Final report

Combinatorial optimization with Reinforcement learning

Digitalization of an epidemic

Yann Cauchepin

Engineer's internship

13.05.2021 - 31.08.2021

24.08.2021

Interactive Summary

Introduction

What is the subject of the internship ?

This internship aims to use reinforcement learning to optimize epidemic crisis management. This therefore includes designing the environment for the epidemic spread and applying reinforcement learning to its assessments. This optimization is multi-criteria since it aims to dissipate or limit infectious exchanges while stabilizing the economy of the territory studied.

This project focuses on Luxembourg society and is coded in Python language since there are many libraries already interesting for the development of the program.

For the sake of perfectionism, the development of the environment has taken time and many elements still remain to be implemented in the program. It was necessary to imagine everything without a guide, integrating the Sumo-Mobility tool, to have a realistic design, efficient and that can be adapted as and when needs and perspectives of evolution.

At the end of this internship, only a simple part of the environment is implemented. It would be necessary to have more time to make the environment more complex/realistic, adapt the simulated results with those of the real data, be able to determine actions of societal control and finally optimize these decisions according to several criteria and constraints.

The idea is to think step by step, perfect and make the environment more and more realistic with design testing. This is the very basis of optimization so it is the most essential !

A quick tour on the environment

The focus here is on epidemics that are transmitted by air. The design of the environment for this project is poorly adapted to other means of transmission such as sexually transmitted diseases such as AIDS.

The spread of infectious agents in society cannot be predicted a priori. It is necessary to wait for the transformation of the environment to characterize the parameters of an epidemic. An epidemic is identified only after discovering infected cases in society and making the link with the emergence of a new epidemic. The evolution of an airborne epidemic extends geographically in a rather chaotic way, with the dynamic entropy of carrier individuals in society. To model this environment, it is necessary to break down society by individuals and spaces/places of passage. Each individual circulates in society through different spaces in his daily life. Infectious exchanges take place with the proximity of healthy and carrier individuals.

Several scales can be distinguished in the representation of the environment. Very locally, it is too absurd to represent the spread of infectious agents within an individual's organism, especially considering the complexity and diversity of organisms. But we will see that we can have a very simplistic approach to the infectious functioning of an organism by considering the duality between antigens and antibodies. At the scale of a confined or semi-enclosed space, it is quite accessible to represent the infectious exchanges between individuals. Then, more generally, there are two scales that can be dissociated at the regional level and then at the national level.

With a view to societal optimization, this project centralizes its approach on a regional scale. We take a specific territory containing several spaces and synthesize the infectious exchanges in these different spaces. In a more developed version, the project can be extended to multi-territorial or national representation, where decisions may be different depending on the region. In parallel, we will also be interested in the mobility of individuals within the architecture of spaces. The outlook for developments is discussed in more detail at the end of this report.

What main tools are used ?

The starting point of the program is the description of a territory by an openstreetmap file that makes it possible to obtain the spaces of interest of a region as well as the tracks for the circulation of individuals. The description of these files is done by contributors and is relatively well detailed. It should be noted that the veracity of the data described or their updating are sometimes questionable. In our use for rich and popular parts of the world like Luxembourg, the files are far more than enough.

The second tool is SumoMobility, which is an implementation dedicated to civil engineering to digitize urban traffic with several distinct types of transport. The program can directly use an openstreepmap file of the input region to get the road network. It is a very complex and interesting tool to reproduce the mobility of a society. However, one of the difficulties is that it is very much oriented around trafficking and not individuals. It is therefore necessary to adapt our implementation. This tool has the advantage of being popular and can provide sufficient complexity to our program with the mobility and traffic constraints of the company. It should also be noted that there are research studies on traffic demands and traffic statistics that may be interesting to deepen the project. One example is Luxembourg Sumo Traffic Scenario (LuST) produced by Laura Codeca, Raphael Frank and Thomas Engel, from the University of Luxembourg.

These tools will be more detailed as the report progresses.

Similarly, we can note the use of a pre-existing library in Python for reinforcement learning called rlib. This implementation is very complete but will not be developed in this report due to lack of time. Of course, the project uses other standard Python libraries that are not like real tools.

Sophisticated approach

Before talking about multicriteria optimization, it is necessary to work on the design of the environment in detail. This is the basis of the project : the more the environment is perfected, the more applicable and realistic its optimization will be. It is from this point of view that my internship focused on the development of the environment. For the sake of perfectionism, it is essential to think carefully about the interlocking designs/ideas to be able to break down the implementation and anticipate its use for related topics or to foresee prospects for evolution.

As a result, there are five main layers in the environment : individuals, spaces, mobility, infectious agents and combinatorial decisions. Each of them will be detailed in the following parts.

In detail, we use cartographic and demographic data to initialize the territory we are studying. This phase is essential and it is absurd to determine an optimization on a little-known environment. This is why we will talk more about "digitalization" than simulation. More precisely, this phase will be called the static digitalization of society, which brings together individuals and spaces.

Secondly, we are concerned with the dynamic digitalization of individuals in different spaces. It is a complex phase. This is where we can find a parallel with a neural network. From an initial environment composed of individuals and spaces, it must be possible to evaluate the dynamics of society over a large set of days, each with different characteristics. It is like a black box where we find the individuals and spaces at the entrance, and the detailed occupations of the spaces at the exit.

So why not simply use a neural network to predict the dynamics of society ?

The very concept of a neuron is to assimilate several data, process them and summarize them into simple information. It's not intuitive about developing a complete dynamic assessment for each individual and each space. So a priori we could summarize mobility and be satisfied with more or less simple statistics. We can find the usefulness of a neural network which is to solve a goal of classification or prediction of a set of data without context. But if we manage to approach the reality of society and observe a real dynamic digitalization, we can have much more relevant and unmodified information. Our data processing is much more precise and leaves very little room for error. You don't make a pastry recipe by playing dice! This is the complex part of any project which is to work on the details of things. When done well, there is only a tiny uncertainty about the final result. With a view to multi-criteria optimization of evaluations, details about the environment are therefore more than essential.

To the question how to make dynamic digitalization reliable and realistic ?

This is the problem, and the challenge of environmental design. The part concerning the mobility of individuals in spaces and dealt with later in the report. We will see that we can find a form of training of this digitalization as we find in a neural network. Of course, this training is not mandatory here since we can already reproduce a totally random mobility and obtain a dynamic of individuals. This is also the current result of the internship but it must be taken into account that it is absurd. But this at least has the merit of representing the potential of this project.

The training could be carried out with traffic demands, observed data on the affluence of places for example. We may be interested in tracing individuals via network coverage but this poses ethical problems for the individual. As a reminder, the interest here is not to spy on or manipulate individuals via their personal data ; The project focuses on predicting the movements of a territory to optimize health risks and maximize economic stability. It must be benevolent for individuals and the population in general.

From there, it is interesting to dissociate digitalization from society and infectious spread. For benevolent side projects or even to simplify the maintenance of the implementation, it is interesting to be able to digitize a territory independently of an epidemic consideration.

The layer concerning infectious agents is an addition of digitalization. Of course, it brings specific characteristics to individuals, spaces and mobility. The interest is not to make these parameters blocking for a simple societal digitalization. It will be seen that we can also train the parameters of each infectious agent with observed data. However, as noted above, the parameters of an infectious agent can only be characterized retrospectively. The whole point is to perfect this adaptation with historical data such as Covid-19. It is by training that you become an expert !

The ideal of this project is to propose as quickly and as appropriately as possible an optimal decision sequence for society, producing a realistic and benevolent territorial digitalization towards the individual.

It is clear, however, that the relevance of the use of observed infectious data is highly dependent on the use of society's dynamic data. One cannot properly approach the reality of an epidemic spread without having a robust and stable basis for the dynamic behaviour of a society.

So what are combinatorial decisions based on ?

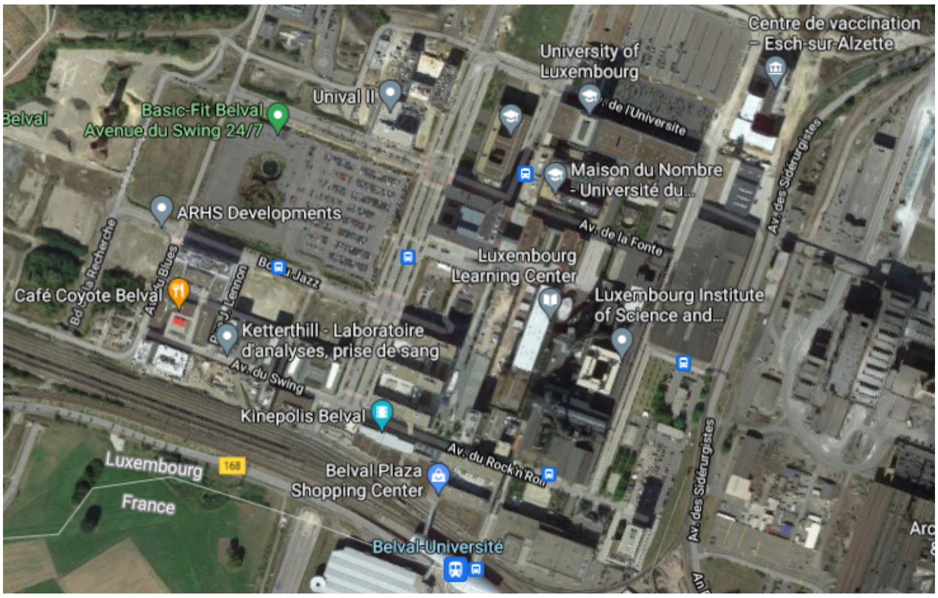
Such an environment cannot be optimized by leaving a large research space. It is necessary to define a predefined number of decisions, with their detailed constraints in the environment. These decisions impact the dynamic digitalization of society and modify the mobility of society. There are differences in the sequence of spaces traversed by each individual as well as in the affluence of these spaces. It is from this data that evaluation scores on several different criteria are established. These scores form the support of multi-criteria combinatorial optimization that establishes the most appropriate sequence of decisions possible.

Combinatorial decisions thus form a final layer of the program and are not essential for simple societal digitalization. One of the problems is therefore to code the static and dynamic digitalization of society, while leaving a possible modification on the part of a higher layer.

Why invest in this project ?

Simplistic simulations can already be used to predict the health evolution of an epidemic. For example, the experience presented with the CRS model in thefollowing presentation includes : *Youtube Example*

With this project, we propose to bring a real approach to society :
we want to consider the places of passage specific to each individual, predict societal behaviors and study infectious exchanges according to different parameters. In a defined region, real data such as map, demographic, epidemic and many other data is used. The objective of this project is to have a multi-criteria optimization perspective of epidemic management. We want to act with kindness towards the individual, limit health risks and maximize the economic stability of a territory. We define this territory on Luxembourg, and more precisely on the Belval campus in Esch-sur-Alzette for this internship. But this project is intended to adapt to any territory.



A satellite view of the Belval campus in Esch-sur-Alzette from Google Maps.

To talk in depth about the motivations of such a project, it is more than necessary to talk about moral principles and human benevolence, especially when talking about artificial intelligence. We go beyond the technical framework of an engineering internship but it is interesting to talk about it in the appendix at the end of this report.

Historically, natural, environmental and epidemic disasters can be classified as the largest humanitarian disasters. What other problems are more essential to optimize ? We do not count wars since there is a voluntary desire on the part of man to destroy his neighbor, which must not be optimized in the ethical sense. The stakes of humanitarian disasters are the most important in the world, whether human, economic, psychological, etc. At these levels, the slightest optimizations at the lowest scales are always crucial.

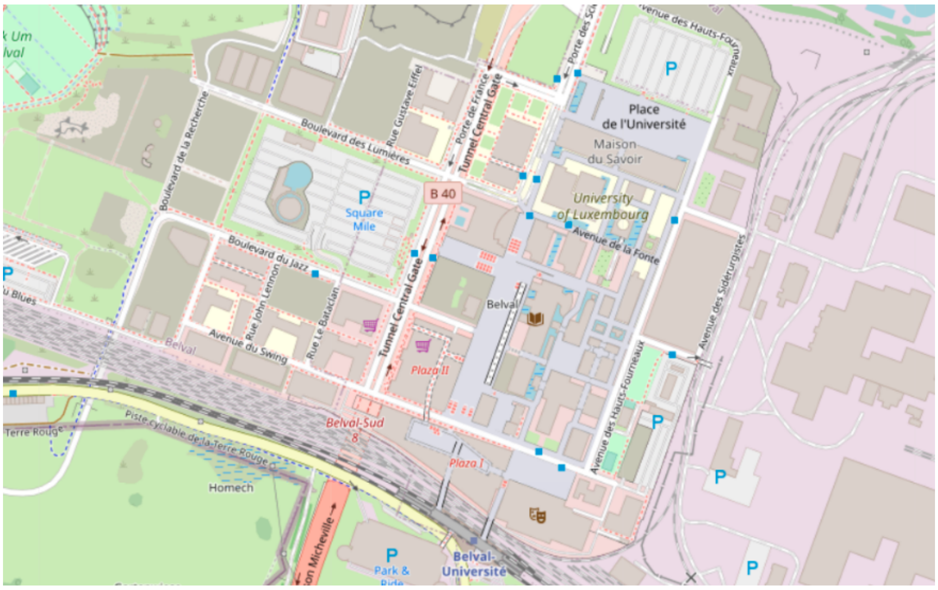
In our case here of epidemic crises, the subject is topical with Covid-19. Current events show us all the importance of optimizing the management of an epidemic crisis. From a historical point of view, no technology at the time of the old crises made it possible to design an optimization of decisions. Remembering the appalling damage of previous health crises, it seems obvious that they must prepare as well as possible to optimize future epidemics/pandemics. In the immediate term, this would also support current decisions in the management of Covid-19.

As mentioned earlier, by breaking down the elements of this project, one can also find other interests such as the modeling of a society and its dynamics, independently of a spread of infectious agents. In addition, beyond the novelties in the digitalization of an epidemic, we also find an interest in optimizing several criteria in epidemic management : it will be interesting to think about looking for other avenues of multi-criteria optimization with reinforcement learning, which are not already addressed in research. These avenues of exploitation may include the complex environment and the possibility of accurately describing actions in the environment. With this project, it is also possible to detail many evaluations with very different criteria.

Chapter 1

Static digitalization of society

1.1 OpenStreetMap



A cartographic overview of the Belval campus in openstreetmap format.

To design a static digitalization of society as authentic as possible, the project bases its design on the use of an openstreetmap file of the territory studied. Everything starts from this starting point. The idea is to use the places and the real road network of a determined area to initialize the spaces, and have the support of the paths between the different places.

To facilitate the execution of the project, it is therefore interesting to be able to retrieve the openstreetmap file directly from the program. To do this, we fill in several information such as the bounds representing the minimum and maximum latitudes and longitudes in a Python implementation named Overpy.

One of the small disadvantages of openstreetmap is the delimitation of the territory studied. A priori, it is possible to recover a region by forming a polygon, most often rectangular but it is not intuitive in our use here. In particular to use demographic statistics for example, which are appropriate for a certain delimitation of a territory. It would be necessary to check the possible purges already existing before deepening this approach but it seems wise to transform the recovered file by filtering the elements that are contained within a well-defined perimeter : we can speak of "administrative" limits of the territories where we look at the official borders between the different regions. In general, it would be very interesting to evaluate the position of the elements in relation to a peripheral boundary. If there is no attribute in the element that indicates its administrative location, it would be necessary to model the delimitation by a sequence of points with acceptable longitudes and latitudes and to check element by element. There are probably already algorithms for this filtration.

Another interesting point to look at is the proximity of an element to the boundaries of the territory. This is already more complex than knowing if the element is contained within the perimeter and therefore requires a little more work. An element located near a border may have a different treatment in digitalization but we will see this point a little later.

In a later version, it might be interesting to see an interest in breaking down a large geographical area into a mosaic of several small areas in order to parallelize the static digitalization of society.

1.2 Spaces

1.2.1 Format openstreetmap

Initially described by an openstreetmap file, spaces have characteristics specific to the file format. It is therefore necessary to read all the elements and rework the characteristics of each of the spaces. A space is primarily described by a set of nodes, each of which has a precise latitude and longitude. This set of nodes forms an enclosed space that delineates the contours of each space. By simply reworking the data, each space can be characterized by common attributes :

- A unique identifier, taken from the osm file.
- The nodes that delimit it ; useless in a simple version but which may possibly be of perfectionist interest.
- Its average location, calculated with the average of its nodes ; The result is questionable in geometric terms but sufficient to summarize the position of the space.
- Its area, calculated with its nodes ; Calculating an area of any polygon with geographic coordinates is very imperfect and requires the use of the Shapely package.
- Labels from the original file that can provide realistic specifications to spaces ; We will find the original classification of the space in OpenStreetMap and sometimes amusing information such as the name of the structure, etc.

The following link provides more information about the openstreetmap format : *Wiki OpenStreetMap*

1.2.2 Database

The reading of the elements of the file is sequential and is not parallelizable a priori. The file structure is broken down into nodes, path, and relationship, preventing independent processing of subsets of the file. Of course, as mentioned above, it is possible with a geographical import pre-treatment to break down the territory into several sub-territories to read them independently.

It is obvious to register the spaces in a database for the use of digitalization. Instead of simply saving the interesting elements of the file, it is therefore advisable to think about the memory management of these spaces.

In the openstreetmap format, spaces are classified according to a key and a value according to the type of place. To best digitize an epidemic, the implementation reclassified the different types of places in order to better group spaces according to their societal role, their economic function or their ability to accentuate the contamination of their occupants. The underlying idea is to treat spaces of the same category with common parameters. The new classification is described by types, which in turn consist of subtypes. The role of room types is more affiliated with a consistent organization of the database. They mainly allow subtypes to be logically grouped together, which can be useful for example to treat subtypes in the same way.

Each subtype therefore has common characteristics that allow, for example, to adjust the health and economic evaluations of each of the spaces associated with them. These parameters make it possible to adapt the value returned by the sanitary and economic functions of the spaces.

1.2.3 Advanced attributes

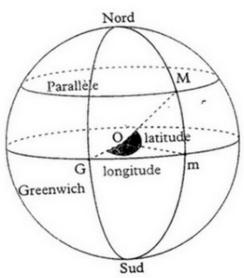
Common attributes can easily be added to spaces, subtypes, or even to the types of spaces themselves. In most cases, it is best to define attributes in subtypes by a Python dictionary. Their uses in practice can be modulated by means the simple attributes of spaces such as area. This is the case, for example, with assessment factors relating to the economy or health risks. After that, we can specify the subtypes more by reducing the roles that their spaces can play. The explanation of the roles will be given with dynamic digitalization.

In terms of attributes, we can extend the prospects of evolution by adding, for example :

- Opening and closing times ; These schedules must be acceptable and logical in relation to the categories of spaces.

- Average passage time of individuals ; to nuance with random parameters and by dissociating the different roles of individuals.
- Obligation to present a valid known condition ; depending on the validity of a health or other test.
- Possibility of teleworking and the associated percentage of performance ; Only working hours are considered in economic terms and this parameter is therefore exclusive to spaces with a work role.
- Limit of space of the occupants at the same time ; by not considering individuals with the role of workers and by nuance with the area of spaces.

1.2.4 Distance management



In modeling, spaces make it possible to add distance design to the digitalization of society. However, all geographic data are given in latitude and longitude. There are two ways to calculate the metric from this data. There is a Shapely package in Python that allows you to have accurate results but takes longer to execute compared to the simple mathematical method. Shapely is therefore applied to obtain exact data to record such as the area of spaces. It is preferable to use the simple method to compare distances. This is the case, for example, to select the nearest space from a location. We can also filter a set of spaces according to their relative distance to a location, but also according to their area or category. Due to some uncertainty, the results may differ significantly from the Shapely long method. It will therefore be necessary to empirically add a factor to the mathematical method to adjust it to the long one.

1.3 Individuals

1.3.1 Constitution of static data

The use of cartographic data from an openstreetmap file does not make it possible to constitute the demographic heritage of a territory. It is necessary to find strategies to identify the number of individuals to initialize as well as to characterize them adequately.

The best method is to combine the delimitation of an openstreetmap territory with the administrative boundaries of a region in order to make the best use of local demographic data. If this is not possible, then we can analyze the openstreetmap file itself and find solid correlations between the number of individuals and elements of the file. For example, a linear regression between the total file size and the population of the region could be sought. Otherwise, by analyzing in more detail, we can look at the total area of dwellings in the file to deduce the number of individuals. There are many possible hypotheses just waiting to be tested. Of course, it is always possible to define a number of individuals manually, independently of the map data.

In our case of the internship, the Grand Duchy of Luxembourg provides a very interesting set of public data on a website : *Public Data of Luxembourg*

In the case of this section, this concerns demographic data or household data for example. To best represent the population in this project, we therefore have the total population by Luxembourg region in 2021.

1.3.2 Representation of individual diversity

How to bring the static diversity of individuals ?

Especially in the context of an epidemic study, it is necessary to be able to represent the diversity of organisms in terms of resistance, fragility, etc. And from the simple point of view of society, it is essential to distinguish individuals with significantly different dynamics. For example, students will not have the same dynamic as retirees.

However, we can find a common factor rather simple : age. By nuanced with classes of individuals with similar behaviors, it seems intuitive to characterize an individual by age and socio-dynamic category.

It is also available to know the distribution of the Luxembourg population according to age, from 0 to 95 years. For the sake of simplicity, the proportion of the Luxembourg population over 95 is added to the initialisation. The population of the implementation in the current state is therefore divided between 0 and 95 years. A priori, it seems unproductive to focus on the evolution of population ages, births and non-infectious mortalities. But we can nevertheless leave the door open to these avenues of exploration for other benevolent projects concerning demographic change, for example.

1.3.3 Database

Like spaces, individuals are classified into types and subtypes. We find the principle of dictionary to list common characteristics.

A subtype of individuals has several basic attributes :

- The minimum age and maximum age of the individuals who compose it.
- The categories of dwellings in which its individuals can sleep.
- The categories of spaces in which its individuals can work.

On the side of individuals, each of them would be characterized by :

- A unique identifier.
- An age.
- A set of data corresponding to the personal base.

How to distribute the ages of individuals in digitalization ?

It is an ambiguous problem. On the one hand we have the socio-dynamic classes of individuals and on the other the age distribution of a territory. We can question the choices of implementation but the chosen solution seems appropriate at first :

At initialization, the total number of individuals is arbitrarily distributed across all existing types of the database. We therefore have a subnumber of individuals to distribute between the different subtypes in each of the types. This is where the age distribution of the territory, in this case Luxembourg, is used. Only the age range identified in the subtype is taken and the proportions to a total of 100% are determined.

In the case of a tangled set of subtypes with overlapping age intervals, it is preferable to respect the distribution of the territory over the entire type of individuals. As a result, all the age intervals of the subtypes are first recovered in the same interval, reproportioned on a scale of 100%. The initialization begins by choosing an age respecting the different probabilities of each of them. When the selected age can be associated with several categories, the choice of subtype is then made randomly with equiprobability.

The demographic distribution of ages is therefore respected at the level of a type, but important nuances can be found when considering a set of types of individuals. It is even possible not to find certain age ranges in static digitization.

1.3.4 External consideration

One of the great complexities in digitalization, whether static or dynamic, is the external modeling of the territory. All territories are permeable, unless we take the world as a whole, which is unthinkable. However, it would be interesting to discuss the correlation between the surface of the territory studied and the static, and dynamic, outdoor affluence to be digitized. For the same surface, we will find an external influence much greater in Paris than in a countryside. So is there rather a correlation with the "internal" population of the territory, which as a reminder can also be estimated from the openstreetmap file itself ?

Now, it is true that we can question the relevance of age and socio-dynamic classes. Are the "external" individuals always the same? As a whole, do we find a distribution of ages that is logical or similar to the territory studied ? Is there even an interest in recording them, or rather representing them by spontaneous individuals with random attributes ?

It is obvious that this approach must be deepened with more hindsight and reflection. From a static point of view, external persons can be identified by their occupation with the role of worker ; the distinction between interns and externals is based on the position of their home in relation to the territory. And then from a dynamic view, we must also add individuals a little randomly, with a frequency and diversity to discuss.

Their digitization is essential but requires more reflection so that it is relevant to talk about it.

1.3.5 Individual base

The individual base are essential assignments. They make it possible to bring a form of determinism and coherence in dynamic digitalization. In an initial version, the base is represented by the individual's housing space and work space. In both, it also seems sensible to specify the frequent local entourage of the individual. In absolute terms, this makes it possible to have a very good modeling of the digitalization of each individual in a case of great dynamic restriction.

In the case of the epidemic study, it is more than essential to have good modelling in case of major political restrictions. This would make it possible, for example, to see the real effectiveness of containment and to observe possible flaws in the connection of contact networks, for example.

Of course, this base can evolve over time but it is assumed that the individual keeps his home, his roommates, his work and his close colleagues in a first version. We will see later ways to add more information to model and anticipate the dynamic base of the individual. This will be the case, for example, with dynamic digitalization, where it will be important to assign a mode of transport to each individual. Among the most relevant tracks, we can be interested in considering the close friendships and family relationships of the individual. However, the question of the exploitation of this base in dynamic terms remains to be deepened. Similarly, we can possibly extend the base with the habits of individuals, which can be compared to needs more precisely here, such as the daily visit of the individual to a gym to cite an example. Playful habits would really not be relevant, or even interesting, to consider to simplify the design.

It should be noted that this conception does not address the age of individuals : children are considered independent persons with housing, workplace, etc. In this case, care should be taken to establish a school as a workplace. We can then think about the workspace of individuals without work. Is it empty and in which case it would be necessary to treat this case separately, or should it be represented by their housing space ?

How to assign the individual's housing and work spaces ?

The objective is simple, to make a join between the database of internal individuals and that of housing spaces. As a reminder, each category of individual has a list of housing categories where he can reside. For example, retirement institutions will not be able to be found in a socio-dynamic category of children. This also leads to other reflections for the classification of categories of individuals to specifically associate disabled individuals with adequate housing establishments. The same principle is found at the level of professional spaces. Each category has defined sets. However, it is necessary to take into account the complexity of the professional distribution of a territory.

In a simple version, the selection is made randomly among the set of categories associated with it, excluding no space specifically. The adjustment is made by weighting the selection with the area of each space. Therefore, a space with one area greater than another will have proportionally more assignment than the second. We can of course review this selection to perfect the static digitization of the observed data. Similarly, it might be wise to look at the possibility of excluding spaces in selections.

This conception is not complete : it is necessary to be able to identify the people really close to the individual in each of these key spaces. For example, when the space in question is a small apartment, all the individuals in the dwelling are indeed in contact. On the other hand, when the dwelling is a large residence as a whole, very few people are really in contact with the individual. This is why it is necessary to list the people said to be close to the individual in each of the assigned spaces. The ideal would be to use real demographic data such as the distribution of the number of people in households or territory. Indeed, this attribution is made iteratively, without wanting to reproduce logic in the construction of households. The program does not seek, for example, to gather two children and two adults in their forties. Conversely, it is totally possible to observe original groupings such as five children in a house. This problem of coherence has less impact on workplaces since there seems to be more heterogeneity. Especially since a certain logic is respected with the association of job categories with that of individuals. That is to say, there is no risk of observing children, whose work is school, working alongside adults ; except when it comes to teachers. Of course, we can also look at the real data of the territory on the distribution of the labor market for more realism.

Chapter 2

Dynamic digitalization of society

2.1 Scenarios of individuals

From an epidemic perspective, the more agitated society is, the more an epidemic spreads. It is therefore necessary to describe the individual dynamics of each individual. The question of time is essential to deal with. Considering the diversity of trips on the same day and the use of SumoMobility, it is obvious to deal with dynamic digitalization day by day, on a minute scale. In practice, a second scale will be applied to adapt to the SumoMobility format.

How do I represent a 24-hour scenario ?

The idea is to present a sequence of spaces with defined slots as well as additional information specifying the movement such as a degree of importance or the role of the individual. All scenarios start at a fixed time in the day. As a rule, more active individuals are observed at 01:00, during the night, than at 05:00 in the morning. It therefore seems wise to specify a fixed start time of the scenarios at 04:00. The scenarios therefore begin with the awakening of the individual and end with his sleep. Of course, in theory, a scenario begins and ends at the same time. The greatest complexity of dynamic digitalization is the generalization of the creation of scenarios for each individual, respecting the diversity of situations. We can resume the parallel with a neural network.

2.1.1 Multiple hyperparameters

A scenario must consider several factors to meet this specific complexity :

- Infectious state of the individual : known states, actual states.
  The infectious state primarily impacts an individual's scenario. The more severe its actual infectious state, the less attention is paid to the condition identified to it. On the other hand, the healthier its real state,

the more interested one is in its known state. We can speak more of the propensity of the individual for a less serious condition and constraints/obligations of the individual for a serious condition. For more details, the infectious reflexes associated with each case are lower in the report.

- Territorial decision :
  Territorial decisions impose constraints on the population. These are the combinatorial solutions on which the reinforcement learning algorithm relies. This is the very basis of this search for multi-criteria optimization of the company. These territorial decisions are to be defined manually by describing the actions they apply to the company. It should be noted that decisions may be associated with constraints/obligations on tests and vaccinations. For more details, the decisions are further discussed in a later section.

- Type of day : weekday, Saturday, Sunday, holidays, seasonal period.
  A person's daily life depends a lot on the calendar. Depending on the type of day, the company can be modeled quite differently. On weekdays, most people work during the day, have little free time in the late afternoon and then go out or stay at home in the evening. Otherwise, on weekends, they have most of their time to do what they want. The type of day is therefore an interesting parameter because it makes it easy to have information on the state of the dynamics in general. In addition, it is obvious that the time of year is a very important parameter to best characterize dynamic propensities. Individuals orient themselves more in parks in summer than in winter. In a later version, we can have fun with this calendar to have more tools and leverage to design realistic scenarios.

- Type of daily life : work, rest, holidays
  Each individual can have different daily newspapers regardless of the type of day. He can either have a daily life where he spends a certain number of hours, defining, at work, or be at complete rest, or even go on vacation and get out of his routine completely.

- Role of spaces : economical, necessary, unnecessary work

The role of spaces is important to clarify. From an economic valuation point of view, it is essential to distinguish between the individuals who work or visit the different spaces. After that, it is also necessary to distinguish between necessary and unnecessary travel. This has little impact in a normal situation, but these specifications are more than necessary in case of restriction. For example, supermarkets, medical places and pharmacies are necessary travel. In a later version, it might be interesting to no longer represent the importance of spaces in binary terms but with a scale of necessity. For example, we could consider that gyms are a psychological "spoiler" for individuals. These spaces could then be considered as semi-necessary places. We return to an evolutionary perspective where we consider the habits of individuals, when these seem to be an individual need. Conversely, a center of attraction will always be denoted as unnecessary for example. Apart from the working role of spaces, this level of importance can be classified directly into space categories. This could also add a new psychological dimension to the project, which will be exposed at the end of the report. It should be noted that this has nothing to do with the ability of spaces to contaminate its occupants.

- Various propensities : external travel, other accommodation, family or friendly visits, etc.
  Each individual has a propensity to move outside the study area. Individuals residing near a boundary of the study area will be more likely to travel outdoors. We find the problem of identifying the proximity of spaces to the geographical limits of the region. Then, the individual may occasionally sleep elsewhere than at home, whether in the study area or not. We can also consider the propensity of the individual to share moments with friends and family. We find the perspective of evolution of the project where we specify the friends and family of each individual. It is in a way the hyperparameter for all the possible evolutions of dynamic digitalization, which will be useful among others to approach the observed affluences.

- Type of space selection : assigned to the individual, random, criteria
  When developing a scenario, you need to know what space to add to the sequence. So the question does not arise as to the base of the

individual where housing and work are well identified. We join the representation of the habits of individuals, which can be associated with certain propensities. As for the rest, the daily life of an individual can be very diverse, and it is necessary to be able to choose spaces from the entire database. From a simplistic point of view, we can say that the individual visits a sequence of space a little randomly. In a certain set of categories, a space can be chosen completely randomly or according to certain criteria. The more random the selection of spaces, the more the circulation of the epidemic will have a significant Brownian motion. It may therefore be wise to perfect this selection by criteria such as the nearest or largest space within a certain radius of distance for example. It is therefore necessary to add filtration functions to select a space from the database. This hyperparameter could be perfected with voluntary population data on their preference, such as surveys.

2.1.2 How do you do it ?

Optimizing agendas for a population is already a complex problem. The idea here is to simplify this generalization of scenario with the parameters mentioned above. In order of importance, the scenarios depend first on the infectious state of the individual and then on territorial decisions, especially during major constraints/obligations. Then, we can identify travel patterns according to the type of day and the type of daily life of the individual. Then we have the various propensities of the individual to organize certain necessary or non-necessary trips. For the rest, it is a question of identifying the spaces to be introduced into the sequence, whether it is already predefined, chosen according to certain criteria, categories or entirely random, etc.

Each individual has certain necessities that he must realize according to a certain rhythm over time. This is the case, for example, with food shopping or routine medical visits. It is not realistic to reproduce this kind of travel every day. It is therefore necessary to establish a system of propensity of the individual to build his daily life. A first approach could be to reproduce more or less the same system as a shared distribution of computer tasks on the same computing core, where the tasks would represent all available travel. Each available move has a score that increments over each day. It will be necessary to separate them by necessity or not. By considering the score of each of them, the individual would select the most appropriate one in his daily life.

The selection lever could be a threshold system from which the individual would place the displacement in his scenario. It would then be necessary to respect the space available and the competition of trips exceeding this threshold. It would seem wise to choose an identical threshold for all spaces and to change the daily increase in the propensity score of each category. Another leverage system could be to integrate these movements while playing with stochastic processes. These stochastic decisions would be made every day, regardless of the weather. An identical increment would then be applied for all categories of spaces. Differentiation would be added by factors specific to each of these categories. Regardless of the selection lever, the scores of each category would be reset to zero each time they are selected.

2.1.3 Problem of time

Time is the major constraint when designing a scenario. A scenario must not exceed 24 hours. There are two options for taking into account the time limit. A bit like the backpack problem, a scenario has a time limit and it must be filled with trips that have a certain importance/obligations, while not exceeding this limit. Then either the scenario can be built iteratively. Either we first select all the trips of the day and then we arrange them with timings. In theory, we should reproduce the conception of a scenario in our daily lives : we must all respect time slots and adapt to the time we have left.

There is a certain complexity to take into account : we can not predict the travel time of the individual, hence the interest of using SumoMobility to have the temporal aspect of the trips. We will see later that the project cadence the scenarios by the departures of individuals in the direction of a space, which therefore includes the unpredictable travel time as well as the time on the spot in space. It is necessary to have estimates for each of its trips, considering the type of transport and the desired duration on site. What interests us in this project is the actual time of each individual in the spaces.

There is a significant problem in poorly timing travel. If a time between two timings is too short, the individual may not have time to complete his journey towards space that he will already have to start the next trip towards the next space. In practice, there would be no problem when running with SumoMobility. But once the simulation is done, the analysis phase will not be able to properly evaluate the scenarios knowing that some spaces will not actually be visited.

It might be interesting to specify a reference on the occupancy times of individuals in spaces. Each category of space could therefore be characterized by a typical duration of passages, whether for a simple visit, a working time, etc. This duration could be represented by a fixed quantity or a positive normal distribution, parameterized by a certain mean and a certain standard deviation. Note that we can find a random duration in practice since we do not anticipate the exact travel time. An interesting estimate of this travel time could be made depending on the type of transport and the distance between the point of departure and the point of arrival. For example, a linear regression between distance, type of transport and actual travel time can be analysed.

Timings must also follow a certain coherence to group individuals in their dynamics. For example, individuals have fairly general slots for meal times. Work timings depend on the type of work of the individual, the weekly duration, etc.

2.1.4 Memory management

The question of memory management of each individual's scenarios is an interesting point. The basic problem is whether a scenario is unique over time ? Realistically, no. But for the sake of ease, we may prefer a memory management of scenarios.

To simplify, we can initialize several scenarios for each individual according to certain conditions, even if it means editing them according to their infectious state or political decisions, for example. This design would therefore be interesting to take into account the individuality of each individual and the calendar. For each day of simulation, this means removing the construction time from each scenario and replacing it with a time to adapt to the conditions.

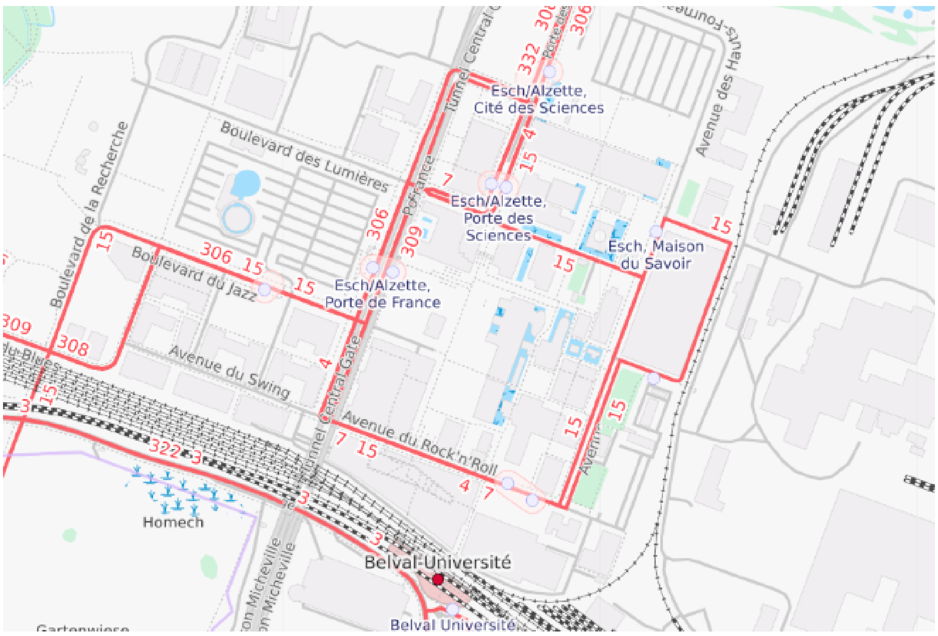
However, we have previously noted that the most impactful parameters when constructing scenarios are political decisions and infectious reflexes. From an execution point of view, this amounts to increasing the memory requirement of the program but accelerating the time of digitization. However, we lose some diversity and an unpredictable side of mobility.

The other conception would be not to keep in memory the scenarios but to build them every day of digitization. Compared to the first option, it would then be necessary to keep the tools for constructing scenarios for each individual such as propensity scores, etc.

2.2 SumoMobility

SumoMobility is an application simulating urban traffic from map data of a road network. By taking the openstreetmap file and applying pre-processing, we can circulate individuals through the different real places. It is possible to use public transport in the area described by the openstreetmap file with the actual positions of the stops. As a reminder, each individual is initialized with a means of transport, according to his age and certain proportions. For the efficiency of the project, it will be limited to a simplistic use of travel. The interest is to consider only the travel times of individuals in daily traffic, and to have an overview of the infectious risks in public transport.

Indeed, public transport is also a space to be taken into account in the spread of infectious agents, although it is not listed in the space database. This consideration can be done globally for simplicity : we can look at the ratio between the proportion of infectious and non-infectious people using public transport as well as the frequency of this public transport. An overall assessment score can thus be provided to the infectious assessment function of the individual ; This score will only be taken into account if the individual takes public transport. Of course, this consideration can be further refined to make it more individualistic.



An overview of public transport data in the openstreetmap format.

To simplify the use of SumoMobility, the simulation of a day in society is carried out without making individuals interact with each other. This avoids unnecessary calculations in our use, and also avoids unpredictable bugs. For example, one of these irregularities can be the collision/traffic jam of several individuals on a pedestrian lane, which causes serious problems on digitalization. Such considerations in the field of civil engineering are unnecessary and even very problematic for the project. Similarly, in a simplistic case, we are not interested in SumoMobility's advanced options such as recalculating routes according to traffic, managing low beams or using taxis. In our simple use, however, some parameters are to be adjusted :

- We must manually choose the frequency of passage of vehicles in the automatic generation of public transport with the stops and lines described by the openstreetmap file. This is an important parameter for individuals who travel by public transport, whether for the speed or the risk of contamination of their travel. However, it is quite easy to find information on the average frequency of public transport in a region, or simply to enter a frequency that seems realistic. It would be interesting to separate the different classes of public transport in a later version in order to be able to apply different frequencies of passage to them.

- It is interesting to look at the parking constraint of vehicles. In a simplistic version, it is assumed that individuals travelling by car do not need to leave their vehicle in a parking lot ; Cars appear and disappear directly at the pedestrian crossing. This parameter is not that important in the digitalization of an epidemic. Of course, in the practical case, this could represent the time that city dwellers spend to find parking in the city, not to mention the time or flat-rate constraints of car parks. Conversely, this option would be very interesting with a view to perfecting dynamic digitalization or civil engineering.

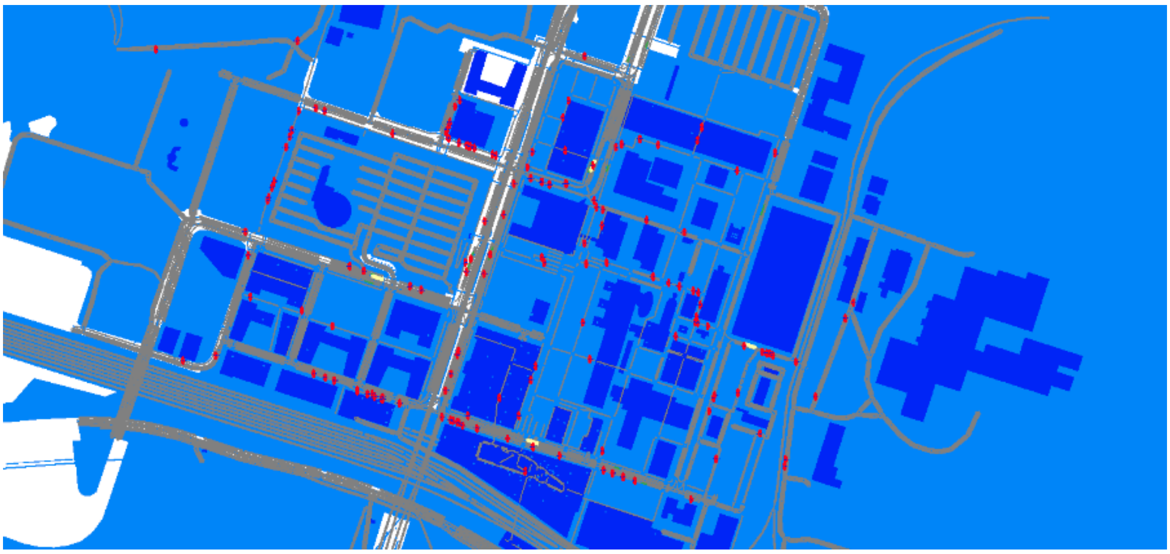
The simulation is continuous over time and travel under SumoMobility is done only by considering the road network. Therefore, to model the stop of an individual in a space, it is necessary to route the individual to the road that is closest to this space. For simplicity, the routes between the different spaces are made with the same type of transport, unique and assigned to each individual.

To make the connection between individuals' scenarios and their implementation as routes in SumoMobility, it is necessary to make some changes. From the scenarios of individuals, it is necessary to be able to write the trips of the individuals which will then be translated into routes by a specific algorithm of SumoMobility, called duarouter. This router is very technical and has many options to identify the fastest sequence of routes to connect a starting and an arrival route. It is possible to adjust many parameters related to optimization such as the vehicles that can be driven on each road, the speed limits on each road, the detailed speeds of each vehicle, etc. The algorithm he uses is Dijkstra. For simplicity, we do not worry about this optimization since it is not the subject of this project. This tool simply allows us to find sensible movements between points of interest. However, we are concerned about the connectivity of the road network, which we can try to correct directly with the router, so as not to have an error in the path search. In this acquisition, the algorithm can use public transport, for individuals using the public transport mode, which will have already been introduced and timed previously. An undeniable advantage of duarouter is the ability to perform calculations on several threads and speed up this process. In particular, this is not negligible when the scenarios of individuals evolve over the days and the routing step must be done regularly.

The router automatically creates vehicles with a unique identifier per individual. Each vehicle then makes the trip corresponding to a route. The total itinerary of the individual therefore designates the unique identifiers of these vehicles to carry out the routes associated with these vehicles. However, this design presents a problem when the individual owns a car and passes through many spaces : the stops of the individual, representing the stop in each space during the scenario, cannot be introduced between the same chain of roads. The individual cannot therefore use his car to make two different routes. The route of his car corresponds to the first movement of the individual, that is to say the first sequence of roads, without stops, of the individual.

One of the avenues for correction would be to model the route of the individual's single car, understanding the individual's stops. However, this does not seem to be feasible since duarouter. It is from this problem that we must separate the total daily itinerary of the individual in a simple itinerary. However, since we cannot predict the travel time of the first route, we must estimate the timings on all subsequent routes. In an ideal case, it would certainly have been more interesting to consider only the occupancy time of individuals in spaces, without worrying about estimates of transport times.

With this design, the scenario of each individual is divided into several sub-paths, identified by the identifier of the person to which are added specific characters that increment with the number of routes. The same individual, under different route identifiers, can use several cars for different routes. A route thus becomes the simple passage from one space to another. As a reminder, we do not take into account the duration that the individual occupies a space but the timings where the individual wishes to move into a space. The duration of the individual's occupation in a space is then reanalyzed after SumoMobility is executed. The decomposition of an individual's scenario into several routes is done contiguously in the routes file, which removes the chronological order of all routes from the file. It is therefore necessary to apply an algorithm specific to SumoMobility to reclassify in order all the routes in the file according only to the departure time of each trip.



En example of graphical execution of SumoMobility here in the network of the Belval campus in Esch-sur-Alzette.

In a first version of the graphical display, it is possible to observe the circulation of individuals in an environment described as in the example above. Regardless of their infectious status, individuals are described by red dots. The shapes in yellow correspond to vehicles such as cars or buses. The spaces that individuals travel through on a daily basis are depicted in dark blue. Of course, it could be interesting to review this display to show more details :

individuals would be represented in different colors according to their infectious state and the same for spaces according to their category. On the other hand, it seems impossible a priori to change the display as a function of time to try to represent the infectious spread during the execution of SumoMobility. It should be noted that we cannot observe the movements of individuals between the road network and the interior of spaces.

It might be possible to display the map by adding graphic elements to represent the final statistical data of a day for example. Otherwise, after a certain number of days, we could graphically represent the spaces where there is the most contamination. There are plenty of illustrative possibilities to provide interesting information.

Dynamic digitalization by SumoMobility can also be parallelized during its execution, which is again not negligible for the efficiency of this project. However, the graphical display has limitations : the execution is much slower and can even saturate during simulation in case of too large region. Execution always has irregularities that are not predictable but have no impact on digitalization. Among them, we can mention the appearance of collisions between vehicles, which can be resolved automatically with SumoMobility parameters.

2.3 Road network

It is necessary to design an intermediary between the identifiers of the spaces and their implementations in journey under SumoMobility. Indeed, under SumoMobility, trips use route identifiers to design routes, and do not recognize spaces.

At initialization, the application imports the road network of the environment with the openstreetmap file, while keeping the identifiers of the import format as well as possible. However, whether through the openstreetmap format or its import into SumoMobility, no direct link is made between spaces and roads. It is therefore necessary to manually assign a route to each space : on one side the routes in SumoMobility and on the other the spaces in the internal database. When importing the network, the openstreetmap identifiers of the roads adapt to the road network according to the arrangement of the roads between them. If a road has junctions along its description, no warning is displayed in the openstreetmap file. On the other hand, SumoMobility breaks down the route into different routes represented by the same original identifier to which are added unique characters. However, we will always find the same logic in these characters : it is therefore accessible to link a road with its openstreetmap identifier by one of its subroutes with SumoMobility identifiers.

For each space, we will therefore be interested in applying a local search to identify an openstreetmap route. To do this, a search loop is carried out that gradually extends the geographical filter perimeters around the position of the space. One of the approximations concerns the location of the roads. Even if we find the principle of nodes forming a polygon, a road under openstreetmap can be very large. Its average location for local search purposes can be very incorrect if you do not focus in detail by its nodes. Especially since it is attached to a sub-road in SumoMobility and this sub-route is not always the portion of the road closest to space. It should be noted, however, that these inaccuracies remain insignificant. Except for special cases, we will remain in a simple search. For a time constraint, the fast distance calculation method is used.

2.4 Interterritorial movements

To digitize interterritorial travel as simply as possible, we do not worry about incident or efferent journeys with the outside. Indeed, travel times are interesting only to better observe the occupation in the spaces and to better assess the contamination of individuals.

Epidemic management outside the network is not "controllable" or predictable and the area we study is somehow subject to external health states. This is why we limit ourselves to a more or less random health assessment of individuals outside the region. However, one can roughly generalize the scenario of outsiders and vary the infectious incidence randomly on a case-by-case basis. For example, it can be assumed that an individual leaves the territory studied, spends six hours in an office to work, goes on to two hours in night clubs, and finally returns to his home inside the territory. His infectious exposure outdoors will be more critical when he is in a nighttime club than in an office.

It is clear that the relevance of the infectious assessment is highly dependent on the length of time before it returns to the territory. It would be absurd to study the infectious assessment of an external individual who spends only weekends in the territory, for example. In this case, it would then be reasonable to estimate the infectious state of the individual a little randomly before the weekend, see if he can accommodate the current decisions and then integrate it into infectious assessments during the weekend.

Another interesting case would be to study interterritorial movements, particularly in the case of serious outpatients. Ultimately, it would be necessary to deepen reflections on external modalities. In particular, with regard to the propensities of internal or external individuals to make interterritorial movements.

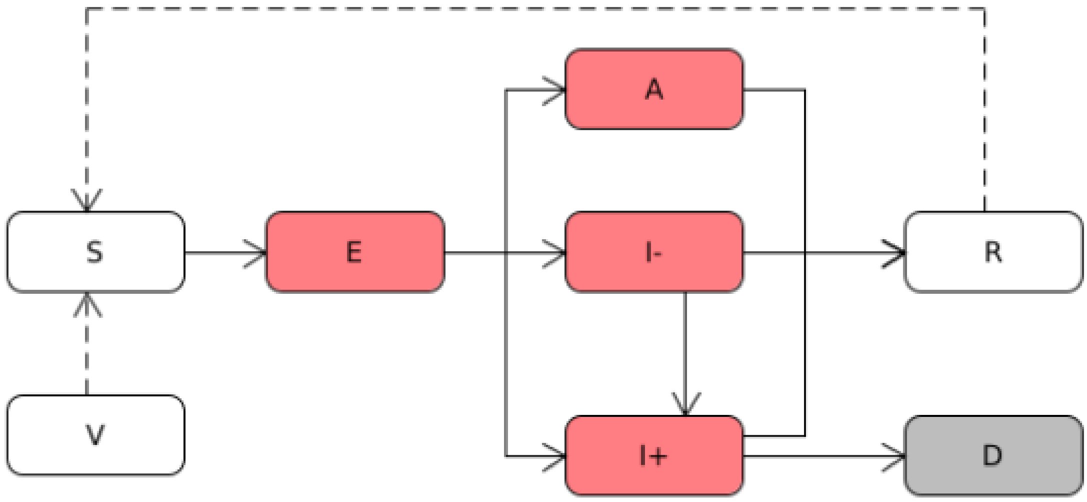
Individuals from outside the territory can, for example, start their arrival at a hotel in the territory. Otherwise, regardless of the location of individuals' homes, when an individual comes from outside, he can start his itinerary with any space directly ; or especially in spaces associated with long journeys such as a train station or airport. In general, there are not necessarily constraints either on arrival or departure. On the other hand, we can direct digitization to reproduce observed data.

In a later version, we can also look at the existence of several different territories, each with different characteristics.

Chapter 3

Dynamics of infectious agents

3.1 Infectious pattern



An overview of infectious pattern.

S : Susceptible	E : Incubation phase
A : Heals	A : Asymptomatic patient
D : Deceased	I- : Standard patient
V : Vaccinated	I+ : Seriously ill

Following the reflections of epidemiologists, one can break down the infectious states of individuals as described above on the diagram. The continuous arrows represent a certain evolution that depends on several temporal and/or stochastic parameters. Those in dotted lines are more anecdotal. The phases in red correspond to the phase where the individual is a carrier and can transmit the infectious agent. The starting point of the population is in the susceptible state.

At the initialization of the infectious agent in the environment, according to a ratio and / or a certain initial amount, individuals are placed in the incubation phase. This selection is carried out randomly on all individuals in the environment. It may be interesting to determine more precisely the first individuals carriers and to study the difference in the spread of infectious agents. This is one of the interesting points of this project.

Each individual is in a specific state for each of the infectious agents in the environment. Competition between these agents can have an impact on their spread. For simplicity, when an individual is a carrier of an infection, he becomes "immune" to all other infections in the environment: he is then in one of three states among susceptible, cured or vaccinated according to his situation and experience.

3.2 State transitions

3.2.1 Susceptible

Each infection is characterized by a contamination threshold that corresponds to the number of particles from which an individual enters the incubation phase. It seems wise to adapt it according to the age of the individual to represent the resistance of the organism. In a simplistic version, it is assumed that the number of particles emitted by a carrier individual is identical regardless of the infection, the infectious state of the carrier individual or the duration for which the individual has been a carrier. This design may be revised in a later version to add a factor in the particle emission based on the infectious score of the carrier individual. Two other parameters characterize the contamination of an infection :

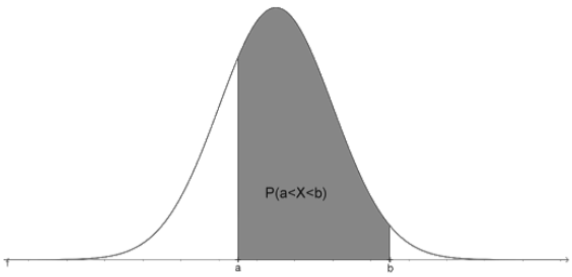
- The time that a particle can persist in space ; We will set this duration even if we can question a Gaussian approach. It was preferred to play on the efficiency of the really persistent particles in space with the factor concerning the surface of space in the infectious evaluation.

- The number of days that a particle can be accumulated in an individual ; This setting can hide another. We could take up the principle of an immediate reaction of the organism: the individual would constantly accumulate external particles, without defining a limited duration, and would produce an immune reaction. It will then be necessary to define an insufficient immune capacity to prevent the total particle from exceeding the contamination threshold. It should be noted that this new approach is inspired by the results obtained during this internship: individuals seem to become infected much too quickly without leaving a certain elasticity.

3.2.2 Incubation phase

When the individual is in the incubation phase, during which no symptoms are visible, two parameters characterize the infection: the mean and the standard deviation of the duration of incubation, which is assumed to follow a normal distribution, regardless of the characteristics of the individual. At the end of this random duration, the infectious status of an individual changes to one of three states : asymptomatic, standard or severe. This transition depends on the age of the individual and the infection.

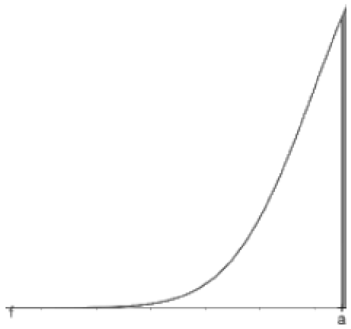
When the incubation period is up, each individual is given a new infection score, such as an amount of infectious antigens in the body. This random score follows a normal distribution according to a mean and standard deviation characteristic of the infection. The idea is to define the infectious state of the individual according to this random amount. Indeed, it is possible to select two thresholds around which we can find three known data in a Gaussian law. This distribution is also to be qualified according to the age of the individuals. Initially, the infectious representation of individuals is therefore described by a single score, assigned using a normal distribution with fixed parameters and regardless of the age of the individuals. The variation in infectious severity is therefore modeled by Gaussian bounds (a,b) that vary according to the age of the individuals. However, it seems more than wise to study the opposite degree of freedom for simplicity in a later version.



An overview of a Gaussian distribution.

Once the infectious score is initialized, the evolution of the infection is done by increasing or decreasing this score, according to parameters specific to the infectious status. In a natural way, the organism already tends to evolve on its own. If the infectious score becomes zero, the individual is cured, otherwise, if it exceeds a certain mortality threshold, the individual has died. To draw a parallel with a very simplistic medical view, this score can represent the difference between the number of antigens and antibodies of the individual.

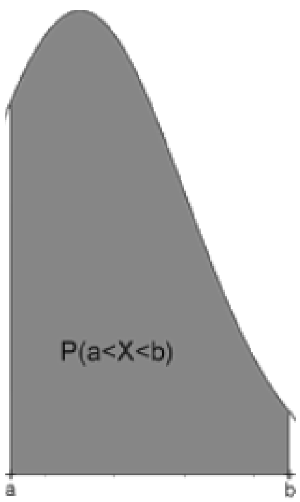
3.2.3 Asymptomatic patient



Each individual can heal on their own according to a certain daily self-healing score, much like an antibody production. For a first version, it is assumed that this score is a fixed amount and independent of the age of the individual. In view of the curve of the infectious score, this means that the healing time does not follow a normal distribution but rather a remission distribution that can be described as late. The healing score can therefore be revised in a later version depending on the observed data and the first results of the conception. It is assumed for simplicity that asymptomatic individuals are not susceptible to infectious care or aggravation.

When he is not aware of his asymptomatic infection, there is no change in behavior on the part of the individual ; this state does not cause infectious reflexes. Obviously, if his infection is identified, the individual will have to change his daily life to limit the epidemic spread according to territorial decisions. In a simple version, we are not interested in the propensity of the asymptomatic sick individual to adopt by himself a protective behavior vis-Ã -vis the population by modifying his mobility. This means that only political decisions influence this state.

3.2.4 Standard Sick



The standard infectious course of an individual follows the same principle of self-healing as the asymptomatic one, except that it is possible/recommended for the individual to follow care. Whether through medical visits or medication, the individual's infectious score can be further reduced. Conversely, it is possible to undergo events that increase the individual's infectious score and eventually move it to the severe stage. Examples include poor care that is involuntarily perceived by the individual as nosocomial diseases or poor self-medication. Each type of care could be either positive or negative depending on a probability. These types of care would also be parameters to be implemented in each infection.

An individual in a state of standard infection has symptoms of infection and his daily life will be impacted by infectious reflexes. These infectious reflexes can push the individual to test himself or to introduce care into his daily life, whether medical visits or medication. Because of its weakened body, it will also tend to limit unnecessary travel. However, it is interesting to note that the symptoms do not necessarily reflect the identification of the infection: the individual will have to follow the policy regarding standard individuals only when it is evaluated by a doctor or an infectious test. Depending on territorial decisions, the daily lives of these individuals are also modified with constraints or obligations. These decisions may guide the individual to identify his condition.

3.2.5 Seriously ill

Individuals in a serious infectious state resume the same principle of care as standard patients. The care would then be different. In addition, no self-curing is performed by the individual and the infectious score increases daily. In practice, it is considered that its production of antibodies is not sufficient compared to that of antigens. This restricts design to simple antigen production, even if it deviates a little from a realistic medical model. The difference between this infectious worsening and the care available characterizes in a way the mortality of each infection. The individual is pushed for his survival to treat himself quickly: beyond a new threshold, the individual has died. An individual in a seriously ill condition does not need to be identified to be known: there is no distinction between the actual and known condition.

The infectious reflexes of these individuals become the daily priority of these people. Their survival instincts are not subject to territorial politics. Moreover, this policy must take into account the limited care resources of the territory to avoid abandoning its individuals in serious condition.

3.2.6 Cured & Vaccinated

For a first conception, we will not be interested in reinfections after recovery or after vaccination of individuals. An individual in a vaccinated state does not need to be identified to be known: there is no distinction between actual and known status. The question arises, however, for the state cures. In a simplistic version, it is considered that when the real carrier state is over, the individual passes into a real and known healing state. It is simply a pragmatic issue. In digitalization, leaving a preservation time, which can be compared to a quarantine time, has very little impact on the study of epidemic spread. The question may arise more on the economic criterion. And it is not interesting to reason about a political decision in this sense in an initial conception.

3.2.7 Deceased

A deceased individual does not need to be identified to be known... In this infectious design, the deaths of individuals do not follow a mortality rate predefined by a data observed in reality. The mortality of individuals depends on the proportion of seriously ill individuals, the gap between the infectious worsening and the available care of the infection, and also on the success of individuals in "surviving" with this available care. It is not easy to directly match a mortality rate observed in reality with digitalization. This is probably one of the most delicate points of the infectious design: the mortality of an infection depends significantly on the dynamics of sick individuals. As a result, the realistic development of the mortality of the infection needs to be deepened.

3.3 Care, Vaccines and Tests

In a first design, we can simplify the parameters related to care, vaccinations and tests. However, it is important to make digitalization realistic and to specify the elements in detail :

- Specify, by a fixed or random variable, the time in day in digitization before the element is available to individuals.
- Constrain the individual to move within a particular category of space to perceive this element. In which case we could add the average time of visit, and assign a necessary role to this trip.
- Add a delay before the effects of the element appear, whether it is the time of effect of a treatment or the waiting time for the results of a test. The case of vaccination is more interesting to treat: each dose of a vaccine has an effect on the body, more or less rapid. One of its effects could be to drastically decrease Gaussian parameters when assigning the infectious score. As medical personnel tell us, you can carry an infectious agent even if vaccinated ; We will then approach an asymptomatic infection. But as a reminder, we are not interested in the case of reinfection after recovery or vaccination in the immediate future.
- Indicate the proportion of positive or negative effects of each type of care ; with reference to a decrease or increase in the infectious score, the value of which must be specified. For simplicity, we propose to inform a fixed amount on the increase or decrease of the score. These two amounts can obviously be different.
- Indicate a confusion matrix for infectious tests, which can be simplified in the first instance. The question of the validity period of the tests would also be interesting to deal with. Especially since it can be included in territorial decisions, and therefore evolve during digitalization.
- Think about the quantitative design of the available elements, on time and geography.

3.4 Interesting details

3.4.1 Assignment of the infectious score

When initializing the infectious score of patients, it is interesting to specify several elements. It is necessary to check the validity of the use of the normal distribution : one does not wish to obtain a negative value for the infectious score. The concern lies in the assignment of thresholds with the objective of finding distribution statistics similar to the observed data.

What to do when the Gaussian parameters produce a significant proportion of negative value ?

There are two options to correct this problem :

- Reset the variable as many times until you get a valid value.
- Take the opposite of the negative value and somehow mirror the normal distribution with respect to the null value.

In both cases, it is necessary to re-examine the effective distribution of probability distribution. It is therefore necessary to add this problem in the search for optimal Gaussian parameters from observed epidemic data.

Even if the Gaussian law seems to be appropriate from a pragmatic point of view, we could leave the door open to the testing of other laws of distribution. Especially since it will always be possible to adapt the thresholds appropriately in an appropriate way.

3.4.2 Infectious severity flexibility

Another interesting point to note is that it is always possible to set the thresholds to remove the case of an asymptomatic infection. For example, the threshold between asymptomatic and standard patients can be set by the value zero. Conversely, it is possible to reproduce the absence of severe cases by defining an upper threshold that is too high. Even if the case is more anecdotal, we can also remove the case of standard patients by choosing the same value for the two thresholds separating infectious states. In summary, this design makes it possible to modulate the different infectious severities quite easily.

In this design, once the infectious score is initialized, a transition from standard to severe state can only be observed during an individual's infection. In all cases, the infection ends once its infectious score is zero, whether it is in asymptomatic, standard or severe condition.

3.4.3 Complexity of the asymptomatic case

There is some difficulty in digitizing an epidemic with regard to asymptomatic cases. How to know the real data of these patients ? It would be interesting to think about it for the continuation of the project. In any case, it is necessary to represent them to have a realistic conception : an epidemic is difficult to control without looking at individuals transmitting the pathogen and presenting no symptoms.

3.5 Memory management of individual states

In the pursuit of a detached implementation, it is more organized and intuitive to group the management of infectious states of individuals in the infectious layer. We would then find an organization with a hash table and keys composed of individual identifiers and different infections of the database. For each individual, it is necessary to distinguish two distinct states : the real one and the known one. Known state is more a matter of simple etiquette. The real one is more complex to model. It is necessary to implement and automate the update of the infectious state, taking into account the specificities of each state. It is therefore necessary to dissociate the management of the real state when it is in danger of the carrier disease, while automating the transitions between the states.

Chapter 4

Exploitation of digitalizations

4.1 Digitalization over time

The challenge of this project is based on the digitalization of the circulation of infectious agents in the mobility of a society. For this, it is necessary to initialize an initial situation of the company and to make it evolve day by day until a certain final point to be defined. The initial state can be obtained by initializing the environment with the various elements mentioned above and contaminate a certain number of individuals, deterministically or randomly. On the other hand, we can stop digitalization under different criteria :

- We can define a predefined number of days to digitize and evaluate the final state of the company. This is one of the important points in the objective of combinatorial optimization: to restrict the research space on time. From an algorithmic point of view, it seems obvious anyway to set a limit beyond which digitalization stops.
- With a view to optimization, we can establish shutdown conditions according to the situation of the company. To make a parallel with a video game, this could represent conditions of victory of the management. Conversely, we can also set conditions of defeat where the management is not sufficiently optimal to maintain a minimum state of society. Among the conditions for victory, we can identify good health assessments of society, where there is little contamination of infectious agents left. From a utopian point of view, one could say that the victory would be to completely eradicate infectious agents from society. However, this seems unimaginable given the external influence of the region whose sanitary conditions are affected by our management. As for the conditions of defeat, in the event of a serious economic crisis in society, digitalization can be stopped. Similarly, we can add health conditions that are too critical for society, whether in terms of mortality, contamination or available resources.

A digital observation unit represents the health and economic evolution of society and its individuals over a defined period of time. This period of time is defined as a fixed number of days, which remains the same throughout digitalization. It is at the end of each of its diagnoses that it is appropriate to modify or not the political strategy. From a realistic point of view, this seems all the more lucid because a policy cannot define a day-to-day strategy: it is necessary to take into account the adaptation of individuals. The project must be pragmatic to be interesting.

To best diagnose the status of the company, it is wise not to limit yourself to a final health and economic assessment at the end of each observation. Data from each observation can be expanded with daily health and economic assessments. This data is particularly relevant in the event that the days follow a calendar logic and where there is a trend on the dynamics of digitalization. Each observation would thus have a set of daily data to analyze in order to determine the appropriate decisions to be made for the next observation period. These diagnoses could be further explored in combinatorial optimization with reinforcement learning.

In practice, the implementation of this digitalization is not so simple to perform. It is necessary to group all the elements of the design mentioned above in a simple, readable and efficient implementation. It is necessary to record the results of the scans over the days. Data memory management can be enriched in static implementations. By associating the number of the day with the hash tables, we can analyze the results of dynamic digitizations, associated with each individual and space. The interest is to observe the times of occupation in the spaces and to evaluate the contaminations as well as the economic stability.

In the hash table of the spaces, there is a list of identifiers of individuals sorted according to their arrival time and associated with each space and day of digitization : we then find the identifier of the individuals who occupied the space, the role/importance of this movement as well as their exit time. On the side of the hash table of individuals, we find the updated scenarios by subtracting the duration of digitized journeys: we then find a list sorted with the time of arrivals, the identification of spaces and the time of exits.

In addition to this data, digitalization also records the health assessments of individuals in the infectious layer itself. The updating of the infectious states of individuals is also almost autonomous in this same layer. The idea is to decentralize the memory management of data to the appropriate layers. We therefore find the same principle with the management of data concerning territorial decisions.

As a general rule, it will always be interesting to keep the data of digitalizations, if only to be able to analyze the weaknesses of the design. We can always ask ourselves the question of the relevance of keeping data. The question is: do we really pretend to believe that the design is sufficient and does not need to be perfected ? However, the more we can trace the process, the more we can reason.

Already, a significant perspective of digitalization is the statistical and graphical representation of results. And by "results", it is implicit to extract the intermediate results of digitalization such as the dynamics of society.

4.2 Infectious assessment

4.2.1 Contamination

The infectious assessment is first calculated at the individual level, within the infectious implementation layer.

During digitalization, a susceptible individual accumulates infectious particles when in contact with carrier individuals or when in a particle-laden space. At the end of each day, for each susceptible individual, we can trace a sequence of spaces with data on :

- Occupancy times : precise arrival times, precise departure times.
- The identifiers of individuals sharing the same spaces during the periods of occupation ; We will then have done a scan to identify the patients carriers among them.
- The ability to facilitate proximity and contamination of the category of each space.
- The quantities of particles present in each space when the individual arrives, according to the particle conservation parameter of each infection in a space.
- The surfaces of each space visited ; This parameter will make it possible to directly nuance the contamination of the individual in a space loaded with particles while remaining pragmatic.

It is necessary to think more deeply about contamination outside a territory before talking about it. Other data are also interesting to integrate into the daily infectious assessment of an individual :

- Contamination in public transport : in the form of a simplistic score according to the type of transport and the frequency of their use by carrier individuals. We could see to detail this consideration from a more individual point of view. -
- The number of particles accumulated in each individual over the last few days according to the conservation parameter of each infection.

- The propensity of each individual, especially internal, to move outside the region. Two phenomena can be observed : the more frequently an individual circulates in the territory, the better we can estimate a random consideration outside ; And the less frequently an individual circulates in the territory, the more absurd it is to find a form of logic in external contamination.

The individual assessment therefore sums these variables weighted according to a generic formula to obtain an individual contamination score for each infection in the database. When at least one of these contamination scores exceeds the contamination threshold of one of the infections, which may depend on the age of the individual, he becomes a carrier of this infection and changes state to be in the incubation phase. As a reminder, it then becomes "immune" to other infectious agents in the environment as long as it is a carrier. Similarly, in a simplistic version, we are not interested in the case where an individual can be reinfected when he is vaccinated or cured : the individual also becomes "immune" to this infection. Of course, even if the object trace of a deceased individual may persist, it is useless to process their cases.

4.2.2 Evolution of states

When the individual assessment function indicates contamination or when the individual changes infectious status during the simulation, the person's infectious reflexes change and modify his mobility. It is necessary to update one's propensities in terms of care, unnecessary travel, etc. Typically, this phase is automatically triggered by the infectious implementation layer and passed indirectly to the dynamic digitization layer. It is agreed to apply a generalist and automatic operation to take into account these updates, on a case-by-case basis for each individual.

4.2.3 Territorial Assessment Scores

To summarize the infectious evolutions at the end of each day, it is necessary to implement specific functions where one places one's observation on society.

These functions depend on several variables to be weighted according to certain weights. For each infection, the following variables can be distinguished :

- The actual and known quantity or proportion of the susceptible population
- The actual and known quantity or proportion of the carrier population
- The amount or proportion of the population cured or vaccinated
- The amount or proportion of the population seriously ill
- The amount or proportion of the population that died
- Immediate change in actual and known number of infections
- A score on the available margin of care for standard patients
- A score on the available margin of care for the seriously ill
- etc.

These different evaluation scores will make it possible to describe the situation of a company, in a multitude of aspects. We would not limit ourselves to a simple score supposed to represent a pseudo-equilibrium, immutable whatever the conditions. We would then start on real multicriteria optimization.

4.3 Economic evaluation

4.3.1 Enterprises

The economic evaluation of spaces can be carried out on the basis of the affluence of individuals and characteristic specific to each category of spaces. There is a general evaluation function where variables must be inserted as parameters according to the category of the space. We can thus individualize the economic functioning of each space to have more concrete and applicable optimizations. For example, the evaluation function of an office will be different from a restaurant. And of course, some categories do not even have an economic function, such as public parks for example. In the notion of economic function, we are talking here about the opportunity of space to make the work of an individual productive in his environment. In any case, it would be just as interesting to study the question of economics properly today. Each space is therefore visited by a certain number of individuals, each with a defined role. Some may work in this space and have a financial dependence on it. While others may use it in another personal addiction, or otherwise as a hobby. The economic evaluation of a space therefore depends on data on :

- The category of space and its economic characteristics.
- The total duration of individuals who have worked in space.
- The number of individuals employed in space.
- The area of space.

Depending on the different types of economic functionality, we can add other data on :

- The number of people who used/visited the space.
- The total duration of individuals who worked from home. The question concerning the economic factor of production associated with telework is, however, to be discussed because it is highly subjective.

We can mention some different economic features such as :

- Offices whose economic dependence can be represented only with the occupation of space by employees. It would then be possible to arrange the management of companies by allowing their employees to telework from home. There is a direct and rather simple correlation between the total working hours of employees and the economic valuation of space.
- Businesses that have an economic dependence very sensitive to the presence of customers. The more individuals who visit these spaces, the more their economic valuation increases. Of course, this evaluation is also very correlated with the presence of employees who sell their products. The effectiveness of having several employees at the same time may be questioned. But we may be going too far in modeling. In a later version, like teleworking, it could be interesting to add a delivery option where businesses send their products directly to individuals. We will not be interested a priori in the terms of delivery.

4.3.2 Individuals

It is essential to also look at the economy from the point of view of the individual. We must not forget the objective of this project and not get too much into an individual problem where we seek to optimize everyone's income. The interest here is to evaluate the difference between the initial state of society and of each individual in relation to the states of restrictions. That is to say, we do not seek to put a score on the income of each individual but rather to observe the loss of working time of each in the event of an epidemic. This is again a debatable point of view since it is necessary to be able to reproduce the distribution of the work of individuals. At the scale of a society, it seems too difficult to reproduce reality. So the interest of the individual economic evaluation is to put a score by taking a step back at the scale of a population.

Note that there are very important elements to observe with this individual evaluation. For example, it will be wise to focus on the cases of children and medical personnel to provide analyses essential to epidemic management, including the stability of the education system.

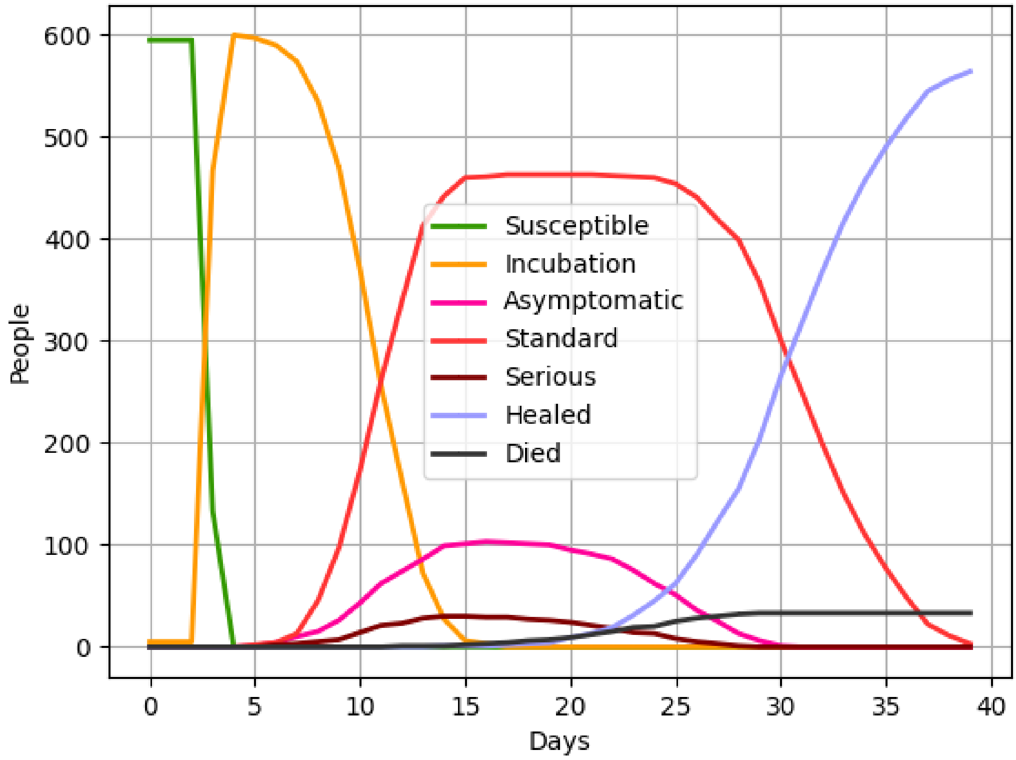
4.3.3 Territorial Assessment Scores

Like the global health assessment, it is necessary to be able to provide illustrative economic scores for each state of society, in order to make the best use of reinforcement learning. With the economy, the focus is on stabilizing production rather than optimizing it. This is why two priority elements must intervene in the overall evaluation : the average economic production of society and the dispersion of production between different spaces and individuals.

Chapter 5

Example of an illustrative application

For the sake of relevance, it is not interesting to worry about the very simple example of the implemented code. To contextualize the practical rendering, it is currently possible to implement static digitization using any openstreetmap file and specifying some parameters arbitrarily for the demography of the territory. Dynamic digitalization is very simplistic and carried out completely randomly. The infectious implementation allows automatic management of infectious transitions. The concrete rendering therefore corresponds to the digitization of the Belval university campus, in Esch-sur-Alzette, over a predetermined number of days, and with an infectious initialization to be arbitrarily defined. The basis of digitalization really needs to be consolidated, whether static or dynamic. This is just the beginning of the work. My last experiments were to play with the parameters of an infection, including the threshold of contamination, and to empirically observe the infectious evolutions of the population on a graph.



An illustrative example of infectious curves with the current implementation.

Quickly, we can get an overview of the design on the digitalization of an epidemic with the graph above. Without going into details, this test was carried out with cartographic data from the Belval campus in Esch-sur-Alzette. Six hundred individuals were implemented, over a period of forty days. Every day their dynamics are totally random and cover the entire network. So we can see it as a totally related graph where all individuals necessarily intersect. There are no outside considerations here. The test was carried out with five individuals in the incubation phase from day one.

There is very unrealistic contamination. Even if all individuals rub shoulders daily without exception, it seems irrational to see the entire population become infected so quickly; Hence the interest of implementing a form of elasticity in the contamination threshold. Indeed, the dynamic profile is almost identical for all individuals: we therefore find contamination scores too similar to observe a dispersion of infections. After, we can find a fairly well-known graph with the news. There is an exponential increase in sick cases as well as a "plateau". In reality, asymptomatic patients are not identified and the network of proximity between individuals is much less connected. We can of course review the proportions between infectious severities; no attempt has been made to replicate actual Covid-19 data. In addition, no care has been implemented, hence the fatality of individuals in serious condition and the increase in the number of individuals who die. Vaccinations have also not been implemented here.

The practical realization and everything that surrounds its production can be treated during a defense. We can then address the structural management of memory management, decisions for static digitalization or problem solving in the dynamics of society. It will be much more explicit and stimulating than evoking all the theoretical reflections.

Chapter 6

Environmental complexities

6.1 Overfitting of reality

By wanting to model too much, we can lose overall logic and robustness. Especially since the more we try to model in detail, the more we could be led to be interested in the personal data of individuals ; which is not the wish here.

We may eventually be interested in finer representations of society even if we are talking about a very high level of detail. This is the case, for example, with contamination during transport. This would therefore be directly related to the capabilities of SumoMobility modeling. For example, when an airborne infection emerges in society, no individual is prepared to protect themselves from contamination and avoid proximity. Proximity is an integral part of collective functioning but the use of sanitary masks is not omnipresent, especially at the beginning of the spread. The question to be asked is therefore whether it is really necessary to model ambient air contamination. However, we can draw a parallel with overfitting to make the link between a neural network and dynamic digitalization. SumoMobility is designed to walk from point A to point B without stopping. At a pinch, we could find a landmark by observing the collisions of individuals on the same pedestrian lane ; It would then be necessary to launch an interactive execution, which as a reminder can significantly cause errors and uncertainties in the scenarios between spaces.

From a realistic point of view, one could question the contamination within personal cars when the driver is accompanied. It seems absurd to approach such a level of detail. This would be a non-essential detail, which would ultimately diminish the robustness of the modeling. Especially since the people accompanying the driver are then very often linked by family ties or already rub shoulders regularly in the spaces we consider.

6.2 Evolution and competition of infections

In the environment, the coexistence of several infections takes place according to two different events :

- The random mutation of an infectious agent that already exists in the environment ; Each infection will then be characterized by parameters related to the frequency of mutation during a new contamination, and by its magnitude on its original characteristics. For simplicity, the appearance of a mutation would occur when an individual exceeds a contamination threshold and enters the incubation phase. According to Bernoulli's law, it can then become the host of a new infectious agent with characteristics similar to the infection it has contaminated. Among these new characteristics, it is possible to observe a contamination threshold different from the original, which would then be applied to the next individual that it contaminates. It is of course possible to remove the possibility of a transfer.
- The manual and programmable introduction of a new infectious agent, with characteristics completely independent of current infections. By deriving the concept a little, we can then manually program the variant of a pre-existing infection, thus mastering the new modified characteristics. This introduction can be programmed on a particular day in the simulation, or when society reaches a certain state.

Note that it is enough to reduce the contamination threshold of an infection to make the pathogen more contagious, reduce the threshold of the level of its serious condition to aggravate its virulence and increase the persistence time of particles in the body or in spaces to increase its resistance. It is therefore quite easy to evolve each infectious agent. This is what brings flexibility to the infectious implementation.

6.3 Diversity of society

The environment must make it possible to include individuals with unusual behaviours. The interest is to amplify the Brownian motion of the epidemic spread in society and to make the digitalization of society very unpredictable. For example, visitors can be represented in prisons in order to better initiate the spread in prisons. Although the individual assessment can be complicated, we can also represent delivery people who cross a lot of space in a day by staying there for very little time. In a similar type of daily life, we can also represent emergency services such as gendarmes, firefighters or emergency medical personnel. Not to mention, for example, gatherings between friends or family. As a general rule, it is necessary to be able to integrate a great diversity in the scenario of individuals, especially for unusual events. Examples include diplomatic sites, burial sites, nativity scenes, car garages, etc. The advantage of the space database is that you can choose from all the spaces listed, without neglecting some of them. It can therefore be interesting to introduce totally random scenarios, which can even go through several dwellings in the same day. A bit like real estate agents. This is certainly not rational but it can contribute to the Brownian motion of pathogens, and potentially bring a better elasticity of dynamics.

The project can therefore draw inspiration from the daily lives of individuals, while modeling non-ordinary behavior through completely random scenarios. We can dissociate two different representations :

- Either each individual would have a certain ratio of ordinary scenarios and totally random scenarios.
- Either we would rationally model a certain population whose scenarios would be totally random.

In these two options, the unpredictable and random factor is executed significantly differently. In the long run, digitalization can approach reality while removing the infinite complexity of modeling the diversity of everyday life. On the other hand, over a short period of observation, where one wishes to choose a decision according to the state of society, this can sufficiently influence the program by a very random parameter.

6.4 Evolution of parameters

In static digitalization, we can question the evolution of individuals. From a pragmatic point of view, it does not seem interesting in a primary version to model births and non-infectious deaths of the environment. On the other hand, it would be an interesting avenue for a demographic side project. Here the interest would be more to evolve the individual base of each. This would involve representing the probabilities of moving and the hiring statistics.

With the execution of dynamic digitalization, we can identify economic and health issues.

In terms of space, it would be interesting to remove bankrupt economic sites due to overly restrictive decisions. It would then be necessary to produce a critical evaluation from which space would be removed in the development of scenarios.

Many parameters are important to take into account in terms of infection. First of all, the evolution of available resources, which can represent an important space for definition. In addition, with care, testing and vaccination, we will choose the rational point of view to set their parameters on the duration of digitalization. But it is necessary to think about the parameters of these elements when adding them during execution: this is the case possible when a new mutation infection appears. All the thought to be done must be done on the manual programming of random or stochastic events. And in both cases, how could this approach reality ? How to find a form of prediction ? How to process historical epidemic data ?

More problems should appear noticeably in the further work of the implementation.

6.5 Impact of external territories

The debate to a greater or lesser extent was mentioned throughout this report. The themes to be interested are :

- Outdoor contaminations for internal individuals.
- Sanitary resources available outside.
- Demands for health resources from outside.
- Decisions on borders.
- External propensities in individual dynamics.
- Random parameters of actual infection states of external individuals.
- Evaluation of the known states of external individuals.
- The amount of external individuals to implement ; In other words, the amount of workers in the territory who come from outside, especially in terms of addition to internal workers. In the same vein, we must also consider internal individuals who work outside.
- etc.

In the immediate future, it is necessary to deepen the reflections.

6.6 Advanced scenario constraints

The problem of scenario design is complex. Not to mention the diversity of parameters, there are time constraints, route estimates, typical structures, etc. This complexity must be able to trace data observed in terms of affluence in spaces, respect socio-dynamic logics, etc. It's already something nice to deal with.

Now, to add realism and accuracy of decisions / optimizations, we can add other constraints, much less fun at first glance. We will not detail them further since the complexities are implicit :

- A limited amount of individuals present simultaneously in a space.
- The need to navigate a space and have to respect a specific available slot, in connection with seat limitations.
- The specific opening hours of certain categories of spaces.
- Individuals' preferences on selecting a space in a defined category, collectively to avoid getting closer to personal data.

6.7 Respect for barrier gestures

How to represent the respect of barrier gestures ? Is there really a typology of individuals who do not respect them in real life ?
And above all, does it really represent an interest in terms of territorial decisions ?

The implementation would require more time to think. But in the event that it seems sensible to consider it, we could bring a random factor, per day and per individual, which would influence infectious evaluations. It would increase the emission parameter of patients and decrease the threshold of contamination of susceptible infections. Respect for barrier gestures works both ways ! An interesting point is the typology of individuals. Are these the same individuals who tend not to respect these gestures or do we have to represent a random diversity to approach reality ?

6.8 Test, alert and protect

At the end of each day, we can identify all the people who shared a space at the same time. However, like the customers of a shop, it is not relevant, nor realistic, to hope to go up the exhaustive list of individuals. To make the link between the individual in the incubation phase and the contact cases, we can find a strategy that brings reality closer together : we would alert the individuals identified in the base of the new infected, and we would find the individuals who have shared a sufficiently long time with him. It would be necessary to specify several parameters :

- The number of days during which contact cases, not associated with the individual base, must be traced.
- The threshold of duration where we consider individuals in the vicinity as contact cases.
- The distribution of contacts during the search period ; whether the nearby individual has been in company with the newly infected person at once or several times in several spaces.
- The decision to be applied to contact cases : a constraint on travel, an obligation to test, etc. Note that this status can be added as an infectious state in its own right in a later version. Combinatorial decisions would then be attributed to Alerted individuals.

Chapter 7

Approval with real data

The idea is to use data that is healthy vis-Ã -vis humans. It is quite possible to model, reproduce and anticipate reality, only with the help of environmental data. One of the constraints of this is therefore not to use personal data such as GPS tracking. We can find a significant amount of healthy data in a territory :

- The age distribution of the local population.
- The distribution of household size.
- Some common affluences of spaces, depending on calendar parameters.
- The number of cars per household.
- The distribution of workers in the different occupational sectors.
- Mapping data as a whole ; See OpenStreetMap.
- Data on frontier workers.
- etc.

This data is intended to be global and requires less memory space. They can cover a large part of the environment.

The whole point is to determine the black boxes between them !
Finally, there are no checks as one could find in a neural network. There is no test data or validation. But the more healthy data you have, the more intuitive it is to find a link. In this research, the easiest way is to look at the causes and consequences of the real environment ; Processes can be identified. We can then run simulations with certain initial conditions and compare with data observed in reality. It is then possible to review little by little the parameters of digitalization to get closer to a reality ; while moving away from personal data and keeping a human ethic.

Chapter 8

Combinatorial decisions

It has not yet been mentioned that several decisions can be taken for the same period. The impacts of decisions can be significantly different and it is therefore possible to add some of them. One example is a decision on the validity of infectious tests, with a curfew from 6 p.m.

It might be interesting to consider non-compliance with these rules with a compliance rate in a later version: with each scenario creation, the individual could avoid these constraints according to a probability ; that is, this non- compliance would be applied regardless of the day and the individual, depending on the category of the individual. Of course, the scope of such non-compliance would depend on the seriousness/extent of the decision. But it seems absurd to characterize a subfield of action, which would apply specifically to a temporary niche of a scenario.

In the immediate future, the implementation is not constructive enough to focus more on decisions. It would produce aberrant results, which we would seek to optimize...

8.1 Constraints on individuals

To apply a specific decision to individuals, one must be interested in the external characteristics of the individual. The idea is to draw a parallel with reality. We cannot therefore use the socio-dynamic classes implemented in the environment. Individuals can then be separated by their ages, known infectious states and dynamic organization. Indeed, individuals whose work is a business can be specifically restricted ; This could then impact several socio-dynamic classes in the implementation for example.

8.2 Constraints on spaces

With spaces, however, it seems mandatory to use the categories in the database to apply specific decisions to them. We can then influence this selection by the area of each space to have a better classification in relation to reality. This is the case, for example, when you want to implement decisions on large shopping centres and leave small shops free. It is on the basis of this limitation of distinction that we can arrange the categories. For example, since originally the keys of the openstreetmap file are more detailed, we can redissociate the category of restorations by interior and exterior restorations.

8.3 Constraints on scenarios

The constraints on the scenarios are more complex since they themselves depend on static digitalization. They could be reduced to time constraints such as curfew or confinement as well as space selection constraints. Indeed, this is in line with the real case of distance limitations. As a reminder, the space database can filter places itself with a starting location and several criteria.

8.4 Constraints on infectious states

The constraints on infections seem obvious from an epidemic management perspective. In general, they focus on known, not real, conditions, as well as on how to access care, tests and vaccines. We can add other decisions such as the validity of the tests, as a whole to simplify. One of the difficulties is in the predetermination of care, tests and vaccines: we cannot predict these elements at the onset of infection. Perhaps by having a better look at medical processes, we could have an approximation on their onset times and their characteristics ; We can find logics between them. The question then arises about randomness.

Chapter 9

Parallelization of calculations

9.1 Initialization phase

Initialization is certainly the longest phase because of the complete reading of the openstreetmap file. This step is visibly difficult to parallelize. This is followed by the allocation of the nearest road for each space. This step takes significantly longer than reading the import file. However, it is possible to parallelize it quite easily : it is enough to simply distribute the local search of each space on several data centers.

Then, initialization follows with the creation of individuals, infections and infectious states, and possibly scenarios when you want to memorize them. These steps are, of course, parallel. It will be necessary to ensure to have a "master" to respect the predefined requests in the initialization.

9.2 Iterative phase

The iterative phase corresponds to the continuation of digitizations on the days. It begins with the creation/adaptation of scenarios, which can be parallelized as in the initiation phase. SumoMobility is then used to translate these scenarios into its format with duarouter. This step can be parallelized by the software itself by specifying a number of threads ; Note that we do not have access to parallelization since it is done internally. There is then a significant step which is to sort the routes according to the departure time.

Now, we can launch the execution of SumoMobility to digitize the dynamics of the company. As a reminder, we can internally parallelize this execution under several threads but we do not really know the details. Once the digitization is done, we then take care of the analysis of the results. You must first read a file and then save the data. The following treatments can almost all be parallelized since the calculations are attached to each space or each individual, etc. These interventions include economic and infectious assessments of individuals/spaces.

Chapter 10

Development prospects

10.1 Local modeling in a space

Modeling inside spaces can be interesting to implement. However, it would not bring a major evolution in digitalization.

From a practical point of view, most epidemic simulations model balls bouncing off a rectangle. We can draw a parallel with a chaotic circulation of individuals in the same space. One could even add impassable shapes in the rectangle to represent the structural architecture in space. For example, one could design the parallel shelves of a supermarket. But again, it is very likely to fall into an overfitting of digitalization, which could weaken the robustness of epidemic modeling in scenarios.

In reality, the dynamic interest of such a consideration does not make it possible to provide more assurance on the relevance of contaminations in a space. Especially since no data could really support the momentum ; And it would probably be personal data, which would be unsympathetic in the eyes of individuals.

On the other hand, from a research point of view, it would be very interesting to study the ability of a space to promote proximity/contamination of its occupants depending on its architecture. In particular, in our digitalization project, it would be more than interesting to better define the contamination factors of the different categories of spaces.

10.2 Exploring potential decisions

Even if research spaces would explode, we could also increase combinatorial choices by replacing territorial decisions with the constraints of digitalization.

10.3 Multi-territorial consideration

We have to be caring. The consideration here will never be to be in a negative competition with the neighboring territory.

The idea is to optimize all territories in a global consideration. The major difference from the original modelling is the separation/duplication of the territory into several regions, each with independent decisions. The term "independent" is, however, to be looked at closely. We can find a form of game theory here, except that we do not want negative competition. From the point of view of benevolent optimization, only positive competition, where both develop, is to be considered. It should be noted that the consideration of interterritorial movements would have to be modified.

10.4 Editing map data

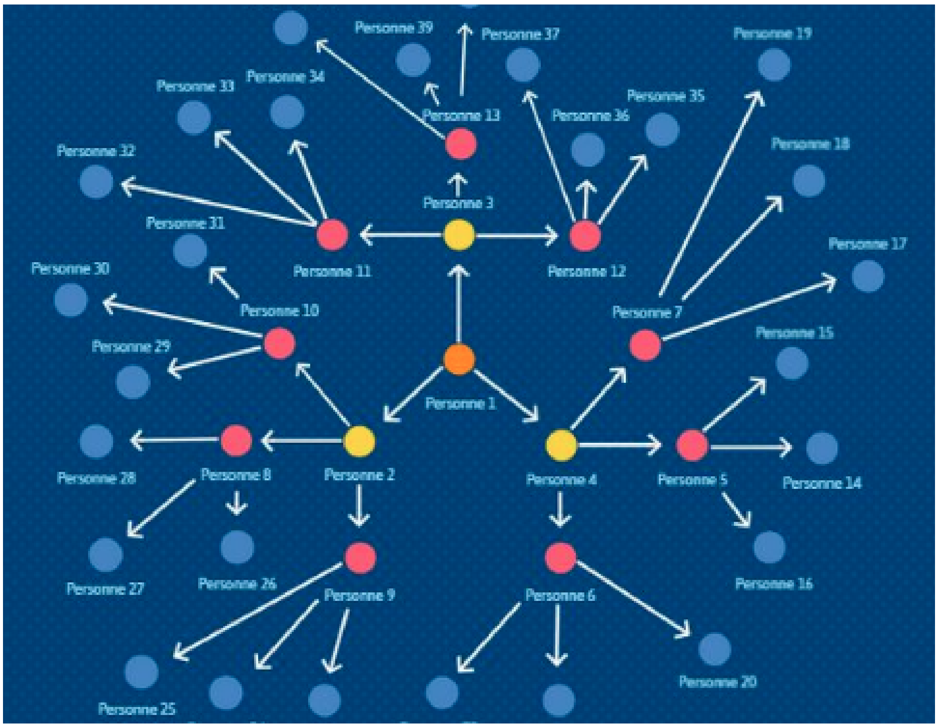
In a later version, it could be interesting to manually add spaces to extend the space for finding solutions to epidemic management. For example, there could be interest in building a new hospital, with a certain area and strategic location, to further optimize the management of infectious patients. We could then play on its location and area to observe its potential impact in digitalization. For this, we can use a Python package named Folium that would position the correct strategic coordinates of the new space by observing the map of the area studied. Its category and area would remain to be defined. It would then be necessary to determine the nearest route, although it would be possible to discuss the connections between the location of the space and those of the roads in the network.

A very wide variety of possibilities can be found with this new consideration. The research space could explode if we add strategic decisions in static digitalization.

10.5 Addition of the psychological criterion

One of the avenues for evolution is also the addition of the psychological dimension to the management of the epidemic. We can see this with the reality of Covid-19 ; There is an increase in depression and suicide as a result of the major restrictions. Then, of course, we can question the real origin of psychological suffering. Do they come directly from territorial decisions ? In any case, it is obvious that the introduction of major restrictions such as lockdowns have impacted the psychological states of individuals. We could therefore try to find a form of psychological evaluation on the continuation of the decisions set. Indeed, there seems to be greater consistency in analyzing decisions in their succession rather than over a given time. We can also add the psychological factor of certain categories of spaces in the dynamics of individuals.

10.6 In-depth analysis of results



Epidemic propagation pattern.

In order to have more details on the design and to perfect the modeling, it may be more than interesting to determine epidemiological scores. For example, we could be interested in calculating the R0 scores of infections ; that is, the reproduction rate of the virus in society. This would give more guidance to better compare and parameterize infections in the implementation.

In general, it would be interesting to collect more data on epidemic digitalization to be able to tend digitalization towards observed infectious parameters. It would be a bit like gradient backpropagation in learning a neural network, more manually here.