained with the fair LQR formulation (see Section 4.3) and that extracted considering the Scenario C discussed in Section 4.2.2 with the new smaller network (see Section 4.3.2) will be analyzed. To this end, we will try to assess the effect of introducing fairness in the Linear Quadratic Regulator. By considering the Index of equity (see Eq. 4.14) that quantifies the level of justice achieved, we discuss the effect induced on both tracking performance and resource allocation by the penalization of unfair control actions according to the proposed framework.

5.1. Free evolution of the networks

Let us first consider the formulation and parameters described in Section 4.1. The results illustrated in Fig. 5.1 show that over the total considered time, both for Milan's and Warsaw's networks, no user has become an adopter.

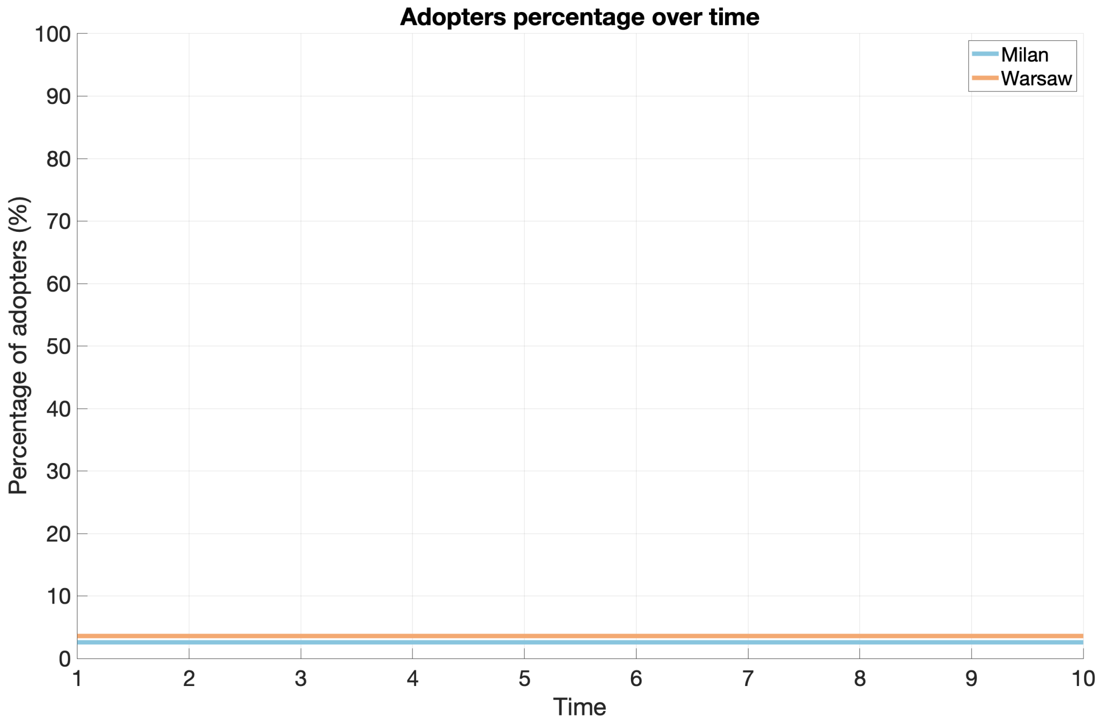


Figure 5.1: Milan's and Warsaw's adopters over time, free evolution. Both for Milan's network and for Warsaw's network the number of adopters don't vary.

This result may be affected by the shape of the network, the set of early adopters and the thresholds. Since the network is very well-connected and the majority of initial adopters are located in the parts of the networks where the degree of centrality is the highest (see Fig. 3.5 and Fig. 3.8), the outcome can be interpreted as a lack of early adopters. In light of the threshold values, the 2.59% for Milan and the 3.59% for Warsaw are not enough to trigger the spread of adoption.

2

5.2. Closed-loop network evolution: Milan vs. Warsaw

As it is clear from the results in the previous section, describing the process of adoption spreading with the Deterministic Irreversible Cascade Model of Eq. 2.3 is not even slightly sufficient to make the network evolve. For this reason, now we will analyze how fostering the adoption affect the evolution of the network.

5.2.1. Scenario A with random values of B

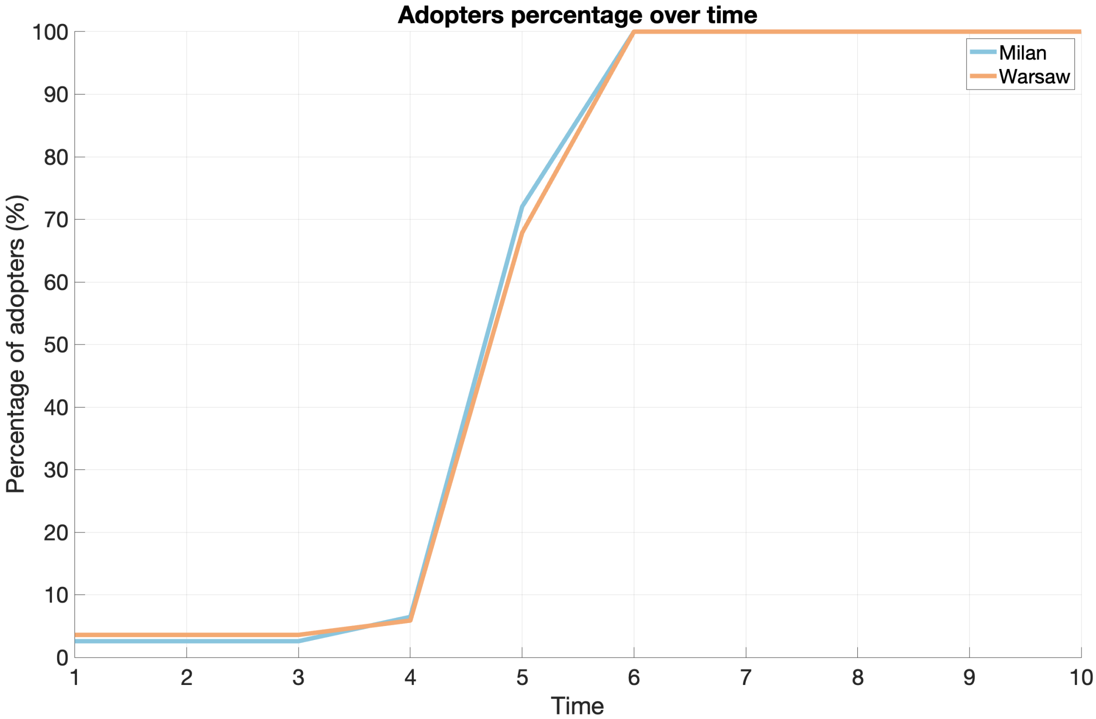


Figure 5.2: Milan's and Warsaw's adopters over time, Scenario A with random B. Both for Milan's network and for Warsaw's network the number of adopters reaches the 100% at t=6.

The results presented in Fig. 5.2 show that now both of the networks evolve over time and the percentage of adoption reaches the 100% of adopter users at t=6, so after 18 months.

5.2.2. Scenario A with selected values of B

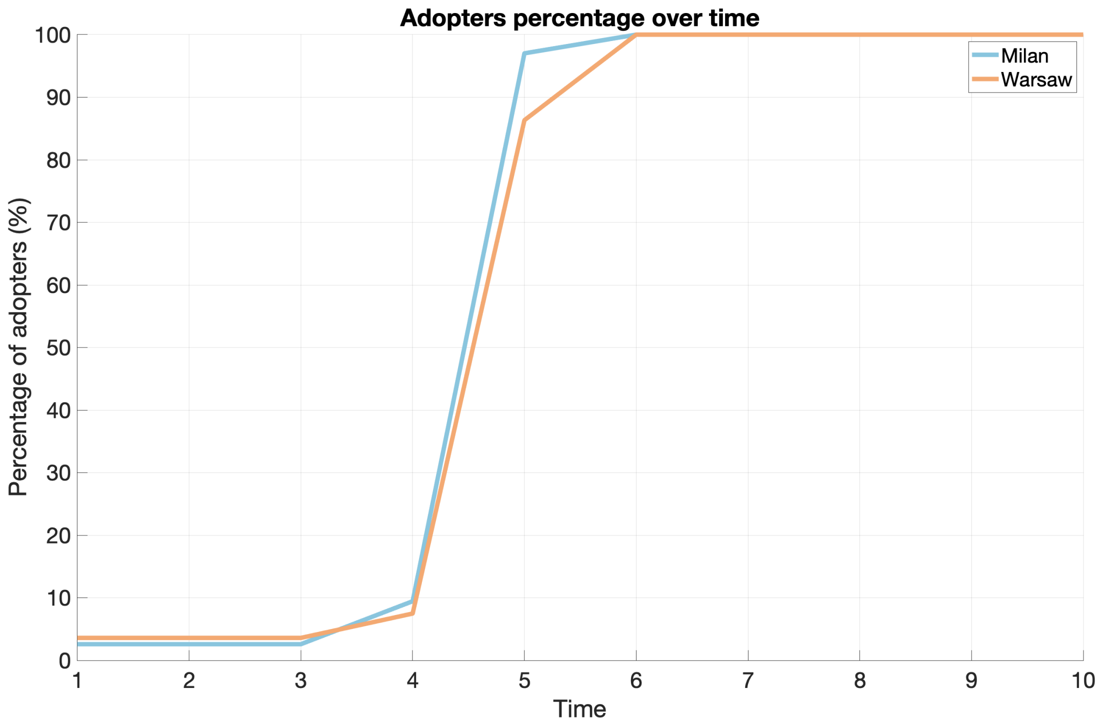
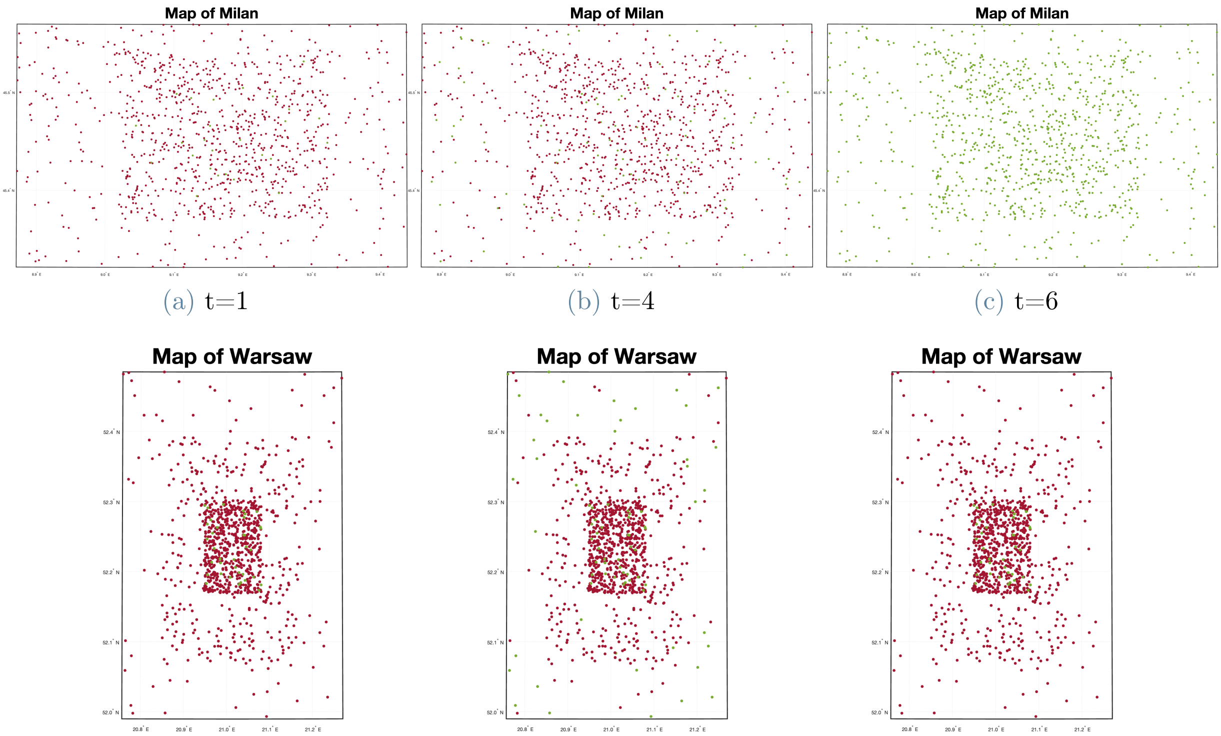


Figure 5.3: Milan's and Warsaw's adopters over time, Scenario A with selected B. Both for Milan's network and for Warsaw's network the number of adopters reaches the 100% at t=6.

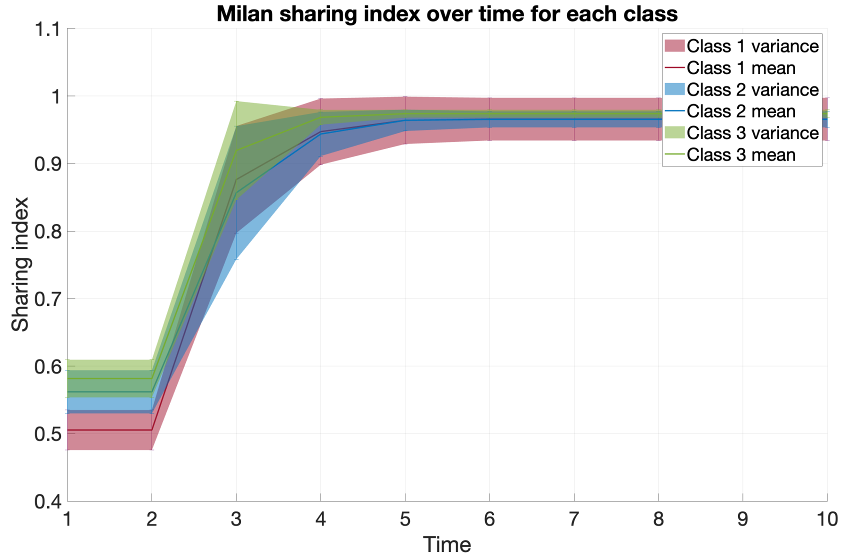
Considering this scenario, the full adoption is reached again at time t=6 (Fig. 5.4) and the actual spread of Car-sharing services in the map is the one depicted on the networks in Fig. 5.4. As it can be seen in Fig. 5.5 the networks evolve as the Sharing indices of the users increase boosted by the incentives.



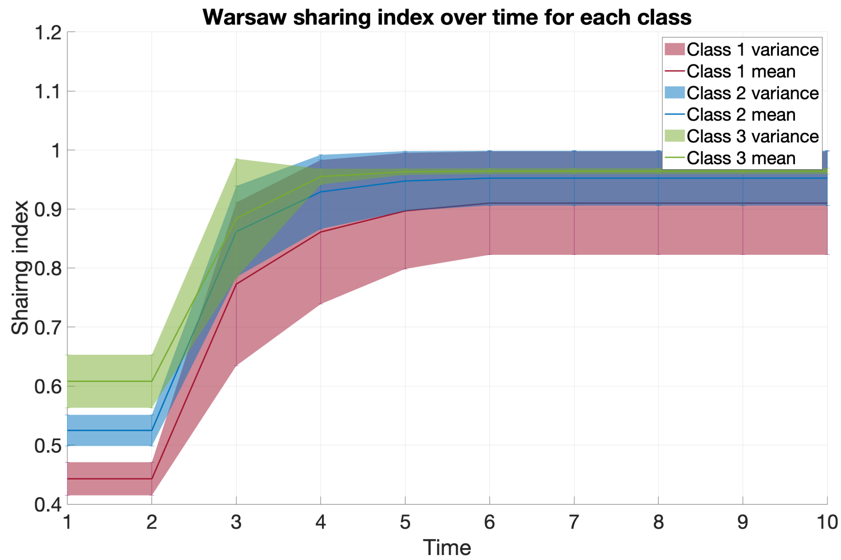
(a) t=1 (b) t=4 (c) t=6

(d) t=1 (e) t=4 (f) t=6

Figure 5.4: Evolution of the Milan's and Warsaw's network over time, Scenario A



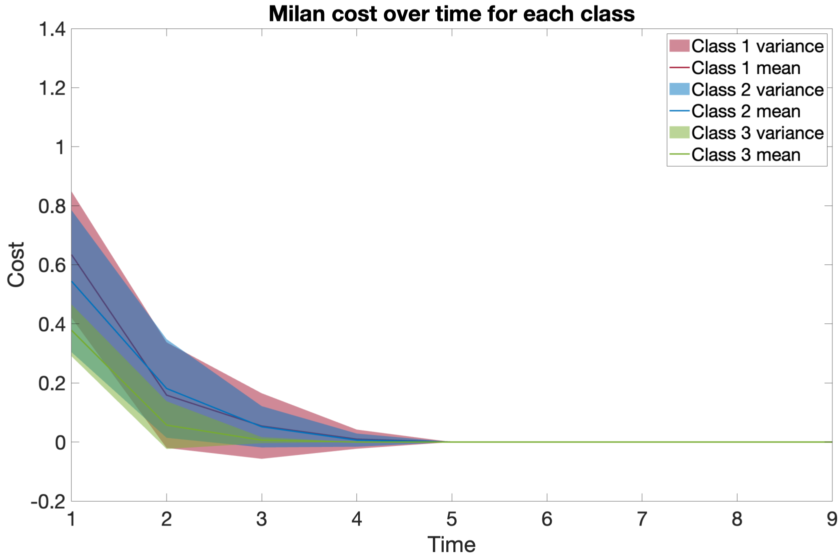
(a) Milan's Sharing index over time per classes, Scenario A.



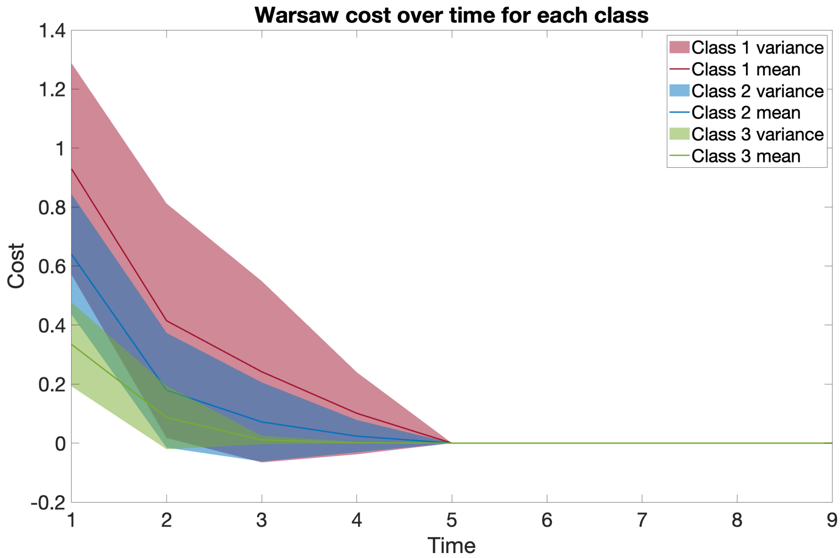
(b) Warsaw's Sharing index over time per classes, Scenario A.

Figure 5.5: Milan's and Warsaw's mean and variance of Sharing index over time per classes, Scenario A.

The policy scheme allocates the resources as shown in Fig. 5.6, and it can be seen that the cost allocated for the case Warsaw (see Fig. 5.6b) are higher than those of Milan (see Fig. 5.6a).



(a) Milan's cost over time per classes, Scenario A.



(b) Warsaw's cost over time per classes, Scenario A.

Figure 5.6: Milan's and Warsaw's mean cost for each feature over time per classes, Scenario A.

Analysing these results we can say that the different choices of the values of B_v as discussed in Section 4.2.2, lead to an evolution of the networks that is a little faster than those presented in the previous scenario. In fact, as can be seen in Fig. 5.3 after 5 time steps in both of the networks the percentage of adopters is above the 80% differently from the previous case (see Fig. 5.2).

The difference between the two networks can be seen by examining Fig. 5.5. In fact, as mentioned when we introduced and talked about the different clusterization outcomes (see Section 3.3), we can now see how Milan's and Warsaw's users are actually different. In Milan, the users are indeed all very similarly high inclined to the sharing mobility and even if the initial Sharing index values of users of different clusters are slightly separate, they end up being all indifferently very high (see Fig. 5.5a). On the other hand, we can clearly see different behaviour of Warsaw's clusters over the entire time horizon (see Fig. 5.5b) with the cluster of Class 1 present from beginning to end smaller values of the Sharing index. The difference between the Milan's and Warsaw's agents explained also the greater allocation for the latter of the inputs (see Table 5.1 and Table 5.2). In addition, in the Warsaw case, we can see clearly that the users of Class 1, i.e. those that are less inclined to make use of Car-sharing, receive more inputs than the others (see Fig.5.6a) precisely due to their bigger distance from the target.

5.2.3. Scenario B with selected values of B

Making reference to the second scenario, Scenario B (see Section 4.2.2), we now see how the direct acting on the features changes the outcomes. In the first attempt of the simulation, considering all of the DNA features with equal weights, as can be seen in Fig. 5.7, the network doesn't evolve and the percentage of acceptance remains to the 2.59% for Milan's network and to 3.59% for Warsaw's network.

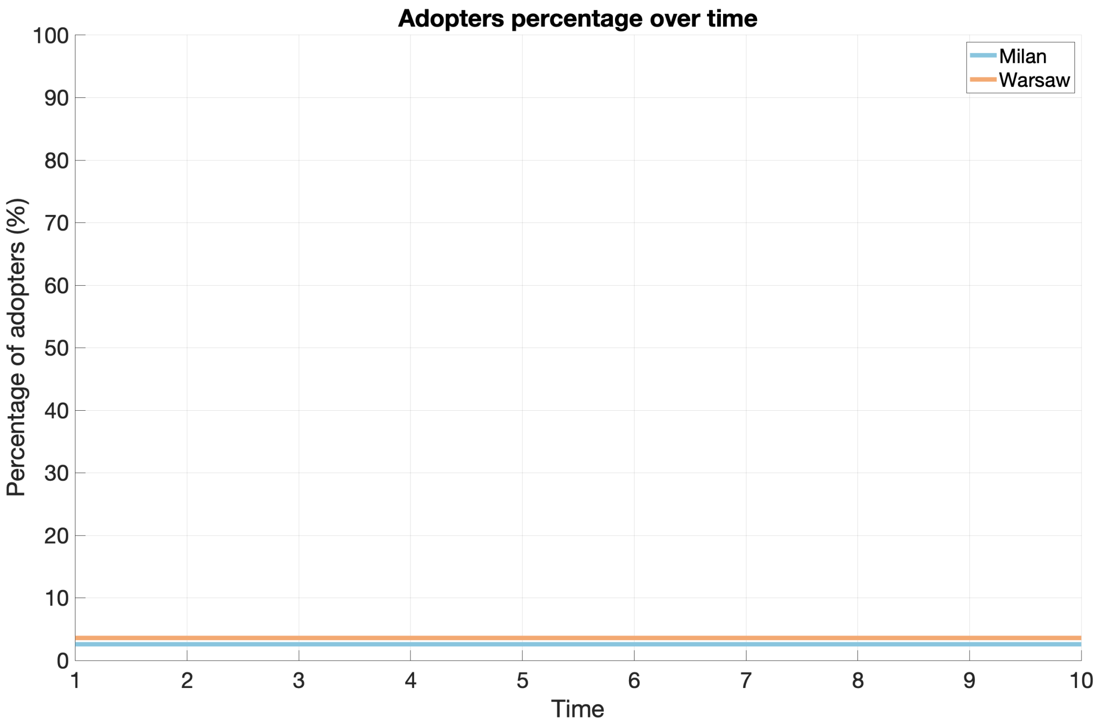


Figure 5.7: Milan's and Warsaw's adopters over time, Scenario B first attempt. Both for Milan's network and for Warsaw's network the number of adopters don't vary.

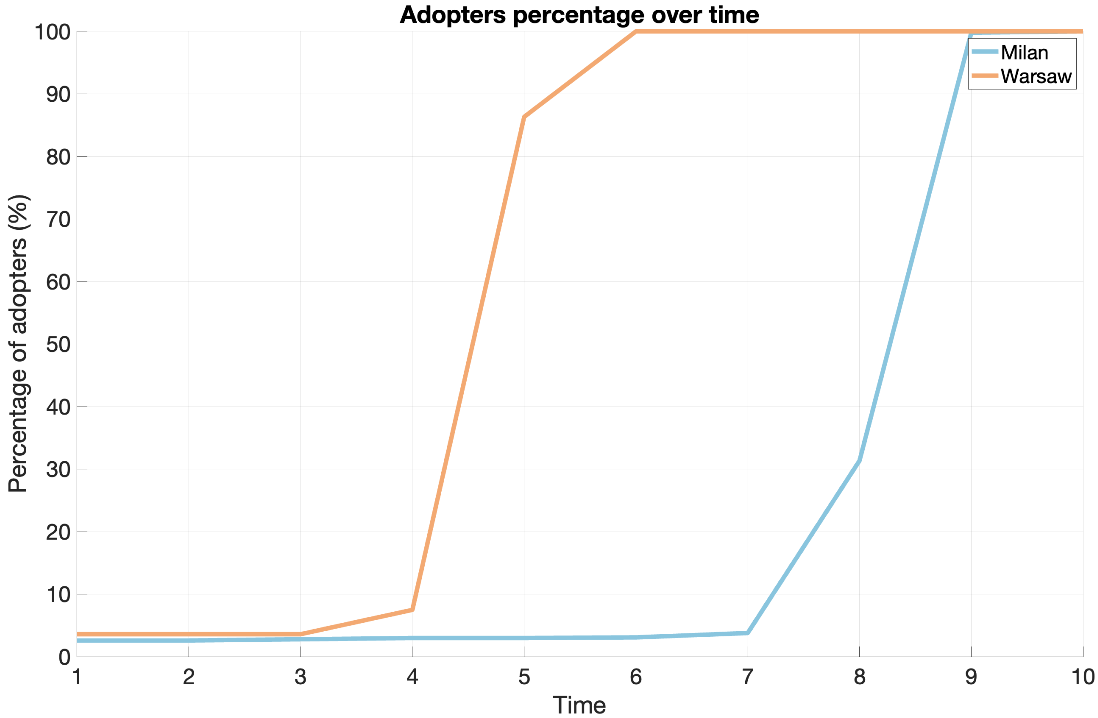
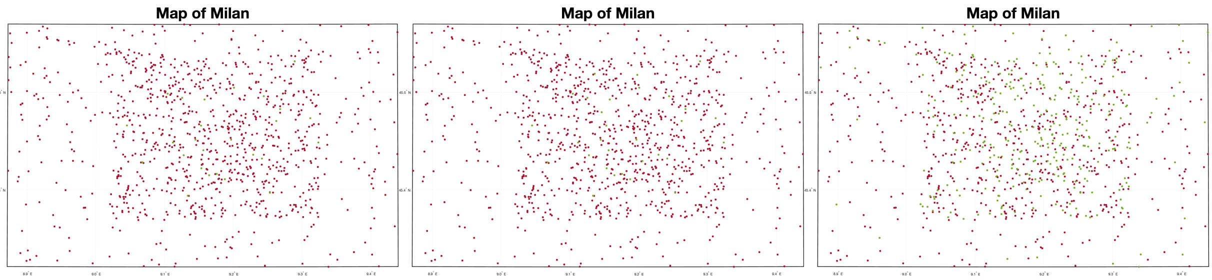
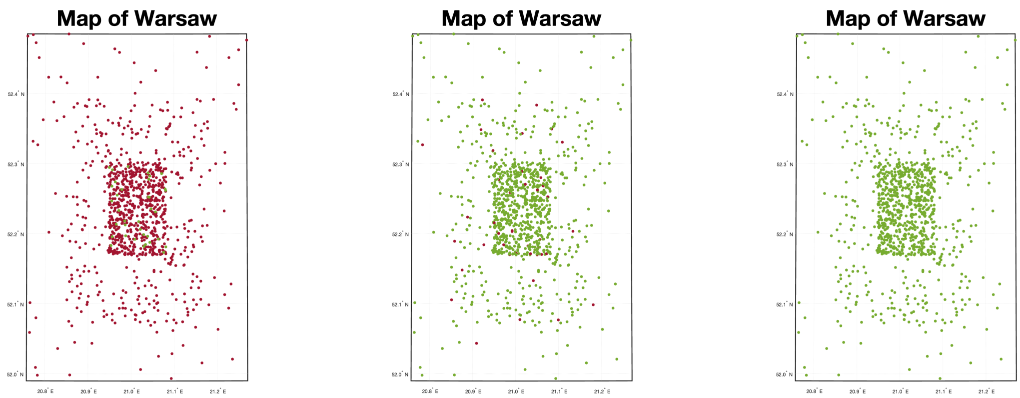


Figure 5.8: Milan's and Warsaw's adopters over time, Scenario B. Both for Milan's network and for Warsaw's network the number of adopters reaches the 100%, respectively at t=9 and at t=6.

Meanwhile, if we consider the weights of the seven indices as in Table 4.1 both of the networks reach 100% of the adopters. However, differently from Scenario A, in this case, the full adoption is achieved at different time steps. In fact, Warsaw gets to full adoption after 18 months (t=3), while Milan only after 27 months (t=9) (see Fig. 5.8). Also in this case the spread of the adoption (see Fig. 5.9) is followed by the increase of the Sharing index in Fig. 5.10 and also by the expansion of the average Sharing-DNA of the three classes (see Fig. 5.11).

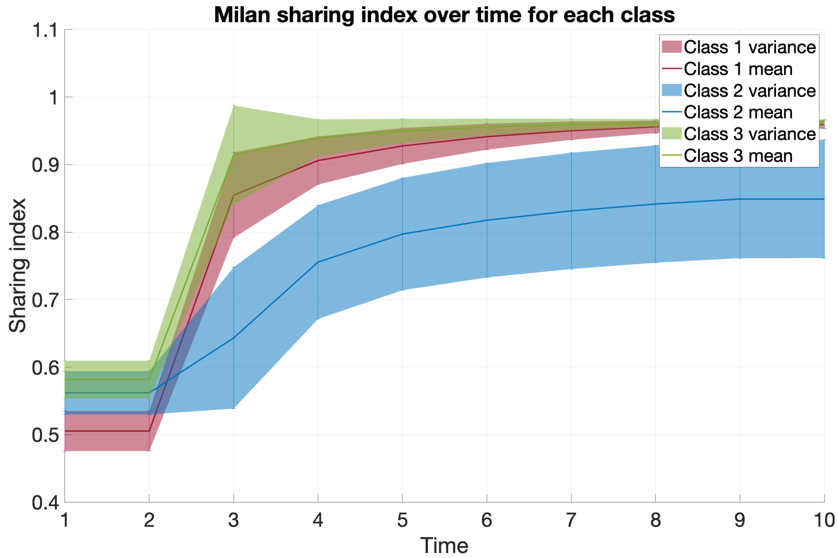


(a) t=1 (b) t=5 (c) t=8

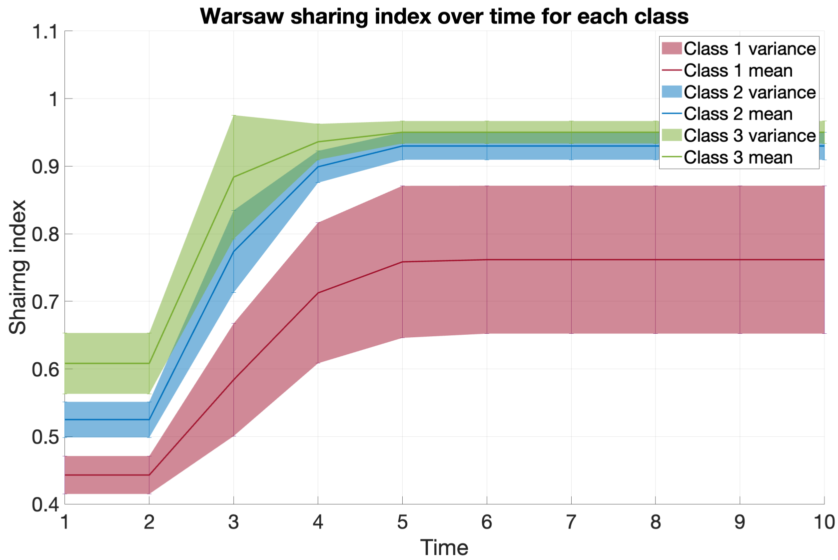


(d) t=1 (e) t=5 (f) t=8

Figure 5.9: Evolution of the Warsaw's and Milan' network over time, Scenario B

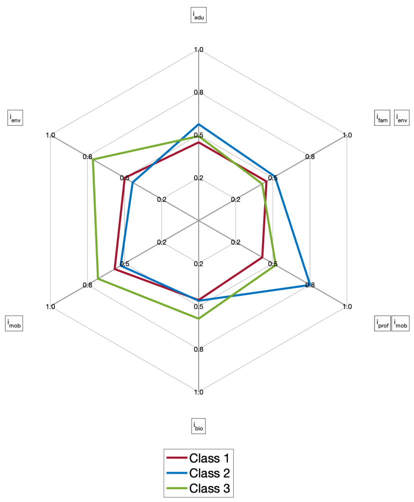


(a) Milan's Sharing index over time per classes, Scenario B.

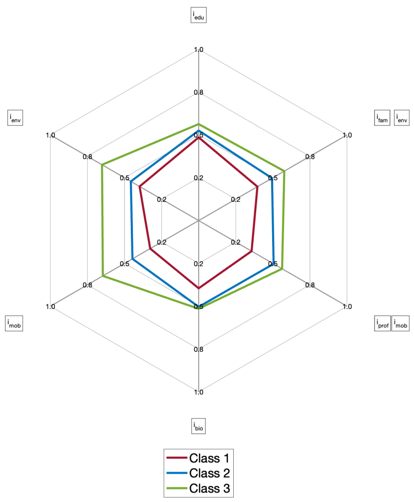


(b) Warsaw's Sharing index over time per classes, Scenario B.

Figure 5.10: Both for Milan's network (Milan's and Warsaw's Sharing index over time per classes, Scenario B.



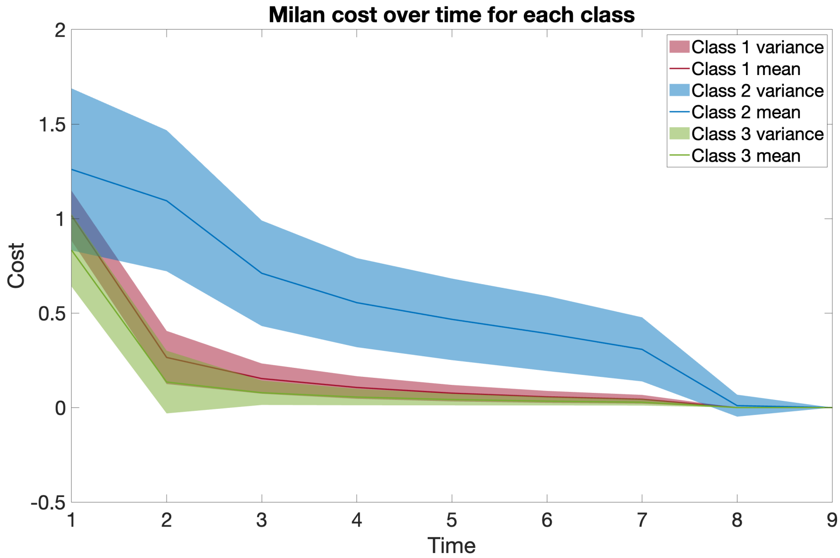
(a) t=1, Milan (b) t=5, Milan (c) t=8, Milan



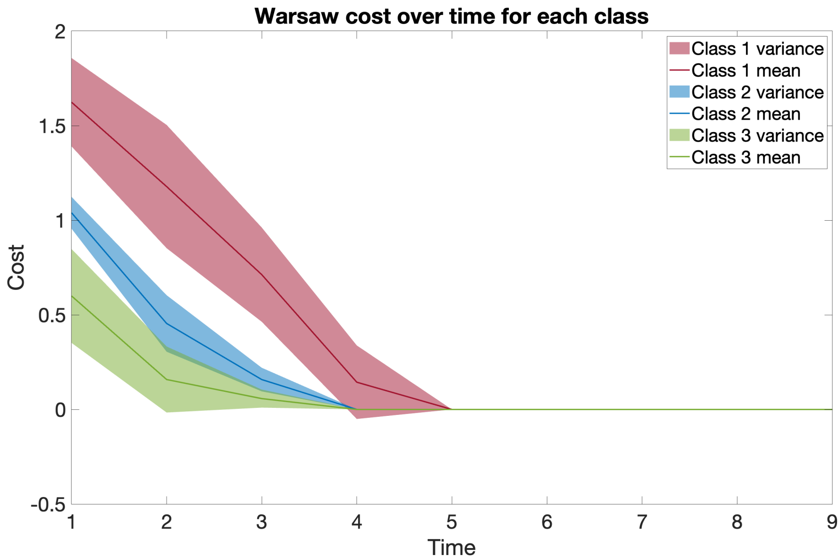
(d) t=1, Warsaw (e) t=5, Warsaw (f) t=8, Warsaw

Figure 5.11: Sharing-DNA of the Warsaw's and Milan's network over time, Scenario B

As can be seen in Fig. 5.12, as in the previous scenario the cost allocated in the case of Warsaw are higher than those of Milan (see Fig. 5.12). In this scenario, concerning the allocation of the inputs we can also see how they get distributed with respect to the features. By seeing Fig. 5.13, we can notice that the features that receive the higher amount of resources are i_{mob} and i_{env}.

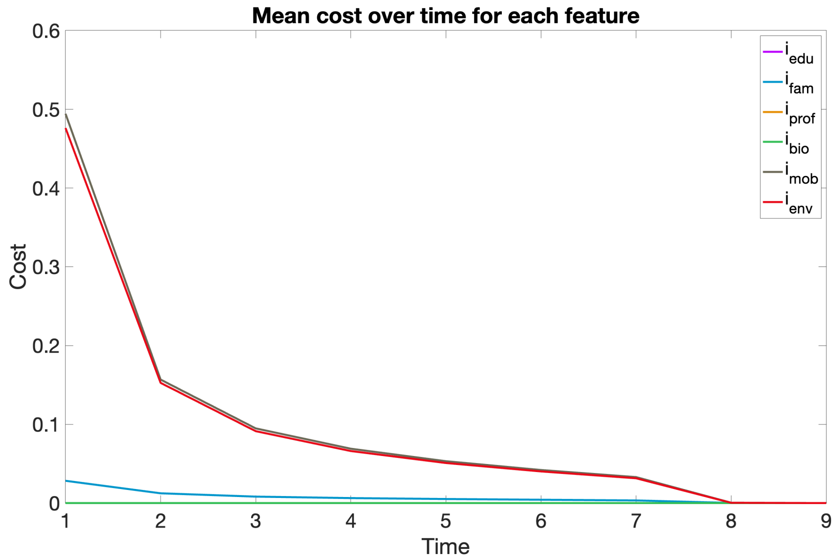


(a) Milan's cost over time per classes, Scenario B.

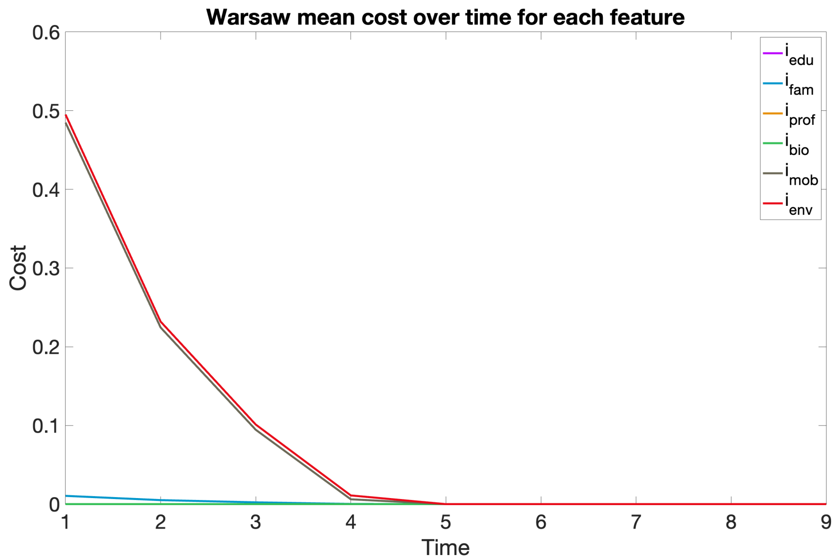


(b) Warsaw's cost over time per classes, Scenario B.

Figure 5.12: Milan's and Warsaw's cost over time per classes, Scenario B.



(a) Milan's mean cost for each feature over time, Scenario B.



(b) Warsaw's mean cost for each feature over time, Scenario B.

Figure 5.13: Milan's and Warsaw's mean cost for each feature over time, Scenario B.

As in the first attempt, no non-adopter user became an adopter (see Fig. 5.7), we are safe to say that the thresholds are all higher than the values of the targets. As we have said in Section 4.2.2 this happens because changing only three features out of the seven DNA features does not allow it to reach a high enough Sharing index value. If, on the other hand, the weights in Table 4.1 are considered, the increasing of the changeable features results enough to trigger the spread of adoption and to reach the full adoption (see Fig. 5.8). For this reason, from now on, we will use the proposed new weights.

As can be seen in the way that the evolution spreads (see Fig. 5.9), the two networks' evolution is not synchronized. In fact, in t=5 the percentage of adoption is very different between Milan's (5.9b) and Warsaw's (5.9e) situation. In addition, using these policy schemes the two networks reach full adoption at different moments in time. Regarding Milan's network in fact, the adoption gets to 100% only after 27 months (t=9) while for Warsaw, the full adoption is reached already after 1 year and a half (t=6) (see Fig. 5.8). The reason why this difference only occurs now can be found in the different index i_{geo} that characterizes the two cities: for Milan, this value is 0.6653 while for Warsaw, it is a little bigger, i.e. 0.7418. Now that, while still tracking the sharing index, the policy only acts on the three changeable features of the DNA and not anymore on the sharing index, the Milan's values of the i_{sha} are inevitably smaller than those of Warsaw. Therefore, for Milan's users, it is harder to reach the adoption thresholds. This can also be noted by the fact that Warsaw spreading of adoption starts significantly earlier (t=2) than those of Milan (t=7) (see Fig. 5.8).

As in the previous scenario, the evolution of the network is matched by an increase in the Sharing index. From the differences that we have found also in Scenario A, we can now see from Warsaw's network that the Sharing indices of the Class 1 users increase, but remain in any case distinctly lower than those of the other classes (see Fig. 5.10b). Nevertheless, all the individuals in the network quickly become adopters. On the contrary, by seeing Fig. 5.10a we notice that the Sharing index of users in Milan has to reach a high value in order for the corresponding node to become an adopter. Therefore, we can say that in Warsaw the adoption of people less inclined to Car-sharing is driven particularly by the relative popularity of the technology, while in Milan it is mainly powered by the increase of the personal Sharing index.

In addition to the evolution of the Sharing index, in this scenario, we can also analyze how the Sharing-DNA changes over time. In Fig. 5.11 in both Milan and Warsaw it can be seen how the editable features on which the policy act the most are i_{env} and i_{mob}, while those related to the individual income (i_{fam}) only slightly increases. This happens consistently with the weight values in Table 4.1 where the weight of i_{fam} is significantly lower than those of i_{env} and i_{mob}. Consistently, these two features are also the ones that receive more inputs than the other editable feature, i_{fam}.

5.2.4. Scenario C with selected values of B

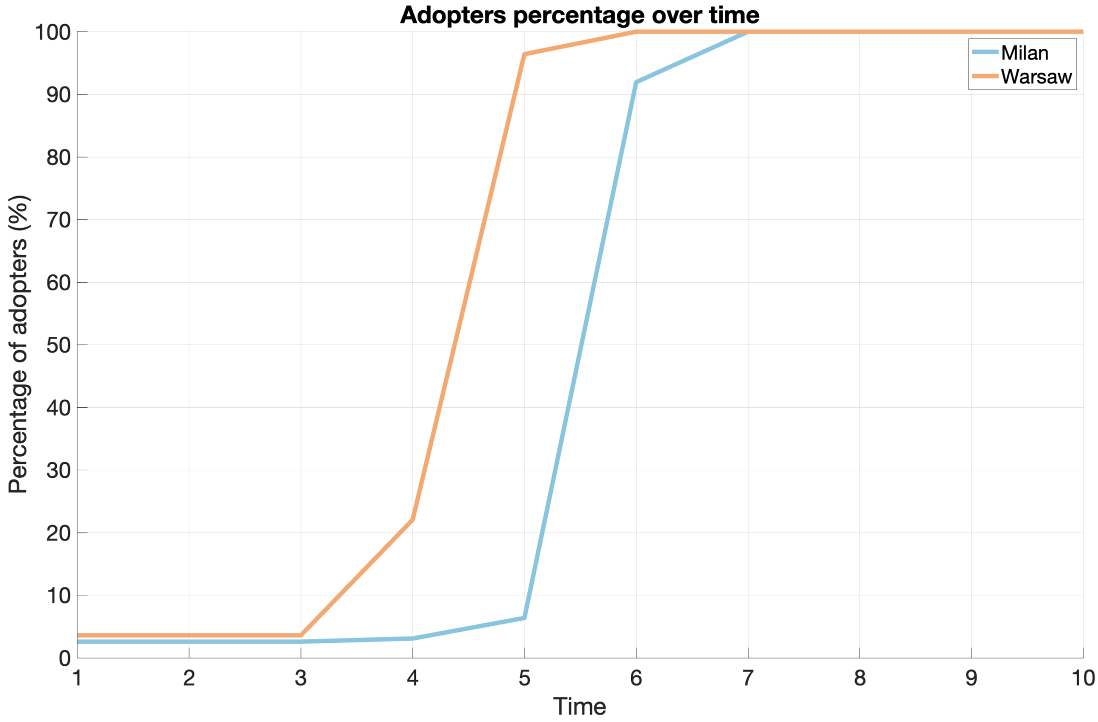
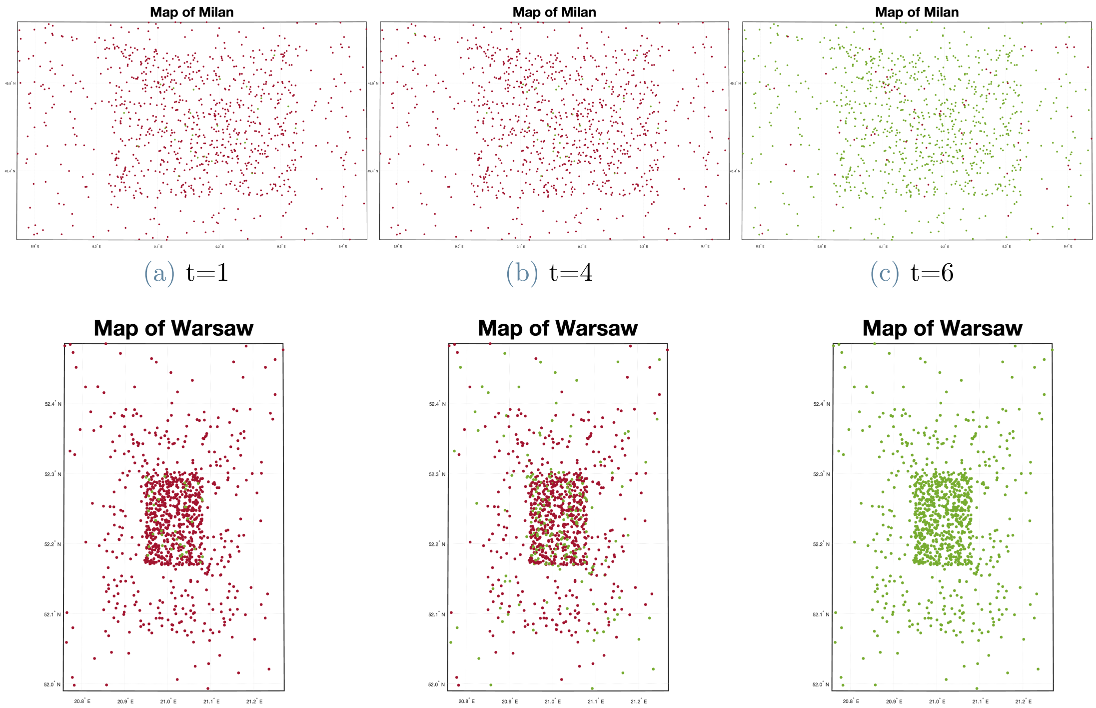


Figure 5.14: Milan's and Warsaw's adopters over time, Scenario C. Both for Milan's network and for Warsaw's network the number of adopters reaches the 100%, respectively at t=7 and at t=6.

The percentage evolution in Fig. 5.14 shows that both of the networks reach full adoption, for what concern Milan at time after 21 months (t=7) and for Warsaw at after 18 months (t=6).

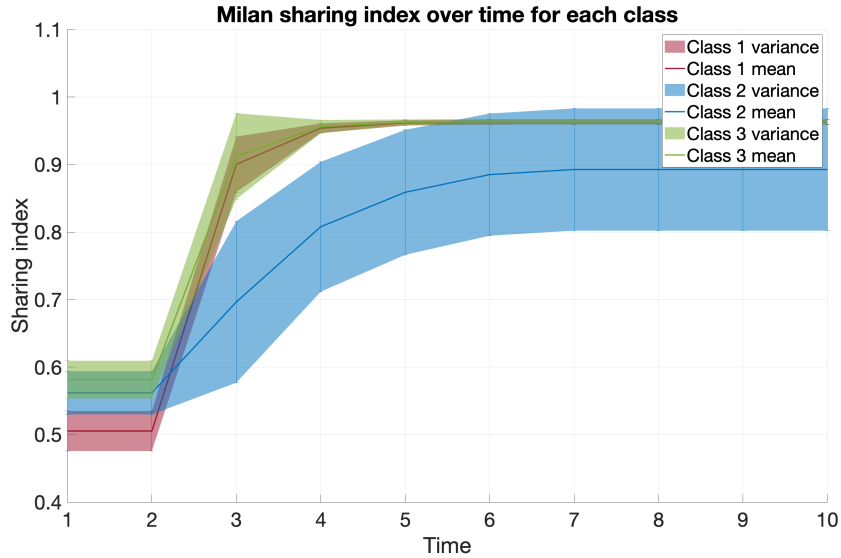


(a) t=1 (b) t=4 (c) t=6

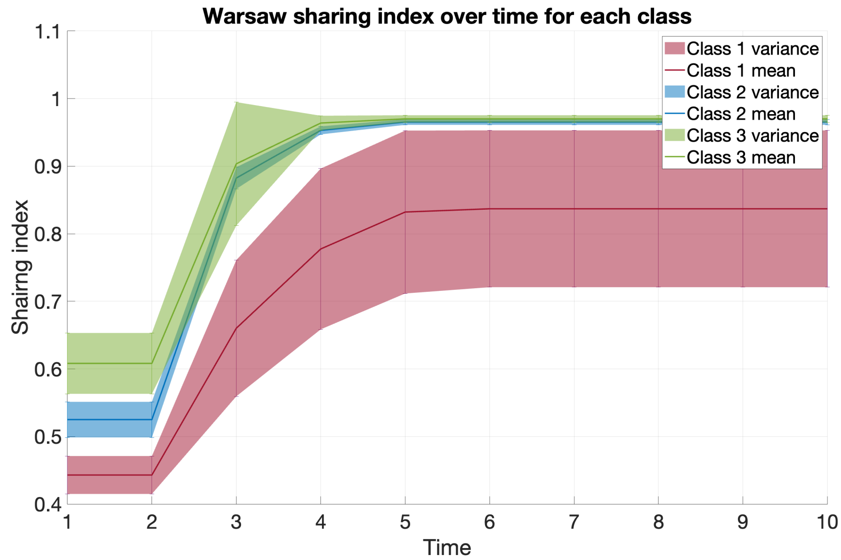
(d) t=1 (e) t=4 (f) t=6

Figure 5.15: Evolution of the Warsaw's and Milan's network over time, Scenario C

The overall outcomes of the policy schemes obtained here are overall very similar to those of the previous case (see Table 5.1 and Table 5.2). Also in this case the actual evolution of the networks shown in Fig. 5.15 are paired with the increase of the Sharing index (see Fig. 5.16) and the Sharing-DNA (see Fig. 5.17).

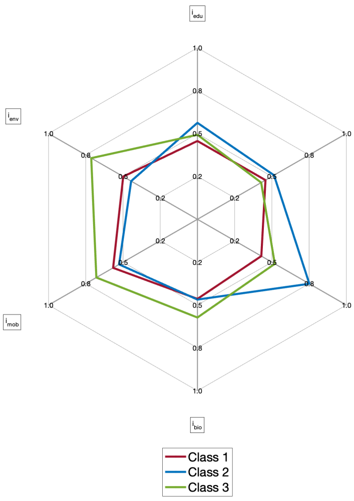


(a) Milan's Sharing index over time per classes, Scenario C.

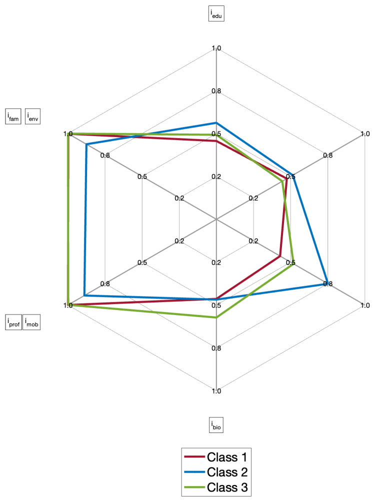


(b) Warsaw's Sharing index over time per classes, Scenario C.

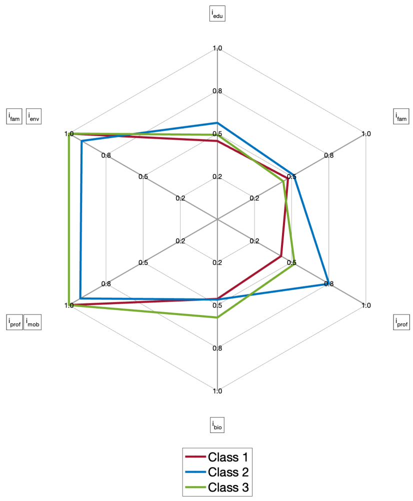
Figure 5.16: Milan's and Warsaw's Sharing index over time per classes, Scenario C.



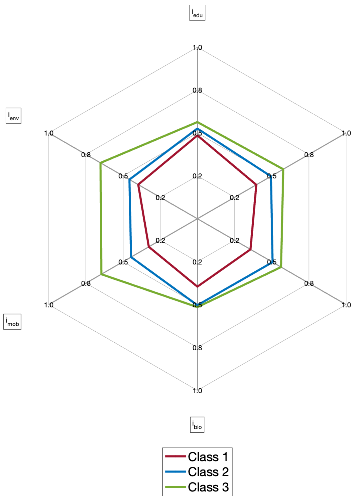
(a) t=1, Milan



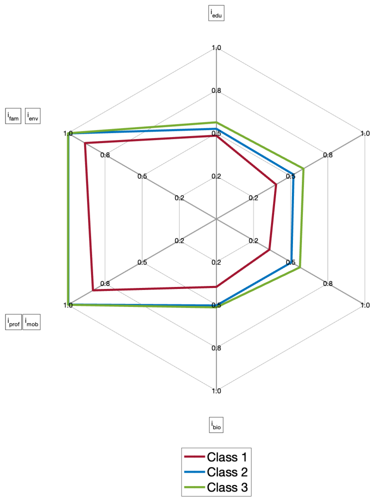
(b) t=4, Milan



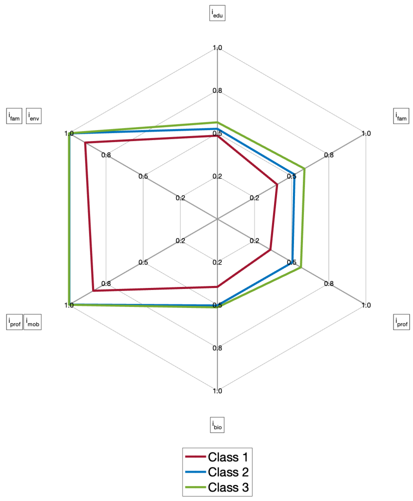
(c) t=6, Milan



(d) t=1, Warsaw



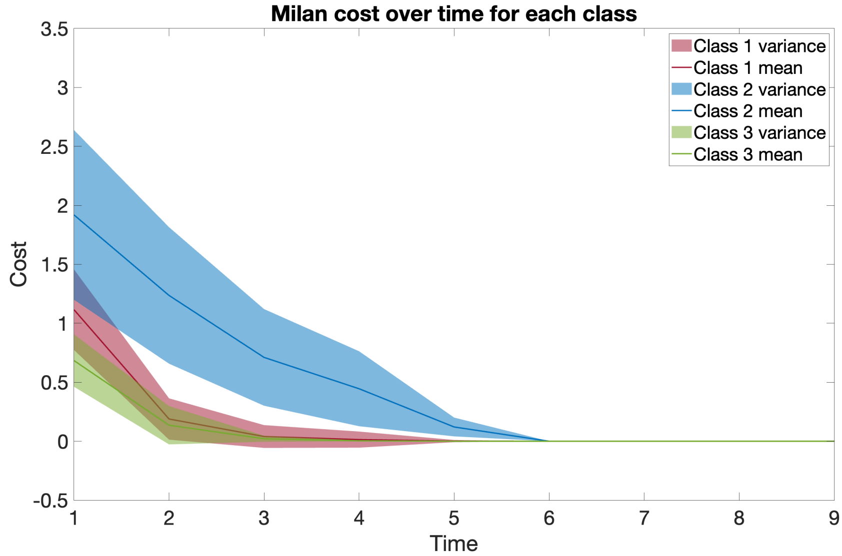
(e) t=4, Warsaw



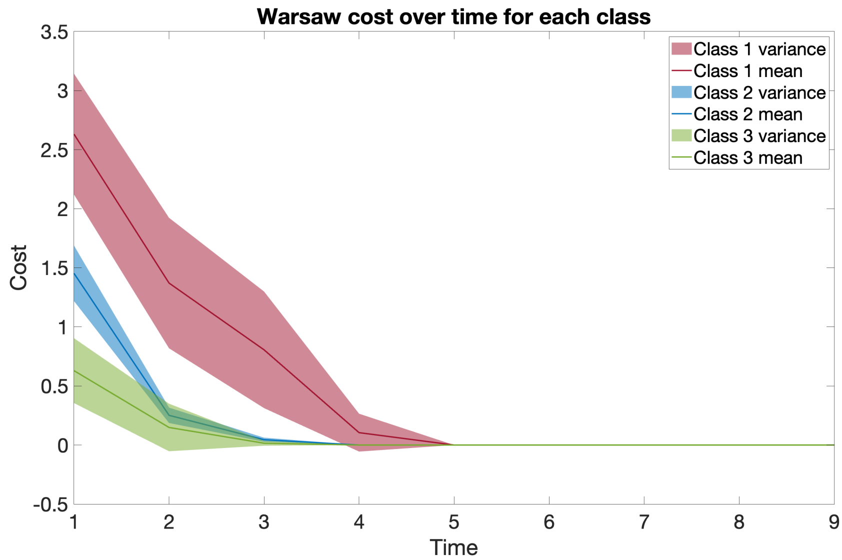
(f) t=6, Warsaw

Figure 5.17: Sharing-DNA of the Warsaw's and Milan' network over time, Scenario C

Again the cost over time for the Warsaw case is higher than those of Milan (see Fig. 5.18) and the features more resource-intensive are the i_{mob} and i_{env} especially in the Warsaw's case (see Fig. 5.19).

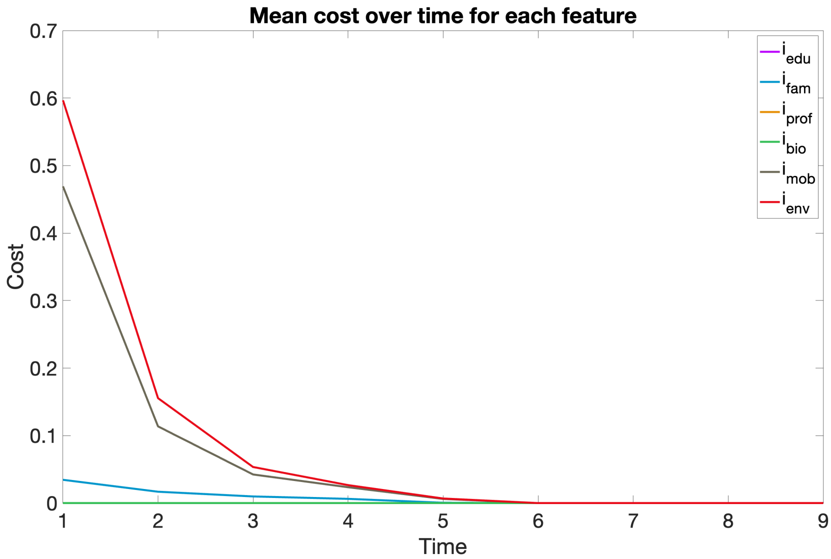


(a) Milan's cost over time per classes, Scenario C.

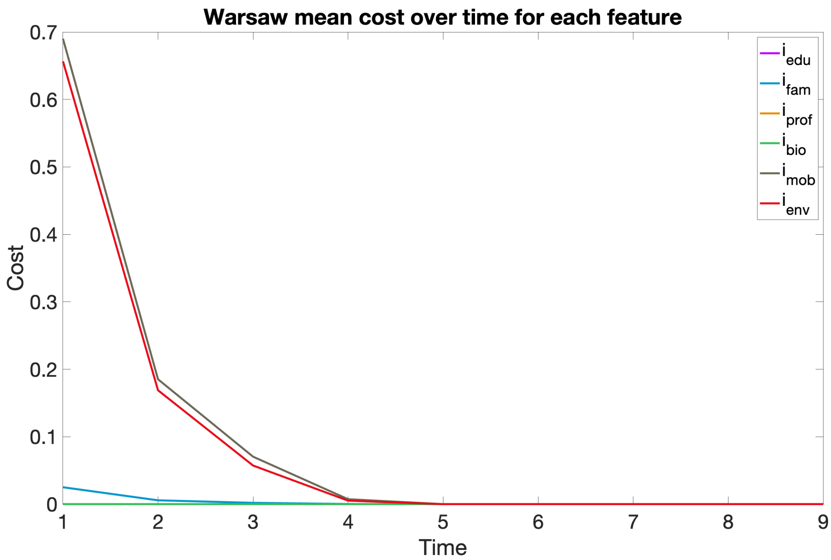


(b) Warsaw's cost over time per classes, Scenario C.

Figure 5.18: Milan's and Warsaw's cost over time per classes, Scenario C.



(a) Milan's mean cost for each feature over time, Scenario C.



(b) Warsaw's mean cost for each feature over time, Scenario C.

Figure 5.19: Milan's and Warsaw's mean cost for each feature over time, Scenario C.

With this third policy scheme, both adoption spreading of the networks (see Fig. 5.14) are faster than the previous ones (see Fig. 5.8). Moreover, the evolution of the two networks is in this case a little bit more comparable in time (see Fig. 5.15) and as both the beginning of the spread and the reaching of the full adoption of Milan are only one step later than those of Warsaw. Therefore we can say that tracking the error between the changeable features and their targets makes the difference between the i_{geo} of the two networks less relevant.

The Sharing index (see Fig. 5.16) and the changeable DNA features (see Fig. 5.17) are in general higher than the previous case at the end of the time horizon and as in the Scenario B in Warsaw the users less inclined to Car-sharing (Class 1) have clearly lower Sharing indices than every other user in the network (see Fig. 5.16b) and as shown in Table 5.2 they also receive the greater amount of inputs (see Fig. 5.18b).

Table 5.1: Milan: Scenario A vs Scenario B vs Scenario C

	B_v	t_{100}	C_{tot}	$\bar{C}$	$\bar{C}_1$	$\bar{C}_2$	$\bar{C}_3$
Scenario A	rand.	6	$1.1594 \cdot 10^3$	115.9368	10.1332	9.5533	8.5737
	cl.	6	750.8355	75.0836	7.3609	6.6674	3.8901
Scenario B	cl.	9	$1.3452 \cdot 10^3$	134.5192	13.7092	33.2454	9.8950
Scenario C	cl.	7	$1.5696 \cdot 10^3$	156.9600	1.5618	12.2738	7.7756

B_v is how the values are selected either in a random way (rand.) or with respect to the class of belonging (cl.), t_{100} is the time needed to the 100% adoption, C_{tot} is the total final cost, $\bar{C}$ is the mean allocated input per time step, $\bar{C}_1$ is the mean allocated input in Class 1, $\bar{C}_2$ is the mean allocated input in Class 2 and $\bar{C}_3$ is the mean allocated input in Class 3.

Table 5.2: Warsaw: Scenario A vs Scenario B vs Scenario C

	B_v	t_{100}	C_{tot}	$\bar{C}$	$\bar{C}_1$	$\bar{C}_2$	$\bar{C}_3$
Scenario A	rand.	6	$1.1989 \cdot 10^3$	119.8893	11.7748	9.5533	9.0202
	cl.	6	874.6458	87.4646	13.4338	7.8215	3.8145
Scenario B	cl.	6	$1.6709 \cdot 10^3$	167.0935	29.8930	14.0984	7.0666
Scenario C	cl.	6	$1.8772 \cdot 10^3$	156.9609	39.6684	15.3097	6.9843

B_v is how the values are selected either in a random way (rand.) or with respect to the class of belonging (cl.), t_{100} is the time needed to the 100% adoption, C_{tot} is the total final cost, $\bar{C}$ is the mean allocated input per time step, $\bar{C}_1$ is the mean allocated input in Class 1, $\bar{C}_2$ is the mean allocated input in Class 2 and $\bar{C}_3$ is the mean allocated input in Class 3.

5.3. Closed-loop Warsaw network evolution: fair vs. unfair

The aim of the fair LQR that we have discussed in Section 4.3 is to allow the agents to achieve their individual control goals while accounting for the fairness of the chosen control action. To understand if such an aim has been reached, it is thus fundamental to compare the different outcomes obtained with the two formulations of the control problem. In order to do that we will not only analyze the network evolution and the allocation of the resources as in the previous sections, but we will also compare the different dynamics of the above defines Index of equity $\mathcal{I}_e$.

Let us point out that in order for the formulations (Scenario C namely unfair LQR, fair LQR with W=1 and fair LQR with W=10) to be comparable we have to implement all the cases using the smaller network presented in Section 4.3.2.

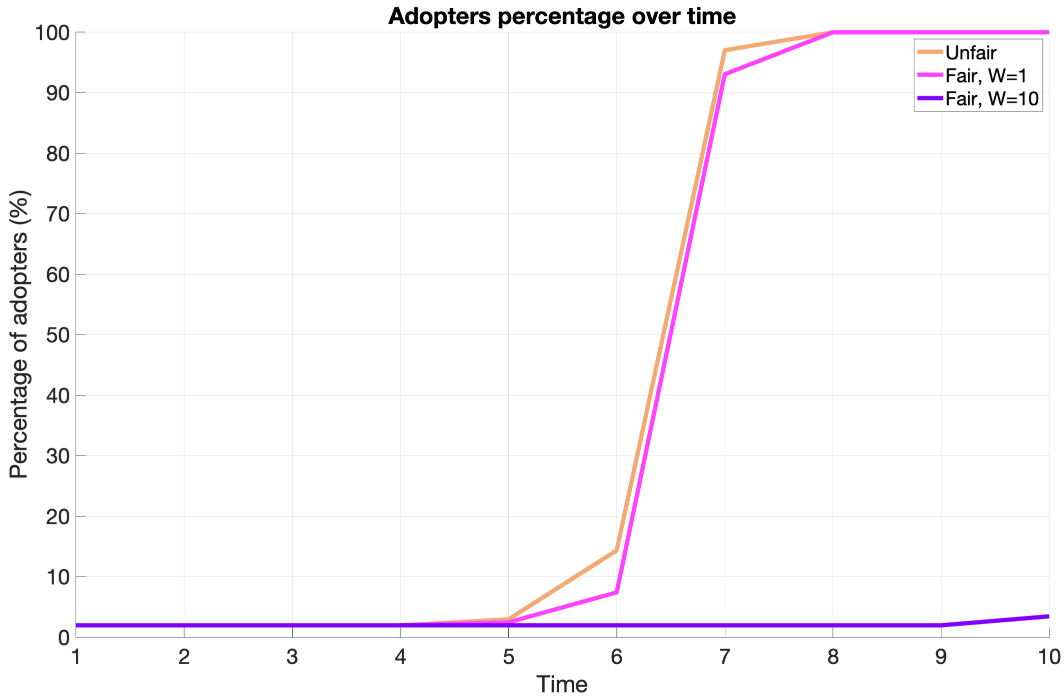
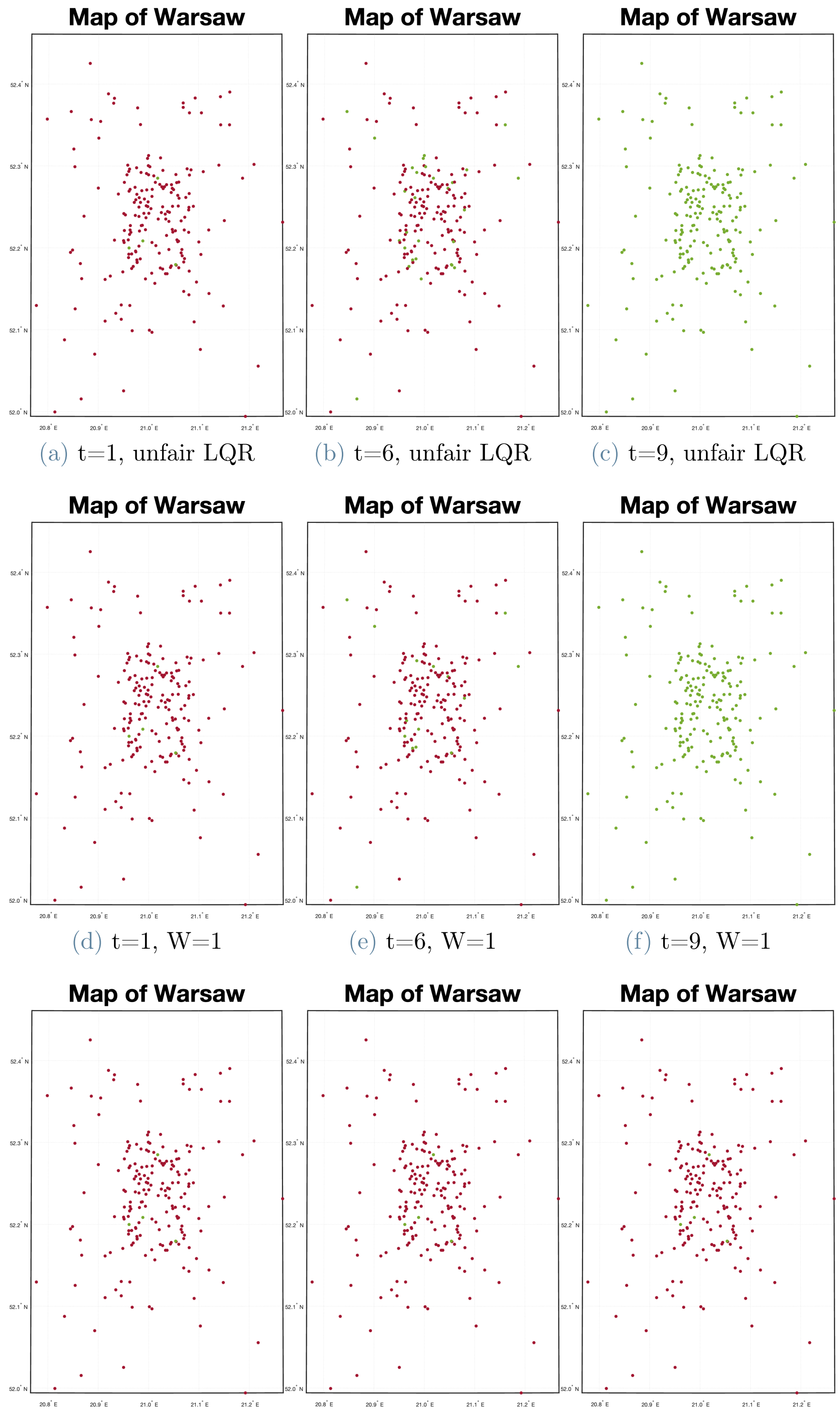


Figure 5.20: Warsaw's adopters over time, unfair vs. fair LQR.

As can be seen in Fig. 5.20 only the case with the value of W set to 10 doesn't reach the 100% of adopters. On the other hand, with regard to the Scenario C case and the fair LQR with W=1 they both reach the full adoption at t=8. For all of the three cases taken under consideration, the Sharing indices and the Sharing-DNA rise over time as shown in Fig. 5.22 and Fig. 5.23 respectively.

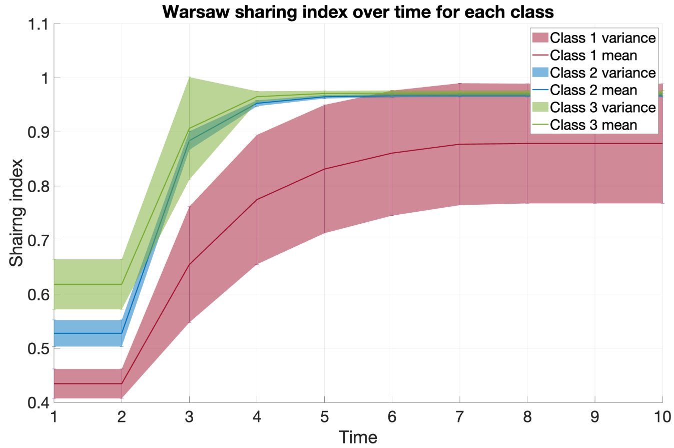


(a) t=1, unfair LQR (b) t=6, unfair LQR (c) t=9, unfair LQR

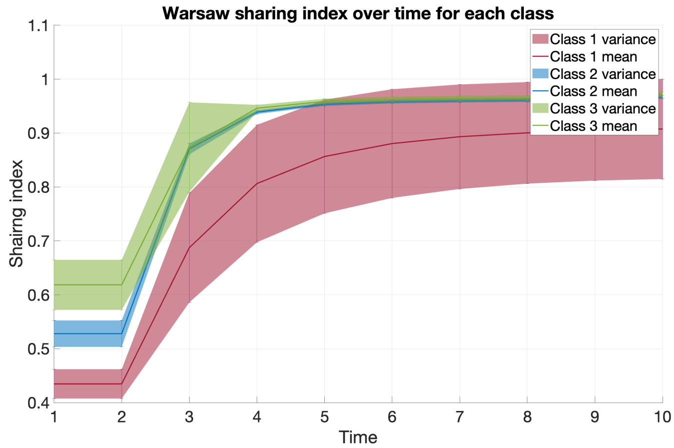
(d) t=1, W=1 (e) t=6, W=1 (f) t=9, W=1

(g) t=1, W=10 (h) t=6, W=10 (i) t=9, W=10

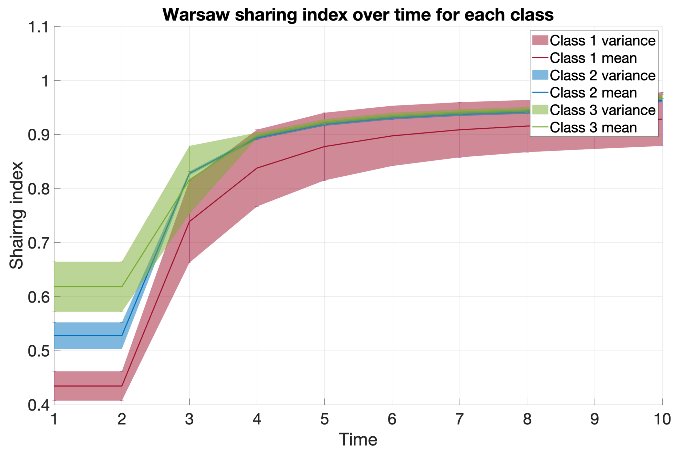
Figure 5.21: Evolution of the Warsaw's network over time, unfair LQR vs fair LQR



(a) Warsaw's Sharing index over time per classes, unfair LQR.

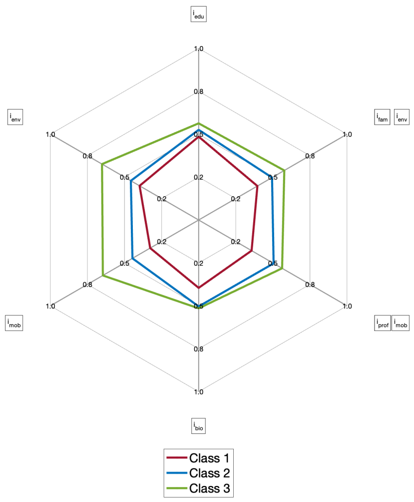


(b) Warsaw's Sharing index over time per classes, fair LQR with W=1.

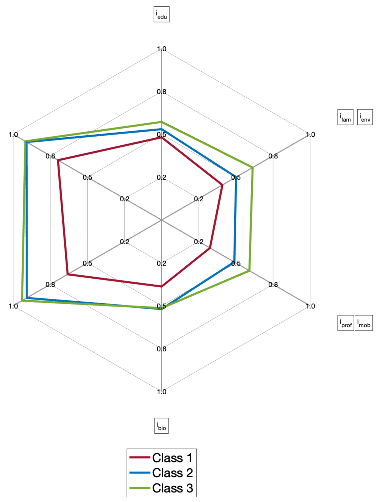


(c) Warsaw's Sharing index over time per classes, fair LQR with W=10.

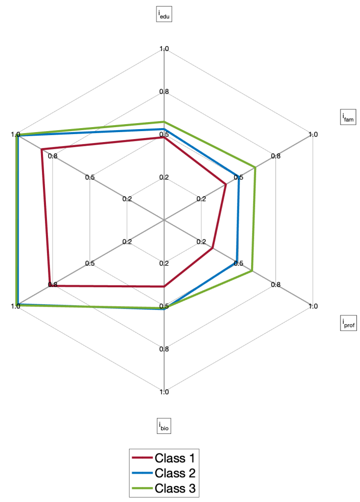
Figure 5.22: Warsaw's Sharing index over time per classes, unfair LQR vs fair LQR.



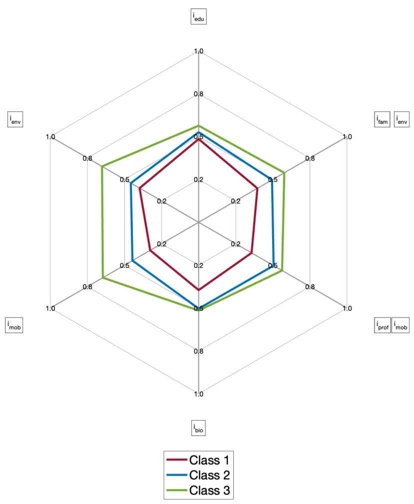
(a) t=1, unfair LQR



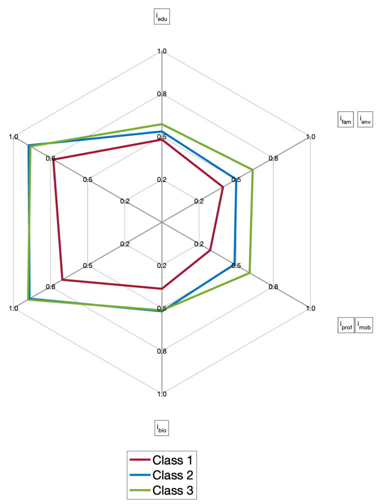
(b) t=2, unfair LQR



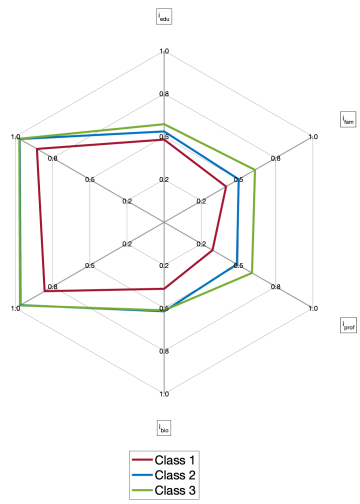
(c) t=3, unfair LQR



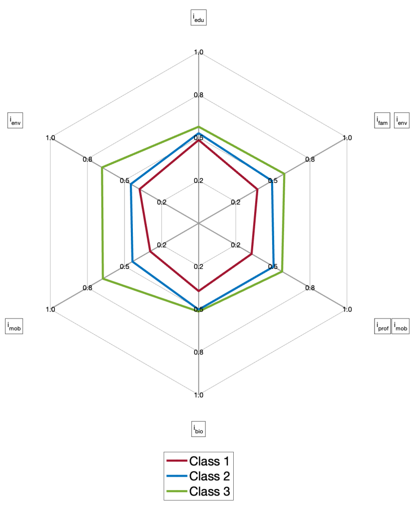
(d) t=1, fair LQR with W=1



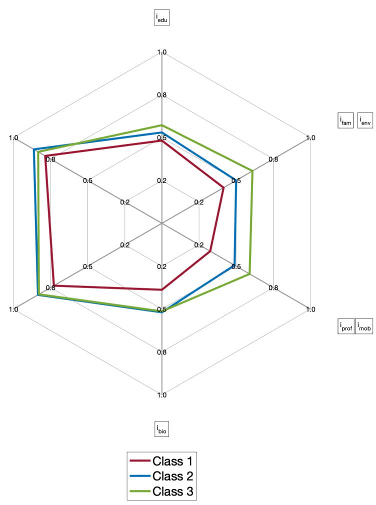
(e) t=2, fair LQR with W=1



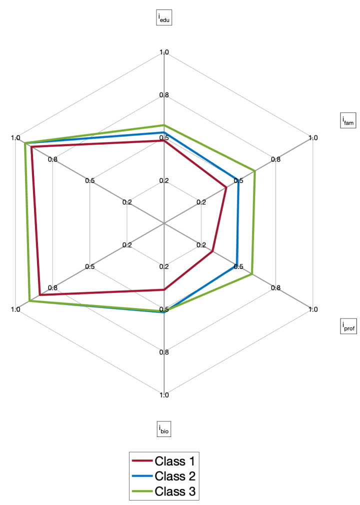
(f) t=3, fair LQR with W=1



(g) t=1, fair LQR with W=10



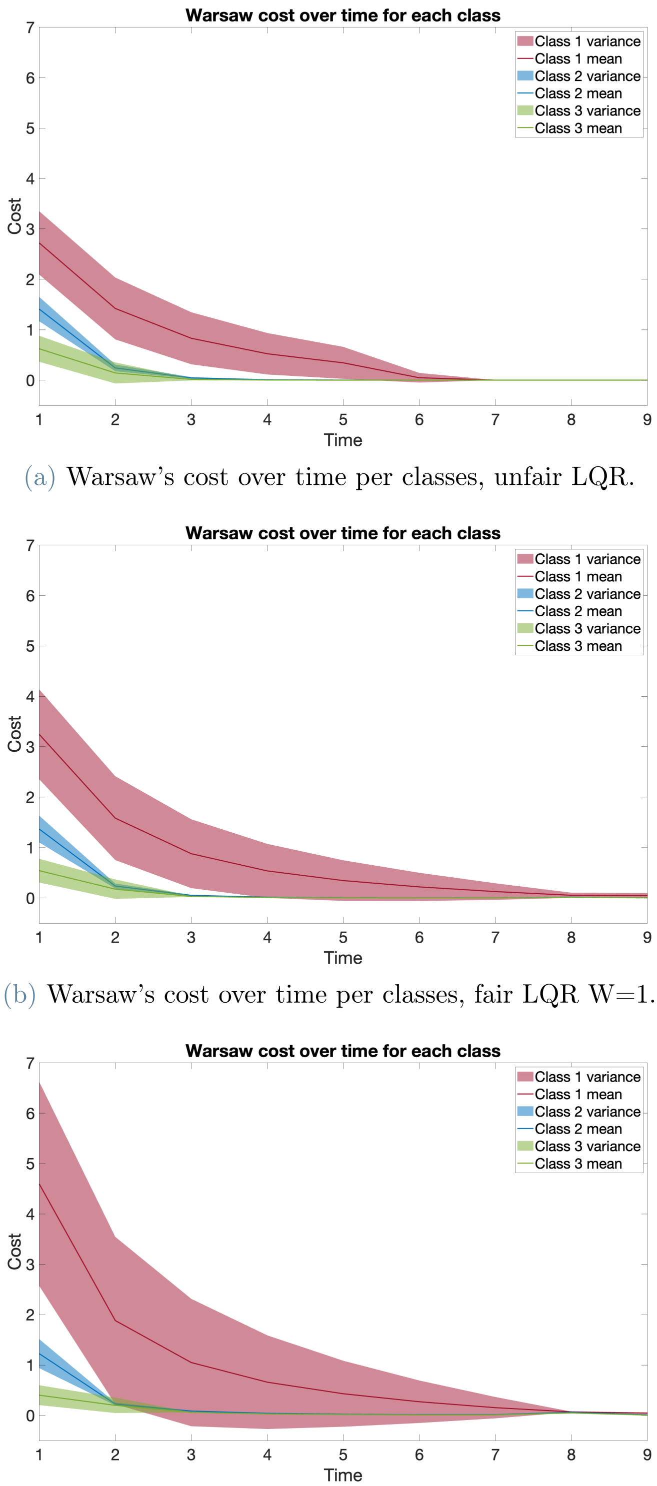
(h) t=2, fair LQR with W=10



(i) t=3, fair LQR with W=10

Figure 5.23: Sharing-DNA of the Warsaw's network over time, unfair LQR vs fair LQR

Meanwhile, regarding the allocation of the resources we can see in Fig. 5.24 that in the case of the fair LQR with W=10, the cost is much higher than the first two cases.



(a) Warsaw's cost over time per classes, unfair LQR.

(b) Warsaw's cost over time per classes, fair LQR W=1.

(c) Warsaw's cost over time per classes, fair LQR W=10.

Figure 5.24: Warsaw's cost over time per classes, unfair LQR vs fair LQR.

Finally, Fig. 5.25 shows how the Index of equity varies over time for the three cases and the values for the case of the unfair LQR are at every time step lower than those of the two cases with the fair LQR.

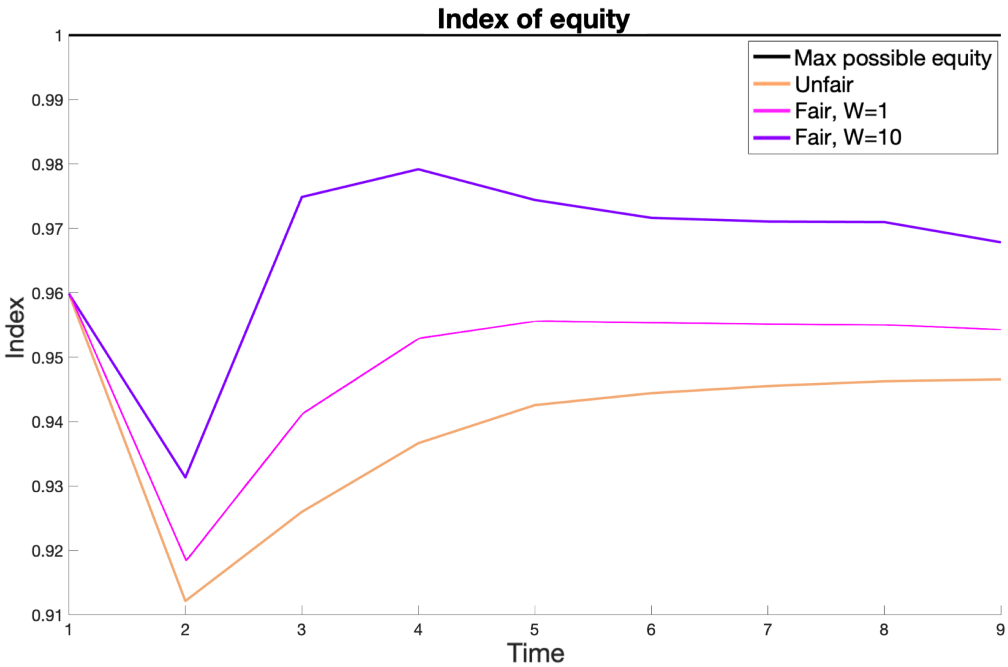


Figure 5.25: Index of equity, unfair LQR vs fair LQR, W=1 vs fair LQR, W=10

In the outcome shown in this section, we can clearly see how the three reported cases behave differently. First of all, as it can be seen in the percentage evolution (see Fig. 5.20) and in the actual spread of the adoption in the map of Warsaw (see Fig. 5.21), the evolution of the network gets slower as the value of W gets higher. This allows us to make a first note of the consequences of this new element, namely the greater it is the importance that we want to give to equity, the more the evolution speed of the network suffers.

Another important thing to note is the differences in how the values of the Sharing index and the Sharing-DNA increase (see Fig. 5.22 and Fig. 5.23). By looking at the Sharing indices dynamics in fact, as the weight W gets higher, the users' values of the Sharing index over time approach each other more. In the case with W=10 the Class 1 values of i_{sha} (see Fig. 5.22c) increase more and get closer to those of Class 3. This is done, however, at the expense of the latter, i.e. the users more inclined to Car-sharing services. In fact, the i_{sha} of Class 3 is in average lower in this last case than in the unfair case (see Fig. 5.22a). This behaviour is the same as what concern the dynamics of the Sharing-DNA. To better analyze the situation in this case we want to take into account the DNA evolution over the first three time steps as it is when the features increase the most (see

Fig. 5.23). It can be observed that the policy scheme in the case of W=10 tends to increase more the Class 1 indices than the other two policy schemes do. This happens again at the expense of the other two classes' values that are in Scenario C and in the case with W=1 higher with respect to the third case.

As regards the allocation of the incentives in Fig. 5.24, the latter is consistent with the behaviour of the Sahring index and the Sharing-DNA. As a matter of fact, the class 1 amount of inputs gets higher with the increase of the importance of the equity factor.

Finally, the analyses carried out during this section are confirmed by the dynamics of the Index of equity (see Fig. 5.25) since it can be shown that the values that are closest to the maximum value 1 are those of the fair LQR with W=10.

6 | Conclusions and future developments

In this thesis, we have introduced a network-based framework to analyze the adoption process of Car-sharing services and to design policies that will help to boost the widespread of this technology. We have used European socio-economic data to characterise how each person felt about using shared mobility services. We propose the simple yet comprehensive tool of the Sharing-DNA to have an easily understandable characterization of each person and to serve as an operational tool for the development of human-centered policies geared towards the promotion of the use of Car-sharing services.

After proposing different control formulations that differ in how the policy schemes act, we have seen that what influences the adoption is the composition of the network along with the characteristics of the users that belongs to it. We can claim that in Warsaw the network evolution is always a little faster than that of Milan as it has more diverse clusters of people. On the other hand, the costs allocated in the Milan case are always lower precisely for the presence of a group of very little inclined individuals in the network of Warsaw. Moreover, we can say that the policy schemes that act on the Sharing index require a lower allocation of resources and result in a faster evolution of the network, but considering the scenarios that enable to act directly on the editable features allows for more intuitive and realistic policy schemes.

We have then provided a preliminary attempt to explicitly account for fairness in the design process. To this end, we have formulated a fair LQR that takes individual performance into consideration while fostering equity among a group of agents. In comparison with the unfair formulation of the LQR problem, we have highlighted how fairness affects both individual performance and the general allocation of the control resources available. In general, it can be said that stronger agents in the group suffer from a little loss in tracking performance, however, the group as a whole exhibits a more balanced behaviour at the end of the time-frame taken into consideration. As a result, hence, introducing a concept of justice in the process of policy design makes it possible to reduce the differences between different types of agents promoting users to be comparably close to their individual targets.

Nevertheless, the research conducted has a significant limitation. In fact, since this work represents only an initial stage of the problem, in the process of policy design, we have not considered any kind of constraint regarding the available resources. In the attempt to minimize the tracking error, this lack of restriction enables the control action to allocate unbounded inputs. As a result, the carried-out simulations turn out to be unrealistic since the resources may not always be available or feasible to implement in real-world applications. Therefore, precisely because we have seen that taking into account a concept of justice during the process of policy formation makes the procedure more resource-intensive, future work needs to examine the role of the limitation of resources in relation to the concept of fairness in such a way that the developed policies are more robust and realistic. To do this, we aim at exploiting the MPC framework for policy design.

Bibliography

[1] A. Altmann, L. Tolos, O. Sander, and T. Lengauer. Permutation importance: a corrected feature importance measure. *Bioinformatics*, 2010.

[2] E. Amirnazmiafshar and M. Diana. A review of the socio-demographic characteristics affecting the demand for different car-sharing operational schemes. *Transportation Research Part D: Transport and Environment*, 14, 2022.

[3] D. Bertsekas. *Dynamic Programming and Optimal Control. Volume I.* Athena Scientific, 1995.

[4] V. Breschi, C. Ravazzi, S. Strada, F. Dabbene, and M. Tanelli. Fostering the mass adoption of electric vehicles: A network-based approach. *IEEE Transactions on Control of Network Systems (Volume: 9, Issue: 4, December 2022)*, pages 1666 – 1678, 2022.

[5] M. . Company. What is diversity, equity, and inclusion?, 2022. URL `https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-diversity-equity-and-inclusion`.

[6] I. Dabbura. K-means clustering: Algorithm, applications, evaluation methods, and drawbacks, 2018. URL `https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03`

[7] A. De Marchi and M. Gerdts. Free finite horizon lqr: A bilevel perspective and its application to model predictive control. *Automatica*, 100:299–311, 2019.

[8] E. E. A. (EEA). Greenhouse gas emissions from transport in europe, 2022. URL `https://www.eea.europa.eu/ims/greenhouse-gas-emissions-from-transport`.

[9] D. Efthymiou, C. Antoniou, and P. Waddell. Factors affecting the adoption of vehicle sharing systems by young drivers. *Transport Policy*, 29:64–73, 2013.

[10] A. Fernandez Peralta, J. Kertész, and G. Iñiguez. Opinion dynamics in social networks: From models to data. 2022.

[11] D. Fiorello and L. Zani. Publications office of the european union. *EU Survey on issues related to transport and mobility*, 2015.

[12] S. Fluchs. The diffusion of electric mobility in the european union and beyond. *Transportation Research Part D: Transport and Environment*, 86, 2020.

[13] J. P. Gleeson. High-accuracy approximation of binary-state dynamics on networks. *Phys. Rev. Lett.*, 2011.

[14] J. Henrich, S. J. Heine, and A. Norenzayan. The weirdest people in the world? *Behavioral and Brain Sciences*, 2010.

[15] F. L. Lewis, D. L. Vrabie, and V. L. Syrmos. Optimal control. 2012.

[16] Y. Li and Z. Xu. A bibliometric analysis and basic model introduction of opinion dynamics. *Applied Intelligence*, 2022.

[17] J. Maeda. Design in tech report 2019, 2019. URL `https://designintech.report/2019/03/09/design-in-tech-report-2019/`.

[18] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54:1–35, 2021.

[19] A. Natekin and A. Knoll. Gradient boosting machines, a tutorial. *International journal of computer applications*, 2017.

[20] M. I. S. of Public Health. Equity vs. equality: What's the difference?, 2020. URL `https://onlinepublichealth.gwu.edu/resources/equity-vs-equality/`.

[21] A. Ortega Hortelano, A. Tsakalidis, A.and Haq, K. Gkoumas, M. Stepniak, F. Marques Dos Santos, M. Grosso, and F. Pekar. Research and innovation in car sharing in europe. 2022.

[22] Ramos, C. Bergstad, and A. Chicco. Mobility styles and car sharing use in europe: attitudes, behaviours, motives and sustainability. *European Transport Research Review*, 2020.

[23] J. Rawls. *A Theory of Justice: Revised Edition.* Harvard University Press, 1999.

[24] N. A. Saxena, K. Huang, E. DeFilippis, G. Radanovic, D. C. Parkes, and Y. Liu. How do fairness definitions fare?: Examining public attitudes towards algorithmic definitions of fairness. *AIES '19: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 99–106, 2019.

[25] V. Singh, V. Singh, and S. Vaibhav. A review and simple meta-analysis of factors

influencing adoption of electric vehicles. *Transportation Research Part D: Transport and Environment*, 86, 2020.

[26] H. Xia, H. Wang, and Z. Xuan. Opinion dynamics: A multidisciplinary review and perspective on future research. *International Journal of Knowledge and Systems Science*, 2:72–91, 2011.

[27] Y. Yousefi. Notions of fairness in automated decision making: An interdisciplinary approach to open issues. *Kő, A., Francesconi, E., Kotsis, G., Tjoa, A.M., Khalil, I. (eds) Electronic Government and the Information Systems Perspective. EGOVIS 2022. Lecture Notes in Computer Science, vol 13429. Springer, Cham*, pages 3–17, 2022.

List of Figures

List of Tables

Acknowledgements

I am very grateful to my supervisor, Mara Tanelli, for the opportunity to work on this thesis project and for the guidance and support I received during the research activity. Additionally, I would like to express my gratitude to my co-advisors Eugenia Villa and Valentina Breschi for their outstanding assistance, support, and patience throughout the research and writing of this thesis. Thanks to their valuable experience and knowledge, they helped me overcome the difficulties during my thesis work and achieve my research goals.