



UNIVERSITÉ DU
LUXEMBOURG

Rapport final

Optimisation combinatoire avec l'apprentissage par renforcement

Digitalisation d'une épidémie

Yann Cauchepin

Stage ingénieur

15.03.2021 - 31.08.2021

Sommaire interactif

Introduction

- Quel est le sujet du stage ?
- Un rapide tour sur l'environnement
- Quels outils principaux sont utilisés ?
- Une approche sophistiquée
- Pourquoi s'investir dans ce projet ?

Digitalisation statique de la société

- OpenStreetMap
- Espaces
 - Format *openstreetmap*
 - Base de données
 - Attributs avancés
 - Gestion des distances
- Individus
 - Constitution des données statiques
 - Représentation de la diversité individuelle
 - Base de données
 - Considération extérieur
 - Socle individuel

Digitalisation dynamique de la société

- Scénarios des individus
 - Plusieurs hyperparamètres
 - Comment s'y prendre ?
 - Problématique du temps
 - Gestion mémoire
- SumoMobility
- Réseau routier
- Mouvements interterritoriaux

Dynamique des agents infectieux

- Schéma infectieux
- Transitions des états
 - Susceptible
 - Phase d'incubation
 - Malade asymptomatique
 - Malade standard
 - Malade grave
 - Guérit & Vacciné
 - Décédé
- Soins, Vaccins et Tests
- Détails intéressants

- Affectation du score infectieux
- Flexibilité de la gravité infectieuse
- Complexité du cas asymptomatique
- Gestion mémoire des états individuelles

Exploitation des digitalisations

- Digitalisation sur le temps
- Evaluation infectieuse
 - Contamination
 - Evolution des états
 - Scores d'évaluations territoriales
- Evaluation économique
 - Entreprises
 - Individus
 - Scores d'évaluations territoriales

Exemple d'application illustratif

Complexités de l'environnement

- Overfitting de la réalité
- Evolution et concurrence des infections
- Diversité de la société
- Evolution des paramètres
- Incidence des territoires extérieures
- Contraintes avancées des scénarios
- Respect des gestes barrières
- Tester, alerter et protéger

Homologation avec des données réelles

Décisions combinatoires

- Contraintes sur les individus
- Contraintes sur les espaces
- Contraintes sur les scénarios
- Contraintes sur les états infectieux

Parallélisation des calculs

- Phase d'initialisation
- Phase itérative

Perspectives de développement

- Modélisation locale dans un espace
- Exploration des décisions potentielles
- Considération multi-territoriale
- Modification des données cartographiques
- Ajout du critère psychologique
- Analyse approfondie des résultats

S'interroger sur l'intelligence

Une projection intéressante

Remerciements

Introduction

Quel est le sujet du stage ?

Ce stage a pour but d'utiliser l'apprentissage par renforcement afin d'optimiser une gestion de crise épidémique. Cela inclut donc la conception de l'environnement concernant la propagation épidémique et l'application de l'apprentissage par renforcement sur ses évaluations. Cette optimisation est multicritère puisqu'elle a pour objectif de dissiper ou limiter les échanges infectieux tout en stabilisant l'économie du territoire étudié.

Ce projet se focalise sur la société luxembourgeoise et est codé en langage Python puisqu'on y retrouve beaucoup de bibliothèques déjà intéressantes pour le développement du programme.

Dans un souci de perfectionnisme, le développement de l'environnement a pris du temps et beaucoup d'éléments restent encore à être implémenter au programme. Il a été nécessaire de tout imaginer sans guide, en intégrant l'outil *SumoMobility*, pour avoir une conception réaliste, efficiente et qui peut s'adapter au fur et à mesure des besoins et des perspectives d'évolutions.

Au terme de ce stage, seule une partie simple de l'environnement est implémentée. Il serait nécessaire d'avoir davantage de temps pour rendre l'environnement plus complexe/réaliste, adapter les résultats simulés avec ceux des données réelles, pouvoir déterminer des actions de contrôle sociétale et enfin optimiser ces décisions selon plusieurs critères et contraintes.

L'idée étant de réfléchir étape par étape, perfectionner et rendre l'environnement de plus en plus réaliste avec les essais de conception. C'est la base même de l'optimisation alors c'est le plus essentiel !

Un rapide tour sur l'environnement

On s'intéresse ici aux épidémies qui se transmettent par voie aérienne. La conception de l'environnement pour ce projet est peu adaptée aux autres moyens de transmission comme les maladies sexuellement transmissibles tel que le SIDA.

La propagation d'agents infectieux dans la société ne peut pas être prédite à priori. Il est nécessaire d'attendre la transformation de l'environnement pour caractériser les paramètres d'une épidémie. On identifie d'ailleurs une épidémie qu'après avoir découvert des cas infectés dans la société et avoir fait le lien avec l'émergence d'une nouvelle épidémie. L'évolution d'une épidémie à transmission aérienne s'étend géographiquement de manière assez chaotique, avec l'entropie dynamique des individus *porteurs* dans la société.

Pour modéliser cet environnement, il faut décomposer la société par des individus et des espaces/lieux de passage. Chaque individu circule dans la société à travers différents espaces dans son quotidien. Les échanges infectieux ont lieu avec la proximité des individus *sains* et *porteurs*.

On peut distinguer plusieurs échelles dans la représentation de l'environnement. Très localement, il est trop aberrant de représenter la propagation des agents infectieux au sein de l'organisme d'un individu, notamment en considérant la complexité et la diversité des organismes. Mais nous verrons qu'on peut avoir une approche très simpliste du fonctionnement infectieux d'un organisme en considérant la dualité entre les antigènes et les anticorps. A l'échelle d'un espace clos ou semi-clos, il est assez accessible de représenter les échanges infectieux entre les individus. Après, plus globalement, on retrouve deux échelles qu'on pourra dissocier au niveau régional puis à celui national.

Dans une optique d'optimisation sociétale, ce projet centralise son approche sur une échelle régionale. On prend un territoire spécifique contenant plusieurs espaces et on synthétise les échanges infectieux dans ces différents espaces. Dans une version plus développée, on pourra étendre le projet à une représentation multi-territoriale ou nationale, où les décisions peuvent être différentes en fonction des régions. En parallèle, on pourra aussi s'intéresser à la mobilité des individus à l'intérieur de l'architecture des espaces. Les perspectives d'évolutions sont abordées plus en détails à la fin de ce rapport.

Quels outils principaux sont utilisés ?

Le point de départ du programme est la description d'un territoire par un fichier *openstreetmap* qui permet d'obtenir les espaces d'intérêts d'une région ainsi que les pistes pour la circulation des individus. La description de ces fichiers se fait par des contributeurs et est relativement bien détaillée. A noter que la véracité des données décrites ou leur mise à jour sont parfois discutables. Dans notre utilisation pour des régions du monde riches et populaires comme le Luxembourg, les fichiers sont bien plus que suffisants.

Le second outil est *SumoMobility* qui est une implémentation dédiée au génie civil pour digitaliser la circulation urbaine avec plusieurs types de transports distincts. Le programme peut directement utiliser un fichier *openstreemap* de la région en entrée pour obtenir le réseau routier. C'est un outil très complexe et intéressant pour reproduire la mobilité d'une société. Toutefois, l'une des difficultés est qu'il est très orienté autour du trafic et non pas des individus. Il est donc nécessaire d'adapter notre implémentation. Cet outil a l'avantage d'être populaire et peut fournir une complexité suffisante à notre programme avec les contraintes de mobilité et de trafic de la société. A noter d'ailleurs qu'il existe des études de recherche sur les demandes de trafic et les statistiques de circulation qui peuvent être intéressantes pour approfondir le projet. On pourra citer par exemple *Luxembourg Sumo Traffic Scenario (LuST)* produit par *Laura Codeca*, *Raphael Frank* et *Thomas Engel*, depuis l'Université du Luxembourg.

Ces outils seront plus détaillés au fil du rapport.

De même, on peut noter l'utilisation d'une librairie préexistante dans Python pour l'apprentissage par renforcement nommée *rllib*. Cette une implémentation très complète mais qui ne sera pas développée dans ce rapport par défaut de temps. Bien entendu, le projet utilise d'autres librairies standard de Python qui ne s'apparentent pas à de réels outils.

Une approche sophistiquée

Avant de parler d'optimisation multicritère, il est nécessaire de travailler sur la conception de l'environnement dans le détail. C'est la base du projet : plus l'environnement sera perfectionné et plus son optimisation sera applicable et réaliste. C'est d'ailleurs de ce point de vue que mon stage s'est orienté sur le développement de l'environnement. Par souci de perfectionnisme, il est essentiel de bien réfléchir aux emboîtements des conceptions/idées pour pouvoir décomposer l'implémentation et anticiper au mieux son utilisation pour des sujets annexes ou prévoir des perspectives d'évolutions.

De ce fait, on distingue cinq principales couches dans l'environnement : les individus, les espaces, la mobilité, les agents infectieux et les décisions combinatoires. Chacune d'entre elles sera détaillée dans les parties suivantes.

Dans le détail, on utilise des données cartographiques et démographiques pour initialiser le territoire qu'on étudie. Cette phase est essentielle et il est aberrant de déterminer une optimisation sur un environnement peu connu. C'est d'ailleurs pour cela qu'on parlera plus de "digitalisation" que de simulation. Plus précisément, on nommera cette phase comme étant la digitalisation statique de la société, qui regroupe les individus et les espaces.

Dans un second temps, on se préoccupe de la digitalisation dynamique des individus dans les différents espaces. C'est une phase complexe. C'est ici que l'on peut trouver un parallèle avec un réseau de neurones. A partir d'un environnement initial composé d'individus et d'espaces, on doit pouvoir évaluer la dynamique de la société sur un grand ensemble de journée ayant chacune des caractéristiques différentes. C'est comme une boîte noire où l'on retrouve les individus et les espaces en entrée, et les occupations détaillées des espaces en sortie.

Alors pourquoi ne pas utiliser simplement un réseau de neurones pour prédire les dynamiques de la société ?

Le concept même d'un neurone est d'assimiler plusieurs données, les traiter et les résumer en une information simple. Ça n'a rien d'intuitif par rapport à l'élaboration d'un bilan dynamique complet pour chaque individu et chaque espace. Alors à priori on pourrait résumer la mobilité et se contenter de statistiques plus ou moins simple. On peut retrouver l'utilité d'un réseau de neurones qui est de résoudre un objectif de classification ou de prédiction d'un ensemble de données sans contexte. Mais si on réussit à approcher la réalité de la

société et observer une réelle digitalisation dynamique, on peut avoir nettement plus d'informations pertinentes et non modifiées. Notre traitement des données est beaucoup plus précis et ne laisse que très peu de marge à l'erreur. On ne fait pas une recette de pâtisserie en jouant aux dés ! C'est la partie complexe de tout projet qui est de travailler sur le détail des choses. Lorsqu'elle est bien réalisée, on ne laisse qu'une infime incertitude sur le résultat final. Dans une optique d'optimisation multicritère des évaluations, les détails sur l'environnement sont donc plus qu'essentiels.

A la question comment faire pour rendre la digitalisation dynamique fiable et réaliste ?

Là est tout le problème, et l'enjeu de la conception de l'environnement. La partie concernant la mobilité des individus dans les espaces est traitée plus bas dans le rapport. On verra qu'on peut retrouver une forme d'entraînement de cette digitalisation comme on retrouve dans un réseau de neurones. Bien entendu, cet entraînement n'est pas obligatoire ici puisqu'on peut déjà reproduire une mobilité totalement aléatoire et obtenir une dynamique des individus. C'est d'ailleurs le résultat actuel du stage mais il faut bien prendre en compte que c'est aberrant. Mais cela a au moins le mérite de représenter le potentiel de ce projet.

L'entraînement pourrait être réalisé avec les demandes de trafic, des données observées sur l'affluence des lieux par exemple. On pourra s'intéresser à un traçage des individus via la couverture réseau mais cela pose des problèmes éthiques sur l'individu. Pour rappel, l'intérêt ici n'est pas d'espionner ou de manipuler les individus via leurs données personnelles ; le projet s'intéresse à prédire les déplacements d'un territoire pour optimiser les risques sanitaires et maximiser la stabilité économique. Cela doit être bienveillant pour les individus et la population en général.

A partir de là, il est intéressant de dissocier la digitalisation de la société et la propagation infectieuse. Pour des projets annexes bienveillants ou ne serait-ce que pour simplifier l'entretien de l'implémentation, il est intéressant de pouvoir digitaliser un territoire indépendamment d'une considération épidémique.

La couche concernant les agents infectieux est un ajout de la digitalisation. Bien entendu, elle apporte des caractéristiques spécifiques aux individus, aux espaces et à la mobilité. L'intérêt est de ne pas rendre ces paramètres bloquants pour une simple digitalisation sociétale. On verra qu'on peut également entraîner les paramètres de chaque agent infectieux avec des données observées. Toutefois, comme il a été noté précédemment, on ne peut caractériser les paramètres d'un agent infectieux qu'à posteriori. Tout l'intérêt est de perfectionner cette

adaptation avec des données historiques telles que la Covid-19. C'est en s'entraînant qu'on devient expert !

L'idéal de ce projet est de proposer le plus rapidement et le plus pertinemment possible une suite de décision optimale pour la société, en produisant une digitalisation territoriale réaliste et bienveillante envers l'individu.

Il est évident toutefois que la pertinence de l'utilisation des données infectieuses observées est grandement dépendante de l'utilisation des données dynamiques de la société. On ne peut pas approcher convenablement la réalité d'une propagation épidémique sans avoir une base robuste et stable du comportement dynamique d'une société.

Alors sur quoi reposent les décisions combinatoires ?

On ne peut pas optimiser un tel environnement en laissant un large espace de recherche. Il est nécessaire de définir un nombre prédéfini de décisions, avec leurs contraintes détaillées dans l'environnement. Ces décisions impactent la digitalisation dynamique de la société et modifient la mobilité de la société. On observe des différences sur la suite d'espaces parcourus par chaque individu ainsi que sur les affluences de ces espaces. Ce sont à partir de ces données que des scores d'évaluations sur plusieurs critères différents sont établis. Ces scores forment le support de l'optimisation combinatoire multicritères qui établit la suite de décisions la plus adéquate possible.

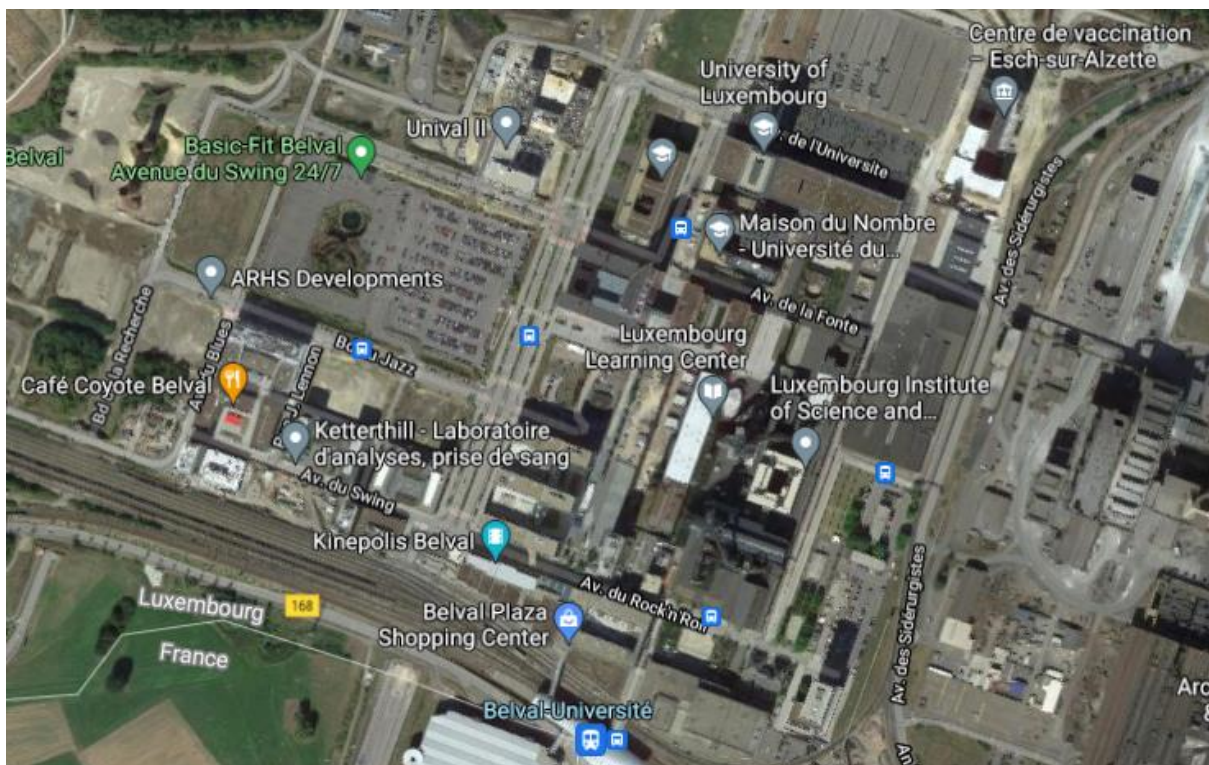
Les décisions combinatoires forment ainsi une dernière couche du programme et ne sont pas essentielles pour une digitalisation sociétale simple. L'une des problématique est donc de coder la digitalisation statique et dynamique de la société, tout en laissant une modification possible de la part d'une couche supérieure.

Pourquoi s'investir dans ce projet ?

On peut d'ores et déjà utiliser des simulations simplistes pour prédire l'évolution sanitaire d'une épidémie. On pourra citer par exemple l'expérience présentée avec le modèle SIR dans la présentation suivante :

<https://www.youtube.com/watch?v=gxAaO2rsdIs>

Avec ce projet, nous nous proposons d'apporter une approche réelle de la société : on souhaite considérer les lieux de passages propres à chaque individu, prédire les comportements sociétaux et étudier les échanges infectieux selon différents paramètres. Dans une région définie, on utilise des données réelles telles que des données cartographiques, démographiques, épidémiques et bien d'autres. L'objectif de ce projet est d'avoir une optique d'optimisation multicritère d'une gestion épidémique. On souhaite agir avec bienveillance envers l'individu, limiter les risques sanitaires et maximiser la stabilité économique d'un territoire. On définit ce territoire sur le Luxembourg, et plus précisément sur le campus Belval à Esch-sur-Alzette pour ce stage. Mais ce projet est destiné à s'adapter à n'importe quel territoire.



Une vue satellite du campus Belval à Esch-sur-Alzette depuis Google Maps.

Pour parler en profondeur des motivations d'un tel projet, il est plus que nécessaire de parler de principes moraux et de bienveillance humaine, surtout

lorsqu'on parle d'intelligence artificielle. On sort du cadre technique d'un stage ingénieur mais il est intéressant d'en parler en annexe à la fin de ce rapport.

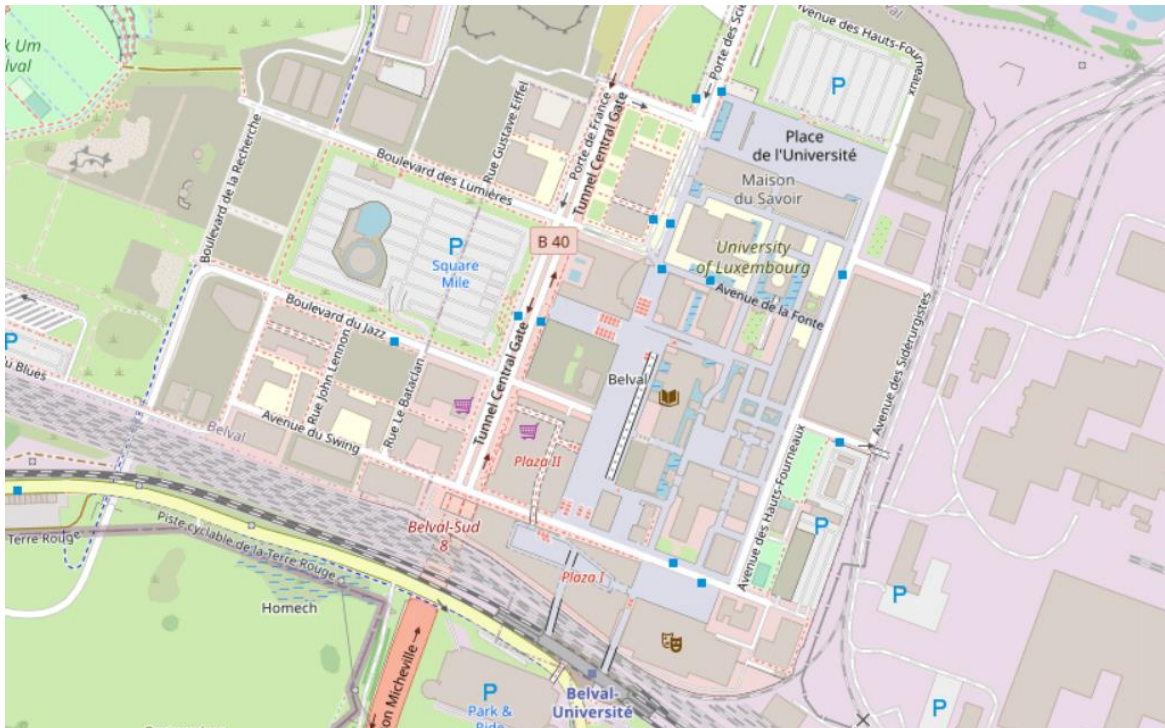
Historiquement, on peut classer les catastrophes naturelles, environnementales et les crises épidémiques parmi les plus grandes catastrophes humanitaires. Quels autres problèmes sont plus essentiels à optimiser ? On ne compte pas les guerres puisqu'il y a un souhait volontaire de la part de l'homme de détruire son prochain, ce qui ne doit pas être optimisé au sens éthique. Les enjeux des catastrophes humanitaires sont les plus importants au monde, qu'ils soient humains, économiques, psychologiques, etc. A ces niveaux, les moindre optimisations aux plus faibles échelles sont toujours capitales.

Dans notre cas ici des crises épidémiques, le sujet est d'actualité avec la Covid-19. L'actualité nous démontre tous l'importance d'optimiser la gestion d'une crise épidémique. D'un point de vue historique, aucune technologie ne permettait à l'époque des anciennes crises de concevoir une optimisation des décisions. En se rappelant des dégâts épouvantables des précédentes crises sanitaires, ils semblent évidents de devoir se préparer au mieux pour optimiser les prochaines épidémies/pandémies. Dans l'immédiat cela permettrait également de soutenir les décisions actuelles dans la gestion de la Covid-19.

Comme cité précédemment, en décomposant les éléments de ce projet, on peut également trouver d'autres intérêts tels que la modélisation d'une société et de sa dynamique, indépendamment d'une propagation d'agents infectieux. De plus, au-delà des nouveautés dans la digitalisation d'une épidémie, on trouve également un intérêt en s'intéressant à optimiser plusieurs critères dans la gestion épidémique : il sera intéressant de réfléchir à chercher d'autres pistes d'optimisation multicritères avec l'apprentissage par renforcement, qui ne sont pas déjà traités dans la recherche. Parmi ces pistes d'exploitation, on pourra s'intéresser à l'environnement complexe et à la possibilité de décrire précisément des actions dans l'environnement. Avec ce projet, il est aussi possible de détailler de nombreuses évaluations avec des critères très différents.

1. Digitalisation statique de la société

1.1. OpenStreetMap



Un aperçu cartographique du campus Belval à Esch-sur-Alzette sous le format openstreetmap.

Pour concevoir une digitalisation statique de la société la plus authentique possible, le projet base sa conception sur l'utilisation d'un fichier *openstreetmap* du territoire étudié. Tout part de ce point de départ. L'idée étant d'utiliser les lieux et le réseau routier réel d'une zone déterminée pour initialiser les espaces, et avoir le support des chemins entre les différents lieux.

Pour faciliter l'exécution du projet, il est donc intéressant de pouvoir récupérer le fichier *openstreetmap* directement depuis le programme. Pour cela, on renseigne plusieurs informations telles que les bornes représentant les latitudes et longitudes minimales et maximales dans une implémentation Python nommée *Overpy*.

L'un des petits inconvénients que représente *openstreetmap* est la délimitation du territoire étudié. A priori, il est possible de récupérer une région en formant un polygone, le plus souvent rectangulaire mais ce n'est pas intuitif dans notre usage

ici. En particulier pour utiliser des statistiques démographiques par exemple, qui sont appropriées pour une certaine délimitation d'un territoire. Il serait nécessaire de vérifier les possibles épurations déjà existantes avant d'approfondir cette approche mais il semble judicieux de retransformer le fichier récupéré en filtrant les éléments qui sont contenus dans un périmètre bien défini : on peut parler de limite "administrative" des territoires où l'on regarde les frontières officielles entre les différentes régions. De manière générale, il serait très intéressant d'évaluer la position des éléments par rapport à une délimitation périphérique. S'il n'existe pas d'attribut dans l'élément qui indique sa localisation administrative, il serait nécessaire de modéliser la délimitation par une suite de points avec des longitudes et latitudes acceptables et de vérifier élément par élément. Il existe déjà probablement des algorithmes pour cette filtration.

Un autre point intéressant à regarder est la proximité d'un élément par rapport aux délimitations du territoire. C'est déjà plus complexe que de savoir si l'élément est contenu dans le périmètre et demande donc un peu plus de travail. Un élément situé à proximité d'une frontière peut avoir un traitement différent dans la digitalisation mais nous verrons ce point un peu plus tard.

Dans une version ultérieure, il pourrait être intéressant de voir un intérêt à décomposer une grande zone géographique en une mosaïque de plusieurs petites zones afin de paralléliser la digitalisation statique de la société.

1.2. Espaces

1.2.1. Format *openstreetmap*

Initialement décrits par un fichier *openstreetmap*, les espaces ont des caractéristiques propres au format du fichier. Il est donc nécessaire de lire l'ensemble des éléments et de retravailler les caractéristiques de chacun des espaces. Un espace est avant tout décrit par un ensemble de nœuds qui ont chacun une latitude et une longitude précises. Cet ensemble de nœuds forme un espace clos qui délimite les contours de chaque espace. En retravaillant simplement les données, chaque espace peut être caractérisé par des attributs courants :

- Un identifiant unique, tiré du fichier *osm*.
- Les nœuds qui le délimitent ; inutile dans une version simple mais qui peuvent éventuellement présenter un intérêt d'ordre perfectionniste.
- Sa localisation moyenne, calculée avec la moyenne de ses nœuds ; le résultat est discutable en terme géométrique mais bien suffisant pour résumer la position de l'espace.
- Son aire, calculée avec ses nœuds ; le calcul d'une aire d'un polygone quelconque avec des coordonnées géographiques est très imparfaite et nécessite l'utilisation du package *Shapely*.
- Les étiquettes du fichier original pouvant apporter des spécifications réalistes aux espaces ; on retrouvera la classification originale de l'espace sous *openstreetmap* et parfois des informations amusantes comme le nom de la structure, etc.

Le lien suivant fournit davantage d'informations sur le format *openstreetmap* : https://wiki.openstreetmap.org/wiki/FR:C3%89l%C3%A9ments_cartographiques

1.2.2. Base de données

La lecture des éléments du fichier est séquentielle et n'est pas parallélisable à priori. La structure du fichier est décomposée en nœuds, chemin et relation, empêchant le traitement indépendant de sous ensemble du fichier. Bien entendu, comme précisé précédemment, il est possible moyennant un prétraitement d'importation géographique de décomposer le territoire en plusieurs sous-territoires pour les lire indépendamment.

Il est évident d'enregistrer les espaces dans une base de données pour l'utilisation de la digitalisation. Au lieu d'enregistrer simplement les éléments intéressants du fichier, il convient donc de réfléchir à la gestion mémoire de ces espaces.

Dans le format *openstreetmap*, les espaces sont classés selon une clé et une valeur selon le type du lieu. Pour digitaliser au mieux une épidémie, l'implémentation a reclassé les différents types de lieu afin de mieux regrouper les espaces selon leur rôle sociétal, leur fonction économique ou leur aptitude à accentuer la contamination de leurs occupants. L'idée sous-jacente est de traiter les espaces d'une même catégorie avec des paramètres communs. La nouvelle classification est décrite par des types, eux-mêmes constitués en sous-types.

Le rôle des types d'espaces est plus affilié à une organisation cohérente de la base de données. Ils permettent principalement de regrouper logiquement des sous-types, ce qui peut être utile par exemple pour traiter des sous-types de la même manière.

Chaque sous-type a donc des caractéristiques communes qui permettent par exemple d'ajuster les évaluations sanitaires et économiques de chacun des espaces qui leurs sont associés. Ces paramètres permettent d'adapter la valeur retournée par les fonctions sanitaire et économique des espaces.

1.2.3. Attributs avancés

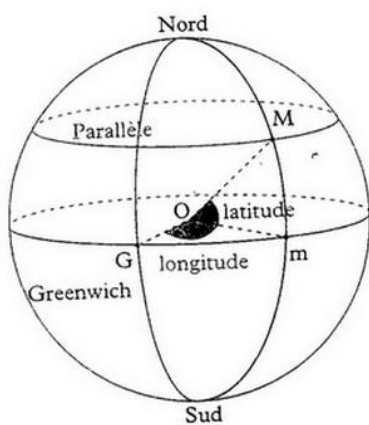
On peut aisément ajouter des attributs communs aux espaces, aux sous-types ou même au types d'espaces eux-mêmes. Dans la plupart des cas, il est préférable de définir les attributs dans les sous-types par un dictionnaire Python. Leurs utilisations en pratique pourront être modulé moyennement les attributs simples des espaces tel que l'aire. C'est le cas par exemple des facteurs d'évaluation portés sur l'économie ou les risques sanitaires. Après, on peut davantage spécifier les sous-types en réduisant les rôles que leurs espaces peuvent jouer. L'explication des rôles sera donnée avec la digitalisation dynamique.

En termes d'attributs, on pourra étendre les perspectives d'évolutions en ajoutant par exemple :

- Horaire d'ouverture et fermeture ; ces horaires doivent être recevables et logiques par rapport aux catégories des espaces.
- Temps de passage moyen des individus ; à nuancer avec des paramètres aléatoires et en dissociant les différents rôles des individus.

- Obligation de présenter un état connu valide ; selon la validité d'un test sanitaire ou autre.
- Possibilité de télétravail et le pourcentage de performance associé ; on ne considère que les heures de travail en terme économique et ce paramètre est donc exclusif aux espaces possédant un rôle de travail.
- Limite de place des occupants en même temps ; en ne considérant pas les individus avec le rôle de travailleurs et en nuanciant avec l'aire des espaces.

1.2.4. Gestion des distances



Dans la modélisation, les espaces permettent d'ajouter la conception de distance à la digitalisation de la société. Toutefois, toutes les données géographiques sont données en latitude et longitude. Deux méthodes permettent de calculer la métrique à partir de ces données. Il existe un package *Shapely* en *Python* qui permet d'avoir des résultats précis mais qui prend davantage de temps à s'exécuter par rapport à la méthode simple mathématiques.

Shapely est donc appliquée pour obtenir des données exactes à enregistrer telles que l'aire des espaces. On préférera utiliser la méthode simple pour comparer des distances. C'est le cas par exemple pour sélectionner l'espace le plus proche à partir d'une localisation. On pourra également filtrer un ensemble d'espaces en fonction de leur distance relative à une localisation, mais aussi selon leur aire ou leur catégorie. Du fait d'une certaine incertitude, les résultats peuvent être significativement différents de la méthode longue *Shapely*. Il conviendra donc d'ajouter empiriquement un facteur à la méthode mathématique pour l'ajuster à celle longue.

1.3. Individus

1.3.1. Constitution des données statiques

L'utilisation des données cartographiques d'un fichier *openstreetmap* ne permet pas de constituer le patrimoine démographique d'un territoire. Il est nécessaire de trouver des stratégies pour identifier le nombre d'individus à initialiser ainsi que de les caractériser de façon adéquate.

La meilleure méthode est d'associer la délimitation d'un territoire *openstreetmap* avec des limites administratives d'une région pour pouvoir utiliser au mieux les données démographiques locales. Si cela n'est pas possible, on pourra alors analyser le fichier *openstreetmap* lui-même et trouver des corrélations solides entre le nombre d'individus et des éléments du fichier. Par exemple, on pourra chercher une régression linéaire entre la taille totale du fichier et la population de la région. Sinon, en analysant plus en détail, on pourra s'intéresser à l'aire totale des logements du fichier pour en déduire la quantité d'individus. Il y a beaucoup d'hypothèses possibles qui ne demandent qu'à être vérifiées. Bien entendu, il est toujours possible de définir un nombre d'individus manuellement, indépendamment des données cartographiques.

Dans notre cas ici du stage, le Grand-Duché du Luxembourg met à disposition un ensemble de données publiques très intéressantes sur un site internet : <https://data.public.lu/fr/>

Dans le cas de cette section, cela concerne les données démographiques ou les données sur les ménages par exemple. Pour représenter au mieux la population dans ce projet, nous disposons donc de la population totale par région luxembourgeoise en 2021.

1.3.2. Représentation de la diversité individuelle

Comment apporter la diversité statique des individus ?

En particulier dans le cadre d'une étude épidémique, il faut pouvoir représenter la diversité des organismes en termes de résistance, de fragilités, etc. Et du point de vue simple de la société, il est essentiel de distinguer les individus avec des dynamiques sensiblement différentes. Par exemple, des élèves n'auront pas la même dynamique que des retraités.

On peut toutefois trouver un facteur commun plutôt simple : l'âge. En nuancant avec des classes d'individus ayant des comportements similaires, il semble intuitif de caractériser un individu par son âge et sa catégorie socio-dynamique.

Il est également accessible de connaître la répartition de la population luxembourgeoise en fonction de l'âge, de 0 à 95 ans. Dans un souci de simplicité, on ajoute à l'initialisation la proportion de la population luxembourgeoise ayant plus de 95 ans à celle de 95 ans. La population de l'implémentation en l'état actuel est donc répartie entre 0 et 95 ans. A priori, il semble non productif de s'intéresser à l'évolution des âges de la population, des naissances et des mortalités non infectieuses. Mais on peut toutefois laisser la porte ouverte à ces pistes d'explorations pour d'autres projets bienveillants concernant l'évolution démographique par exemple.

1.3.3. Base de données

A l'instar des espaces, les individus sont classés en types et sous-types. On retrouve le principe de dictionnaire pour répertorier des caractéristiques communes.

Un sous-type d'individus possède plusieurs attributs de base :

- L'âge minimum et l'âge maximum des individus qui le composent.
- Les catégories des logements dans lesquels ses individus peuvent dormir.
- Les catégories des espaces dans lesquels ses individus peuvent travailler.

Du côté des individus, chacun d'entre eux serait caractérisé par :

- Un identifiant unique.
- Un âge.
- Un ensemble de données correspondant au socle personnel.

Comment répartir les âges des individus dans la digitalisation ?

C'est un problème ambiguë. D'un côté nous avons les classes socio-dynamiques des individus et de l'autre la répartition des âges d'un territoire. On pourra remettre en cause les choix de l'implémentation mais la solution choisie semble convenir dans un premier temps :

A l'initialisation, on répartit arbitrairement le nombre total d'individus à travers tous les types existant de la base de données. On a donc un sous-nombre d'individus à répartir entre les différents sous-types dans chacun des types. C'est là qu'on utilise alors la répartition des âges du territoire, en l'occurrence le Luxembourg. On ne prend que l'intervalle d'âge identifié dans le sous-type et on détermine les proportions par rapport à un total de 100 %.

Dans le cas d'un ensemble enchevêtré de sous-types dont les intervalles d'âges se chevauchent, on préférera respecter la répartition du territoire sur l'ensemble du type d'individus. De ce fait, on récupère d'abord l'ensemble des intervalles d'âges des sous-types en un même intervalle, repropportionné sur une échelle de 100 %. L'initialisation commence par choisir un âge en respectant les différentes probabilités de chacun d'eux. Lorsque l'âge sélectionné peut être associé à plusieurs catégories, le choix du sous-type est alors fait aléatoirement avec équiprobabilité.

La répartition démographique des âges est donc respectée à l'échelle d'un type mais on peut trouver d'importantes nuances en considérant un ensemble de type d'individus. Il est même possible de ne pas retrouver certains intervalles d'âges dans la digitalisation statique.

1.3.4. Considération extérieur

L'une des grandes complexités dans la digitalisation, qu'il soit statique ou dynamique, est la modélisation extérieure du territoire. Tous les territoires sont perméables, à moins de prendre le monde dans sa globalité ce qui est impensable. Il serait toutefois intéressant de discuter la corrélation entre la surface du territoire étudié et l'affluence extérieur statique, et dynamique, à digitaliser. Pour

une même surface, on trouvera une influence extérieure très supérieure à Paris que dans une campagne. Alors il y a-t-il plutôt une corrélation avec la population “interne” du territoire, qui pour rappel peut aussi être estimé à partir du fichier *openstreetmap* lui-même ?

Maintenant, il est vrai qu’on peut se questionner sur la pertinence de l’âge et des classes socio-dynamiques. Est-ce que les individus “externes” sont toujours les mêmes ? Dans leur ensemble, est-ce qu’on retrouve une répartition des âges logiques ou similaires au territoire étudié ? Y a-t-il même un intérêt à les enregistrer, ou plutôt les représenter par des individus spontanés aux attributs aléatoires ?

Il est évident de devoir approfondir cette approche avec plus de recul et de réflexion. D’un point de vue statique, on peut identifier les personnes externes par leur occupation avec le rôle de travailleur ; la distinction entre les internes et les externes étant basée sur la position de leur foyer par rapport au territoire. Et puis d’un vue dynamique, il faut également ajouter des individus un peu aléatoirement, avec une fréquence et une diversité à discuter.

Leur digitalisation est essentielle mais demande plus de réflexions pour qu’il soit pertinent d’en parler.

1.3.5. Socle individuel

Le socle individuel sont des assignations essentielles. Elles permettent d’apporter un forme de déterminisme et de cohérence dans la digitalisation dynamique. Dans une version initiale, on représente le socle par l’espace de logement et celui de travail de l’individu. Dans chacun des deux, il semble aussi judicieux de spécifier l’entourage local fréquent de l’individu. Dans l’absolu, cela permet d’avoir une très bonne modélisation de la digitalisation de chaque individu dans un cas de grande restriction dynamique.

Dans le cas de l’étude épidémique, il est plus qu’essentiel d’avoir une bonne modélisation en cas de grandes restrictions politiques. Cela permettrait par exemple de voir la réelle efficacité d’un confinement et d’observer d’éventuelles failles dans la connexion des réseaux de contact par exemple.

Bien entendu, ce socle peut évoluer en fonction du temps mais on suppose que l’individu conserve son logement, ses colocataires, son travail et ses collègues proches dans une première version. On pourra voir plus tard des pistes pour

ajouter davantage d'informations pour modéliser et anticiper la base dynamique de l'individu. Ce sera le cas par exemple avec la digitalisation dynamique où il sera important d'attribuer un mode de transport à chaque individu. Parmi les pistes les plus pertinentes, on pourra s'intéresser à considérer les relations amicales proches et familiales de l'individu. La question sur l'exploitation de ce socle en terme dynamique reste toutefois à approfondir. De même, on pourra éventuellement étendre le socle avec les habitudes des individus, qu'on pourra comparer à des besoins plus précisément ici, tel que la visite quotidienne de l'individu à une salle de sport pour citer un exemple. Les habitudes d'ordre ludique ne seraient vraiment pas pertinentes, ni même intéressantes, à considérer pour simplifier la conception.

A noter que cette conception ne traite pas l'âge des individus : les enfants sont considérés comme des personnes indépendantes avec un logement, un lieu de travail, etc. Dans ce cas, on veillera à initialiser une école comme lieu de travail. On pourra alors réfléchir à l'espace de travail des individus sans travail. Est-il vide et auquel cas il serait nécessaire de traiter ce cas à part, ou doit-il être représenté par leur espace de logement ?

Comment assigner les espaces de logement et de travail de l'individu ?

L'objectif est simple, faire une jointure entre la base de données des individus internes à celle des espaces de logements. Pour rappel, chaque catégorie d'individu à une liste de catégories de logements où il peut résider. Par exemple, on ne pourra pas retrouver les établissements de retraite dans une catégorie socio-dynamique d'enfants. Cela amène d'ailleurs d'autres réflexions pour la classification des catégories des individus pour associer spécifiquement les individus handicapés aux établissements de logements adéquats. On retrouve le même principe aux niveaux des espaces professionnels. Chaque catégorie a des ensembles définis. Il faut toutefois prendre en compte la complexité de la répartition professionnelle d'un territoire.

Dans une version simple, la sélection est faite aléatoirement parmi l'ensemble catégories qui lui sont associées, en n'excluant aucun espace spécifiquement. L'ajustement est fait en pondérant la sélection avec l'aire de chaque espace. Par conséquent, un espace avec une aire supérieur à une autre aura proportionnellement plus d'assignation que le second. On pourra bien entendu revoir cette sélection pour perfectionner la digitalisation statique aux données observées. De même, il pourrait être judicieux de s'intéresser à la possibilité d'exclure des espaces dans les sélections.

Cette conception n'est pas complète : il faut pouvoir identifier les personnes réellement proches de l'individu dans chacun de ces espaces clés. Par exemple, lorsque l'espace en question est un petit appartement, tous les individus du logement sont en effet en contact. En revanche, lorsque le logement est une grande résidence dans son ensemble, très peu de personnes sont réellement en contact avec l'individu. C'est pourquoi il est nécessaire de répertorier les personnes dites proches de l'individu dans chacun des espaces assignés. L'idéal serait de reprendre des données démographiques réelles tel que la répartition du nombre de personnes dans les foyers ou du territoire. En effet, cette attribution est faite de manière itérative, sans vouloir reproduire de logique dans la construction des ménages. Le programme ne cherche pas, par exemple, à rassembler deux enfants et deux adultes dans la quarantaine. A l'inverse, il est totalement possible d'observer des regroupements originaux tels que cinq enfants dans une maison. Ce problème de cohérence impacte moins les lieux de travail puisqu'il semble y avoir plus d'hétérogénéité. D'autant plus qu'une certaine logique est respectée avec l'association des catégories d'emploi à celle des individus. C'est-à-dire qu'on ne risque pas d'observer des enfants, dont le travail est l'école, à travailler aux côtés d'adultes ; excepté lorsqu'il s'agit d'enseignants. Bien évidemment, on pourra également s'intéresser aux données réelles du territoire sur la répartition du marché du travail pour plus de réalisme.

2. Digitalisation dynamique de la société

2.1. Scénarios des individus

D'un point de vue épidémique, plus la société est agitée, plus une épidémie se propage. Il est donc nécessaire de décrire les dynamiques individuelles de chaque individu. La question temporelle est essentielle à traiter. En considérant la diversité des déplacements sur une même journée et l'utilisation de *SumoMobility*, il est évident de traiter la digitalisation dynamique jour par jour, à l'échelle de la minute. En pratique, on appliquera une échelle de seconde pour s'adapter au format *SumoMobility*.

Comment représenter un scénario de 24 heures ?

L'idée est de présenter une suite d'espace avec des créneaux définis ainsi que des informations supplémentaires précisant le déplacement tel qu'un degré d'importance ou le rôle de l'individu. Tous les scénarios commencent à un temps fixe dans la journée. En règle générale, on observe plus d'individus actifs à 01h00, pendant la nuit, qu'à 05h00 du matin. Il semble donc judicieux de préciser une heure fixe de début des scénarios à 04h00. Les scénarios commencent donc par le réveil de l'individu et se finissent par son sommeil. Bien entendu, en théorie, un scénario débute et se termine à la même heure. La plus grande complexité de la digitalisation dynamique est la généralisation de la création des scénarios pour chaque individu, en respectant la diversité des situations. On pourra reprendre le parallèle avec un réseau de neurones.

2.1.1. Plusieurs hyperparamètres

Un scénario doit tenir compte de plusieurs facteurs pour respecter cette complexité spécifique :

- Etat infectieux de l'individu : états connus, états réels.
L'état infectieux impacte prioritairement le scénario d'un individu. Plus son état infectieux réel est grave et moins on s'intéresse à l'état qui lui est identifié. En revanche, plus son état réel est sain et plus on s'intéresse à son état connu. On pourra davantage parler de propension de l'individu pour un état peu grave et de

contraintes/obligations de l'individu pour un état grave. Pour plus de détails, les réflexes infectieux associés à chaque cas sont plus bas dans le rapport.

- Décision territoriale :

Les décisions territoriales imposent des contraintes à la population. Ce sont les solutions combinatoires sur lesquelles l'algorithme d'apprentissage par renforcement se repose. C'est la base même de cette recherche d'optimisation multicritères de la société. Ces décisions territoriales sont à définir manuellement en décrivant les actions qu'elles appliquent à la société. A noter qu'on pourra associer des décisions à des contraintes/obligations sur les tests et les vaccinations. Pour plus de détails, les décisions sont plus commentées dans une section ultérieure.

- Type de jour : jour de semaine, samedi, dimanche, jours fériés, période saisonnière.

Le quotidien d'une personne dépend beaucoup du calendrier. Selon le type du jour, la société peut être modélisée de façon assez différentes. En semaine, la plupart des individus travaillent pendant la journée, ont peu de temps libre en fin d'après-midi puis sortent ou restent chez eux le soir. Sinon, le week-end, ils disposent de la plupart de leur temps pour faire ce qu'ils veulent. Le type de jour est donc un paramètre intéressant parce qu'il permet facilement d'avoir des informations sur l'état de la dynamique en général. De plus, il est évident que la période de l'année est un paramètre très important pour caractériser au mieux les propensions dynamiques. Les individus s'orientent plus dans les parcs en été qu'en hiver. Dans une version ultérieure, on pourra s'amuser avec ce calendrier pour avoir plus d'outils et de levier pour concevoir des scénarios réalistes.

- Type de quotidien : travail, repos, vacances

Chaque individu peut avoir des quotidiens différents indépendamment du type de jour. Il peut soit avoir un quotidien où il passe un certain nombre d'heures, à définir, au travail, soit être au repos complet, ou même partir en vacances et sortir complètement de sa routine.

- Rôle des espaces : travail économique, nécessaire, non nécessaire

Le rôle des espaces est important à préciser. D'un point de vue de l'évaluation économique, il est essentiel de distinguer les individus qui travaillent ou qui visitent les différents espaces. Après, il faut également distinguer les déplacements nécessaires de ceux non nécessaires. Cela a peu d'incidence dans une situation normale, mais ces spécifications sont plus que nécessaire en cas de restriction. Par exemple, les supermarchés, les lieux médicaux et les pharmacies sont des déplacements nécessaires.

Dans une version ultérieure, il pourrait être intéressant de ne plus représenter l'importance des espaces en terme binaire mais avec une échelle de nécessité. On pourrait par exemple considérer que les salles de sport sont un "défouloir" psychologique pour les individus. On pourrait alors considérer ces espaces comme des lieux semi-nécessaires. On en revient à une perspective d'évolution où on considère les habitudes des individus, lorsque celles-ci semblent être un besoin individuel. A l'inverse, un centre d'attraction sera toujours dénoté comme non nécessaire par exemple.

Excepté le rôle de travail des espaces, on pourra ainsi classer ce niveau d'importance dans les catégories d'espace directement. Cela pourrait ajouter d'ailleurs une nouvelle dimension psychologique du projet, qui sera exposé à la fin du rapport. A noter que cela n'a rien à voir avec l'aptitude des espaces à contaminer ses occupants.

- Diverses propensions : déplacements extérieurs, autre logement, visite familiale ou amicale, etc.

Chaque individu a une propension à se déplacer à l'extérieur de la zone étudiée. Les individus résidant à proximité d'une limite frontalière de la zone étudiée seront plus enclins à se déplacer à l'extérieur. On retrouve la problématique quant à identifier la proximité des espaces par rapport aux limites géographiques de la région.

Ensuite, l'individu peut occasionnellement dormir ailleurs que chez lui, que ce soit dans la zone étudiée où non. On peut également considérer la propension de l'individu à partager des moments avec ses amis et sa famille. On retrouve la perspective d'évolution du projet où l'on spécifie les amis et la famille de chaque individu.

C'est en quelque sorte l'hyperparamètre pour toutes les possibles évolutions de la digitalisation dynamique, qui sera utile en autres pour approcher les affluences observées.

- Type de sélection des espaces : assigné à l'individu, aléatoire, critères

Lors de l'élaboration d'un scénario, il faut savoir quel espace ajouter à la séquence. Alors la question ne se pose pas quant au socle de l'individu où le logement et le travail sont bien identifiés. On rejoint la représentation des habitudes des individus, que l'on peut associer avec certaines propensions. Quant au reste, le quotidien d'un individu peut être très diversifié, et il faut pouvoir choisir des espaces parmi l'ensemble de la base de données.

D'un point de vue simpliste, on peut dire que l'individu visite une suite d'espace un peu aléatoirement. Dans un certain ensemble de catégorie, un espace peut être choisi de façon totalement aléatoire ou alors selon certains critères. Plus la sélection des espaces sera aléatoire et plus la circulation de l'épidémie aura un mouvement brownien important. Il peut donc être judicieux de perfectionner cette sélection par des critères tels que l'espace le plus proche ou le plus volumineux dans un certain rayon de distance par exemple. Il convient donc d'ajouter des fonctions de filtration pour sélectionner un espace parmi la base de données.

Cet hyperparamètre pourrait être perfectionné avec des données volontaires de la population sur leur préférence, comme des sondages.

2.1.2. Comment s'y prendre ?

L'optimisation des agendas pour une population est déjà un problème complexe. L'idée ici est de simplifier cette généralisation de scénario avec les paramètres cités ci-dessus. Dans l'ordre d'importance, les scénarios dépendent d'abord de l'état infectieux de l'individu puis des décisions territoriales, particulièrement lors de grosses contraintes/obligations. Ensuite, on peut identifier des structures de déplacements selon le type de jour et le type de quotidien de l'individu. On a ensuite les diverses propensions de l'individu à organiser certains déplacements nécessaires ou non. Pour le reste, il s'agit d'identifier les espaces à introduire dans la séquence, qu'il soit déjà prédéfini, choisi selon certains critères, catégories ou entièrement aléatoire, etc.

Chaque individu possède certaines nécessités qu'il doit réaliser selon un certain rythme au cours du temps. C'est le cas par exemple des courses alimentaires ou de visites médicales de routine. Il n'est pas réaliste de reproduire ce genre de déplacements tous les jours. Il convient donc d'établir un système de propension de l'individu à construire son quotidien. Une première approche pourrait être de

reproduire plus ou moins le même système qu'une distribution partagée de tâches informatiques sur un même cœur de calcul, où les tâches représenteraient l'ensemble des déplacements disponibles. Chaque déplacement disponible à un score qui s'incrémente au fil de chaque jour. Il sera nécessaire de les séparer par le caractère nécessaire ou non. En considérant donc le score de chacun d'eux, l'individu sélectionnerait celui le plus adéquat dans son quotidien.

Le levier de sélection pourrait être un système de seuil à partir duquel l'individu placerait le déplacement dans son scénario. Il serait alors nécessaire de respecter la place disponible et la concurrence des déplacements dépassant ce seuil. Il semblerait judicieux de choisir un seuil identique pour tous les espaces et de modifier l'augmentation journalière du score de propension de chaque catégorie. Un autre système de levier pourrait être d'intégrer ces déplacements en s'amusant avec des processus stochastiques. Ces décisions stochastiques seraient effectuées chaque jour, indépendamment du temps. On appliquerait alors une incrémentation identique pour toutes les catégories d'espaces. La différenciation serait ajoutée par des facteurs spécifiques à chacune de ces catégories. Quel que soit le levier de sélection, les scores de chaque catégorie seraient remis à zéro à chacune de leur sélection.

2.1.3. Problématique du temps

Le temps est la contrainte majeure lors de la conception d'un scénario. Un scénario ne doit pas dépasser 24h. Il y a deux options pour prendre en compte la limite de temps. Un peu à la manière du problème du sac à dos, un scénario a une limite de temps et il faut le remplir avec des déplacements qui ont une certaine importance/obligations, tout en ne dépassant pas cette limite. Alors soit le scénario peut être construit de manière itérative. Soit on sélectionne d'abord l'ensemble des déplacements de la journée puis on les agence avec des timings. En théorie, il faudrait reproduire la conception d'un scénario dans notre vie au quotidien : nous devons tous respecter des créneaux horaires et nous adapter au temps qu'il nous reste.

Il y a une certaine complexité à prendre en compte : on ne peut pas prédire le temps de trajet de l'individu, d'où l'intérêt d'utiliser *SumoMobility* pour avoir l'aspect temporel des déplacements. Nous verrons plus tard que le projet cadence les scénarios par les départs des individus en direction d'un espace, ce qui comprend donc le temps de trajet non prévisible ainsi que le temps sur place dans l'espace. Il est nécessaire d'avoir des estimations pour chacun de ses déplacements, en considérant le type de transport et le durée souhaité sur place.

Ce qui nous intéresse dans ce projet est le temps effectif de chaque individu dans les espaces.

Il y a une problématique non négligeable à mal cadencer les timings des déplacements. Si une durée entre deux timings est trop courte, l'individu pourrait ne pas avoir le temps de terminer son trajet en direction de l'espace qu'il devra déjà commencer le prochain déplacement en direction de l'espace suivant. En pratique, il n'y aurait pas de problème lors de l'exécution avec *SumoMobility*. Mais une fois la simulation effectuée, la phase d'analyse ne pourra pas évaluer correctement les scénarios sachant que certains espaces ne seront pas réellement visités.

Il pourrait être intéressant de spécifier une référence sur les temps d'occupation des individus dans les espaces. Chaque catégorie d'espaces pourrait donc être caractérisée par une durée type de passages, que ce soit pour une simple visite, un temps de travail, etc. Cette durée pourrait être représentée par une quantité fixe ou un loi normale positive, paramétré par une certaine moyenne et un certain écart-type. A noter qu'on peut retrouver une durée aléatoire en pratique puisqu'on ne pas anticiper le temps exact de trajet. Une estimation intéressante de ce temps de trajet pourrait être réalisée en fonction du type de transport et de la distance entre le point de départ et le point d'arrivée. On pourra par exemple analyser une régression linéaire entre la distance, le type de transport et le temps effectif de transport.

Les timings doivent également suivre une certaine cohérence pour regrouper les individus dans leur dynamique. Par exemple, les individus ont des créneaux assez généraux pour les périodes de repas. Les timings de travail dépendent du type de travail de l'individu, de la durée hebdomadaire, etc.

2.1.4. Gestion mémoire

La question de la gestion mémoire des scénarios de chaque individu est un point intéressant. La problématique de base est de savoir si un scénario est unique au cours du temps ? D'un point de vue réaliste, non. Mais dans un souci de facilité, on pourra préférer une gestion mémoire des scénarios.

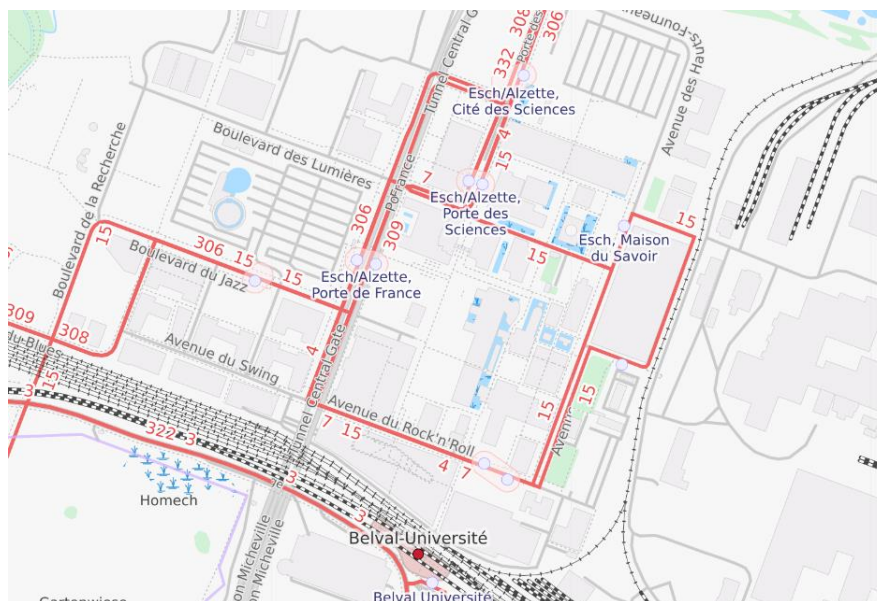
Pour simplifier, on peut initialiser plusieurs scénarios pour chaque individu selon certaines conditions, quitte à les éditer en fonction par exemple de leur état infectieux ou des décisions politiques. Cette conception serait donc intéressante pour prendre en compte l'individualité de chaque individu et du calendrier. Pour chaque jour de simulation, cela revient à enlever le temps de construction de chaque scénario pour le remplacer par un temps d'adaptation aux conditions. Or, nous avons précédemment noté que les paramètres les plus impactant lors de la construction des scénarios sont les décisions politiques et les réflexes infectieux. D'un point de vue de l'exécution, cela revient à augmenter le besoin mémoire du programme mais d'accélérer le temps de la digitalisation. On perd toutefois une certaine diversité ainsi qu'un côté imprévisible de la mobilité.

L'autre conception serait de ne pas garder en mémoire les scénarios mais de les construire à chaque jour de digitalisation. Par rapport à la première option, il serait alors nécessaire de conserver les outils de constructions des scénarios pour chaque individu tels que les scores de propension, etc.

2.2. SumoMobility

SumoMobility est une application simulant la circulation urbaine à partir de données cartographiques d'un réseau routier. En reprenant le fichier *openstreetmap* et en appliquant un pré-traitement, on peut faire circuler les individus à travers les différents lieux réels. Il est possible d'utiliser les transports en commun de la région décrite par le fichier *openstreetmap* avec les réelles positions des arrêts. Pour rappel, chaque individu est initialisé avec un moyen de transport, selon son âge et certaines proportions. Pour l'efficacité du projet, on se limitera à une utilisation simpliste des déplacements. L'intérêt est de considérer uniquement les temps de trajet des individus dans la circulation au quotidien, et d'avoir un aperçu des risques infectieux dans les transports en commun.

En effet, les transports en commun sont également des espaces à prendre en compte dans la propagation d'agents infectieux, bien qu'ils ne soient pas répertoriés dans la base de données des espaces. Cette considération peut se faire de façon globale par simplicité : on peut regarder le rapport entre la proportion des personnes infectieuses et non infectieuses utilisant les transports publics ainsi que la fréquence de ces transports publics. Un score d'évaluation globale peut ainsi être fourni à la fonction d'évaluation infectieuse de l'individu ; ce score sera pris en compte uniquement si l'individu prend les transports en commun. Bien évidemment, on pourra davantage perfectionner cette considération pour la rendre plus individualiste.



Un aperçu des données sur les transports publics du format *openstreetmap*.

Pour simplifier l'utilisation de *SumoMobility*, la simulation d'une journée dans la société se réalise sans faire interagir les individus entre eux. Cela évite les calculs inutiles dans notre utilisation, et évite aussi des bugs non prévisibles. Par exemple, l'un de ces irrégularités peut être la collision/l'embouteillage de plusieurs individus sur une voie piétonne, ce qui provoque de sérieux problèmes sur la digitalisation. De telles considérations dans le domaine du génie civil sont inutiles et même très problématiques pour le projet. De même, dans un cas simpliste, on ne s'intéresse pas aux options avancées de *SumoMobility* tel que le recalcul des itinéraires en fonction de la circulation, la gestion des feux de croisement ou l'utilisation de taxis. Dans notre utilisation simple, quelques paramètres sont toutefois à régler :

- Nous devons choisir manuellement la fréquence de passage des véhicules dans la génération automatique des transports publics avec les arrêts et les lignes décrites par le fichier *openstreetmap*. C'est un paramètre important pour les individus qui se véhiculent par les transports publics, que ce soit pour la rapidité ou le risque de contamination de leurs déplacements. Toutefois, on peut assez aisément trouver des informations sur la fréquence moyenne des transports publics d'une région, ou simplement renseigner une fréquence qui semble réaliste. Il serait intéressant de dissocier les différentes classes de transports publics dans une version ultérieure pour pouvoir leur appliquer des fréquences de passage différentes.
- Il est intéressant de s'intéresser à la contrainte de stationnement des véhicules. Dans une version simpliste, on suppose que les individus se déplaçant en voiture n'ont pas besoin de laisser leur véhicule dans un stationnement ; les voitures apparaissent et disparaissent directement au niveau du passage piéton. Ce paramètre n'est pas si important que cela dans la digitalisation d'une épidémie. Bien entendu, dans le cas pratique, cela pourrait représenter le temps que les citoyens dépensent à trouver un parking en ville, sans oublier les contraintes horaires ou forfaitaires des parkings. Cette option serait à l'inverse très intéressante dans une optique de perfectionnement de la digitalisation dynamique ou du génie civil.

La simulation est continue dans le temps et les déplacements sous *SumoMobility* se font uniquement en considérant le réseau routier. Par conséquent, pour modéliser l'arrêt d'un individu dans un espace, il faut acheminer l'individu au niveau de la route qui est la plus proche de cet espace. Par simplicité, les itinéraires entre les différents espaces sont réalisés avec le même type de transport, unique et assigné à chaque individu.

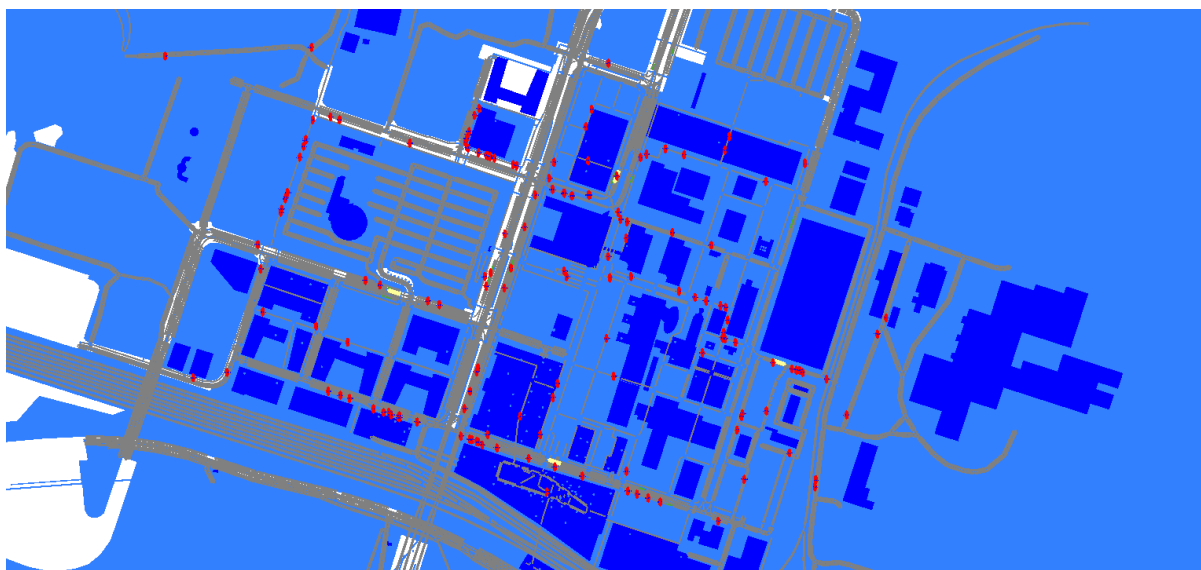
Pour faire le lien entre les scénarios des individus et leur implémentation sous forme d'itinéraires dans *SumoMobility*, il est nécessaire de procéder à quelques

modifications. A partir des scénarios des individus, il faut être capable d'écrire les *trips* des individus qui seront ensuite traduits en *routes* par un algorithme spécifique de *SumoMobility*, nommée *duarouter*. Ce routeur est très technique et possède de nombreuses options pour identifier l'enchaînement le plus rapide des routes pour relier une route de départ et une route d'arrivée. Il est possible d'ajuster beaucoup de paramètres liés à l'optimisation tels que les véhicules pouvant circuler sur chaque route, les vitesses limites sur chaque route, les vitesses détaillées de chaque véhicule, etc. L'algorithme qu'il utilise est Dijkstra. Par simplicité, on ne se préoccupe pas de cette optimisation puisque ce n'est pas le sujet de ce projet. Cet outil nous permet simplement de trouver des déplacements judicieux entre des points d'intérêt. On se soucie toutefois de la connectivité du réseau routier, que l'on peut tenter de corriger directement avec le routeur, pour ne pas avoir d'erreur dans la recherche de chemin. Dans cette acquisition, l'algorithme peut utiliser les transports publics, pour les individus utilisant le mode de transport public, qui auront déjà été introduits et chronométrés précédemment. Un atout indéniable de *duarouter* est la possibilité d'effectuer ses calculs sur plusieurs threads et d'accélérer ce processus. En particulier, cela est non négligeable lorsque les scénarios des individus évoluent au fil des jours et que l'étape de routage doit être faite régulièrement.

Le routeur crée automatiquement des véhicules avec un identifiant unique par individu. Chaque véhicule effectue alors le déplacement correspondant à un itinéraire. L'itinéraire total de l'individu désigne donc les identifiants uniques de ces véhicules pour réaliser les itinéraires associés à ces véhicules. Toutefois, cette conception présente un problème lorsque l'individu possède une voiture et passe à travers de nombreux espaces : les stops de l'individu, représentant l'arrêt dans chaque espace au cours du scénario, ne peuvent pas être introduits entre un même enchaînement de routes. L'individu ne peut donc pas utiliser sa voiture pour réaliser deux itinéraires différents. L'itinéraire de sa voiture correspond au premier déplacement de l'individu, c'est-à-dire le premier enchaînement de route, sans stops, de l'individu.

Une des pistes de correction serait de modéliser l'itinéraire de la voiture unique de l'individu, en comprenant les arrêts de l'individu. Toutefois, cela ne semble pas être réalisable depuis *duarouter*. C'est à partir de cette problématique que l'on doit séparer l'itinéraire journalier total de l'individu en suite d'itinéraire simple. Or, puisqu'on ne peut pas prédire le temps de trajet du premier itinéraire, on doit estimer les timings sur l'ensemble des itinéraires suivants. Dans un cas idéal, il aurait été certainement plus intéressant de considérer uniquement le temps d'occupation des individus dans les espaces, sans se préoccuper des estimations des temps de transport.

Avec cette conception, le scénario de chaque individu est divisé en plusieurs sous-trajets, identifiés par l'identifiant de la personne auquel s'ajoutent des caractères spécifiques qui s'incrémentent avec le nombre d'itinéraires. Un même individu, sous différents identifiant de routes, peut ainsi utiliser plusieurs voitures pour des itinéraires différents. Un itinéraire devient donc le passage simple d'un espace à un autre. Pour rappel, on ne tient pas compte de la durée que l'individu occupe un espace mais des *timings* où l'individu souhaite se diriger dans un espace. La durée d'occupation de l'individu dans un espace est alors réanalysé après l'exécution de *SumoMobility*. La décomposition d'un scénario d'un individu en plusieurs itinéraires se fait de manière contigüe dans le fichier *routes*, ce qui enlève l'ordre chronologique de tous les itinéraires du fichier. Il convient donc d'appliquer un algorithme propre à *SumoMobility* pour reclasser dans l'ordre tous les itinéraires du fichier en fonction uniquement du temps de départ de chaque trajet.



Un exemple d'exécution graphique de *SumoMobility* ici dans le réseau du campus Belval à Esch-sur-Alzette.

Dans une première version de l'affichage graphique, il est possible d'observer la circulation des individus dans un environnement décrit comme dans l'exemple ci-dessus. Quel que soit leur état infectieux, les individus sont décrits par des points rouges. Les formes en jaune correspondent aux véhicules tels que les voitures ou les bus. Les espaces que les individus parcourent au quotidien sont représentés en bleu foncé. Bien entendu, il pourrait être intéressant de revoir cet affichage pour montrer plus de détails : les individus seraient représentés en différentes couleurs selon leur état infectieux et de même pour les espaces en fonction de leur catégorie. Il semble en revanche impossible à priori de faire évoluer l'affichage en fonction du temps pour essayer de représenter la propagation infectieuse pendant l'exécution de *SumoMobility*. A noter qu'on ne peut pas

observer les déplacements des individus entre le réseau routier et l'intérieur des espaces.

Il pourrait être possible d'afficher la carte en ajoutant des éléments graphiques pour représenter les données statistiques finales d'une journée par exemple. Sinon, au terme d'un certain nombre de jours, on pourrait représenter graphiquement les espaces où il y a le plus de contamination. Il existe plein de possibilités illustratives pour apporter des informations intéressantes.

La digitalisation dynamique par *SumoMobility* peut également être parallélisé lors son exécution, ce qui est encore une fois non négligeable pour l'efficacité de ce projet. L'affichage graphique présente toutefois des limites : l'exécution est nettement plus lente et peut même saturer en cours de simulation en cas de région trop volumineuse. L'exécution présente toujours des irrégularités non prévisibles mais qui n'ont aucun impact dans la digitalisation. Parmi elles, on pourra citer l'apparition de collision entre véhicules, qui peut se résoudre automatiquement avec des paramètres de *SumoMobility*.

2.3. Réseau routier

Il est nécessaire de concevoir un intermédiaire entre les identifiants des espaces et leurs implémentations en trajet sous *SumoMobility*. En effet, sous *SumoMobility*, les déplacements utilisent des identifiants de route pour concevoir ses itinéraires, et ne reconnaissent pas les espaces.

A l'initialisation, l'application importe le réseau routier de l'environnement avec le fichier *openstreetmap*, tout en conservant au mieux les identifiants du format d'import. Toutefois, que ce soit par le format *openstreetmap* ou son importation dans *SumoMobility*, aucun lien direct n'est fait entre les espaces et les routes. Il est donc nécessaire d'attribuer manuellement une route à chaque espace : on trouve d'un côté les routes sous *SumoMobility* et de l'autre les espaces dans la base de données interne.

Lors de l'importation du réseau, les identifiants *openstreetmap* des routes s'adaptent au réseau routier en fonction de l'agencement des routes entre elles. Si une route à des croisements le long de sa description, aucun avertissement n'est affiché dans le fichier *openstreetmap*. De l'autre côté, *SumoMobility* décompose la route en différentes routes représentés par le même identifiant original auquel s'ajoutent des caractères uniques. On retrouvera toutefois toujours une même logique dans ces caractères : il est donc accessible de relier une route avec son identifiant *openstreetmap* par l'une de ses sous-routes avec les identifiants *SumoMobility*.

Pour chaque espace, on s'intéressera donc à appliquer une recherche locale pour identifier une route *openstreetmap*. Pour cela, on effectue une boucle de recherche qui étend petit à petit les périmètres de filtre géographique autour de la position de l'espace. L'une des approximations concerne la localisation des routes. Même si on retrouve le principe de nœuds formant un polygone, une route sous *openstreetmap* peut être très grande. Sa localisation moyenne dans un optique de recherche local peut être très incorrecte si on ne s'attache pas dans le détail par ses nœuds. D'autant plus qu'elle est rattachée à une sous-route dans *SumoMobility* et que cette sous-route n'est pas toujours la portion de la route la plus proche par rapport à l'espace. On notera toutefois que ces imprécisions restent insignifiantes. Excepté pour des cas particuliers, on restera dans une recherche simple. Pour une contrainte de temps, on utilise la méthode rapide de calcul de distance.

2.4. Mouvements interterritoriaux

Pour digitaliser au plus simple les déplacements interterritoriaux, on ne se préoccupe pas des trajets incidents ou efférents avec l'extérieur. En effet, les temps de trajet sont intéressants uniquement pour observer au mieux l'occupation dans les espaces et évaluer au mieux la contamination des individus.

La gestion épidémique à l'extérieur du réseau n'est pas "contrôlable", ni prévisible et la zone que l'on étudie subit en quelque sorte les états sanitaires extérieurs. C'est pourquoi on se limite à une évaluation sanitaire plus ou moins aléatoire des individus à l'extérieur de la région. On peut toutefois généraliser grossièrement le scénario des personnes à l'extérieur et faire varier l'incidence infectieuse aléatoirement au cas par cas. Par exemple, on peut supposer qu'un individu sort du territoire étudié, passe six heures dans un bureau pour y travailler, enchaîne par deux heures dans des clubs nocturnes, pour enfin rentrer dans son logement à l'intérieur du territoire. Son exposition infectieuse à l'extérieur sera plus critique lorsqu'il est dans un club nocturne que dans un bureau.

Il est évident que la pertinence de l'évaluation infectieuse est hautement dépendante de la durée avant qu'il revienne dans le territoire. Il serait aberrant d'étudier l'évaluation infectieuse d'un individu externe qui ne passe que le week-end dans le territoire par exemple. Dans ce cas précis, il serait alors raisonnable d'estimer l'état infectieux de l'individu un peu aléatoirement avant le week-end, voir s'il peut s'accommoder avec les décisions actuelles puis l'intégrer aux évaluations infectieuses pendant le week-end.

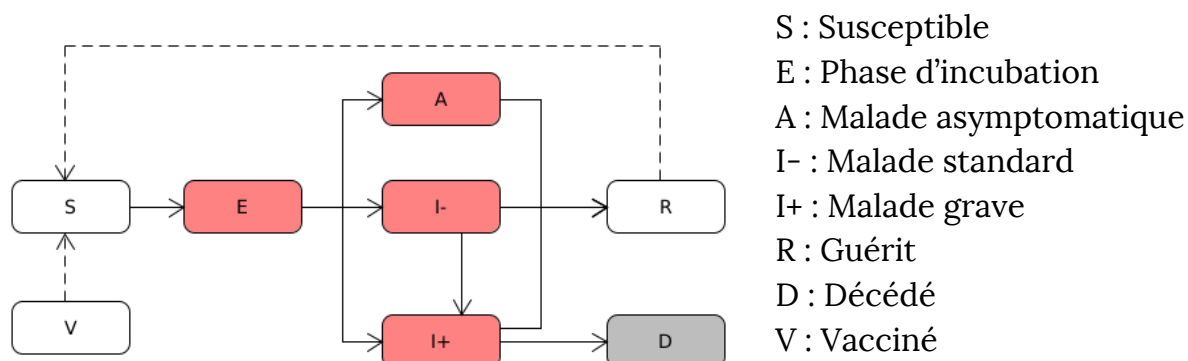
Un autre cas intéressant serait d'étudier les mouvements interterritoriaux, notamment dans le cas des malades externes graves. En fin de compte, il serait nécessaire d'approfondir les réflexions sur les modalités extérieures. Notamment en ce qui concerne les propensions des individus internes ou externes à faire des déplacements interterritoriaux.

Les individus externes au territoire peuvent par exemple débiter leur arrivée par un hôtel dans le territoire. Sinon, quel que soit l'emplacement des logements des individus, lorsqu'un individu provient de l'extérieur, il peut commencer son itinéraire par n'importe quel espace directement ; ou en particulier dans des espaces associés aux longs transports tels qu'une gare ou un aéroport. De manière générale, il n'y a pas nécessairement de contraintes que ce soit à l'arrivée ou au départ. En revanche, on pourra orienter la digitalisation pour reproduire des données observées.

Dans une version ultérieure, on pourra aussi s'intéresser à l'existence de plusieurs territoires différents, ayant chacun des caractéristiques différentes.

3. Dynamique des agents infectieux

3.1. Schéma infectieux



En suivant les réflexions des épidémiologistes, on peut décomposer les états infectieux des individus comme décrit ci-dessus sur le schéma. Les flèches en continue représente une évolution certaine qui dépend de plusieurs paramètres temporels et/ou stochastiques. Celles en pointillés sont plus anecdotiques. Les phases en rouge correspondent au phase où l'individu est *porteur* et peut transmettre l'agent infectieux. Le point de départ de la population se situe à l'état *susceptible*.

A l'initialisation de l'agent infectieux dans l'environnement, selon un ratio et/ou une certaine quantité initiale, des individus sont placés à l'état *phase d'incubation*. Cette sélection est réalisée de façon aléatoire sur l'ensemble des individus de l'environnement. Il pourrait être intéressant de déterminer plus précisément les premiers individus *porteurs* et d'étudier la différence dans la propagation des agents infectieux. C'est un des points intéressants de ce projet.

Chaque individu est dans un état spécifique pour chacun des agents infectieux dans l'environnement. La concurrence entre ces agents peut avoir une incidence sur leur propagation. Par simplicité, lorsqu'un individu est *porteur* d'une infection, il devient "immunisé" pour toutes les autres infections de l'environnement : il est alors dans l'un des trois états parmi *susceptible*, *guérit* ou *vacciné* selon sa situation et son vécu.

3.2. Transitions des états

3.2.1. Susceptible

Chaque infection est caractérisée par un seuil de contamination qui correspond au nombre de particules à partir duquel un individu entre en *phase d'incubation*. Il semble judicieux de l'adapter selon l'âge de l'individu pour représenter la résistance de l'organisme. Dans une version simpliste, on présume que le nombre de particules émis par un individu *porteur* est identique quel que soit l'infection, l'état infectieux de l'individu *porteur* ou la durée depuis laquelle l'individu est *porteur*. On pourra revoir cette conception dans une version ultérieure pour ajouter un facteur dans l'émission de particule en fonction du score infectieux de l'individu *porteur*. Deux autres paramètres caractérisent la contamination d'une infection :

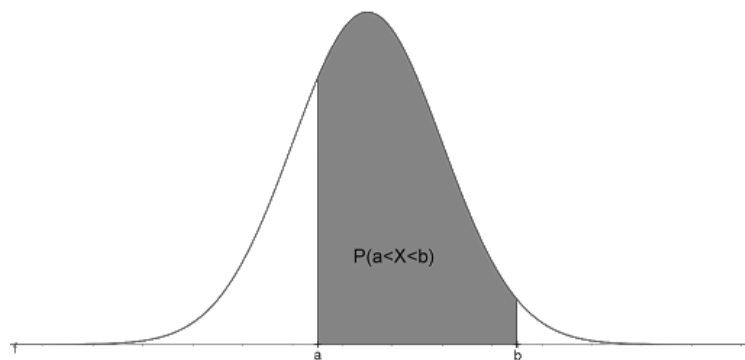
- Le temps qu'une particule peut persister dans un espace ; on fixera cette durée même si on pourra se questionner sur une approche gaussienne. On préférera jouer sur l'efficacité des particules réellement persistantes dans l'espace avec le facteur concernant la surface de l'espace dans l'évaluation infectieuse.
- Le nombre de jours qu'une particule peut être cumulé dans un individu ; ce paramètre peut en cacher un autre. On pourrait reprendre le principe d'une réaction immédiate de l'organisme : l'individu accumulerait constamment les particules extérieures, sans définir une durée limitée, et produirait une réaction immunitaire. Il conviendra alors de définir une capacité immunitaire insuffisante pour empêcher le total de particule de dépasser le seuil de contamination. A noter que cette nouvelle approche est inspirée des résultats obtenus lors de ce stage : les individus semblent beaucoup trop rapidement s'infecter sans laisser une certaine élasticité.

3.2.2. Phase d'incubation

Lorsque l'individu est en *phase d'incubation*, lors duquel aucun symptôme n'est visible, deux paramètres caractérisent l'infection : la moyenne ainsi que l'écart-type de la durée de l'incubation, que l'on suppose suivre une loi normale, indépendamment des caractéristiques de l'individu. A la fin de cette durée aléatoire, le statut infectieux d'individu change pour se diriger vers l'un des trois

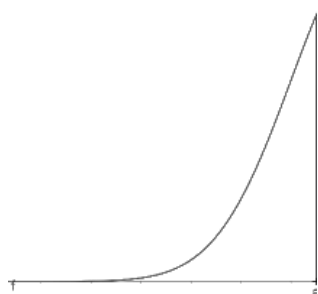
états : *asymptomatique*, *standard* ou *grave*. Cette transition dépend de l'âge de l'individu et de l'infection.

Lorsque la durée d'incubation est écoulée, chaque individu se voit octroyer un nouveau score d'infection, tel un montant d'antigènes infectieux dans l'organisme. Ce score aléatoire suit une loi normale selon une moyenne et un écart-type caractéristique de l'infection. L'idée est de définir l'état infectieux de l'individu en fonction de ce montant aléatoire. En effet, il est possible de sélectionner deux seuils autour desquelles on peut retrouver trois données connues dans une loi gaussienne. Cette répartition est également à nuancer en fonction de l'âge des individus. Dans un premier temps, la représentation infectieuse des individus est donc décrite par un score unique, assigné à l'aide d'une loi normale avec des paramètres fixes et indépendamment de l'âge des individus. La variation de la gravité infectieuse est donc modélisée par des bornes gaussiennes (a,b) variables selon l'âge des individus. Toutefois, il semble plus que judicieux d'étudier le degré de liberté inverse pour plus de simplicité dans une version ultérieure.



Une fois le score infectieux initialisé, l'évolution de l'infection se fait en augmentant ou en réduisant ce score, selon des paramètres propres au statut infectieux. De manière naturelle, l'organisme tend déjà à évoluer de lui-même. Si le score infectieux devient nul, l'individu est *guérit*, sinon, s'il dépasse un certain seuil de mortalité, l'individu est *décédé*. Pour faire le parallèle avec une vision médicale très simpliste, ce score peut représenter la différence entre le nombre d'antigènes et d'anticorps de l'individu.

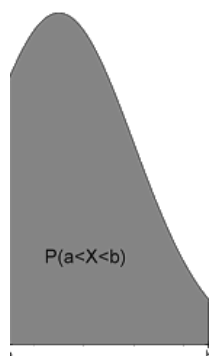
3.2.3. Malade asymptomatique



Chaque individu peut guérir de lui-même selon un certain score d'auto-guérissons journalier, un peu comme une production d'anticorps. Pour une première version, on suppose que ce score est un montant fixe et indépendant de l'âge de l'individu. Au vue de la courbe du score infectieux, cela signifie que la durée de guérison ne suit pas une loi normale mais plutôt une distribution de rémission que l'on pourra qualifier de tardive. Le score de guérison pourra donc être revu dans une version ultérieure en fonction des données observées et des premiers résultats de la conception. On suppose par simplicité que les individus *asymptomatiques* ne sont pas sensibles à des soins ou à des aggravations infectieuses.

Lorsqu'il n'est pas conscient de son infection *asymptomatique*, il n'y a pas de changement de comportement de la part de l'individu ; cet état ne provoque pas de réflexes infectieux. Bien évidemment, si son infection est identifiée, l'individu sera amené à changer son quotidien pour limiter la propagation épidémique selon les décisions territoriales. Dans une version simple, on ne s'intéresse pas à la propension de l'individu *malade asymptomatique* à adopter par lui-même un comportement protecteur vis-à-vis de la population en modifiant sa mobilité. Ce qui signifie que seules les décisions politiques influencent cet état.

3.2.4. Malade standard



L'évolution infectieuse *standard* d'un individu reprend le même principe d'auto-guérissons que celui *asymptomatique*, excepté qu'il est possible/recommandé à l'individu de suivre des soins. Que ce soit par des visites médicales ou une médication, le score infectieux de l'individu peut être réduit davantage. Il est possible à l'inverse de subir des événements augmentant le score infectieux de l'individu pour éventuellement le faire passer au stade *grave*. On pourra citer l'exemple de mauvais soins perçus involontairement par l'individu comme des maladies nosocomiales ou une mauvaise automédication. Chaque type de soin pourrait être soit positif ou négatif selon une probabilité. Ces

types de soins seraient également des paramètres à implémenter dans chaque infection.

Un individu en état d'infection *standard* présente des symptômes de l'infection et son quotidien sera impacté par des réflexes infectieux. Ces réflexes infectieux peuvent pousser l'individu à se tester ou à introduire des soins dans son quotidien, que ce soit des visites médicales ou de la médication. De par son organisme affaibli, il tendra également à limiter ses déplacements non nécessaires. Il est toutefois intéressant de préciser que les symptômes ne traduisent pas nécessairement l'identification de l'infection : l'individu devra suivre la politique concernant les individus *standard* que lorsque celui-ci sera évalué par un médecin ou un test infectieux. Selon les décisions territoriales, le quotidien de ces individus est également modifié avec des contraintes ou des obligations. Ces décisions pourront notamment orienter l'individu à identifier son état.

3.2.5. Malade grave



Les individus en état infectieux *grave* reprennent le même principe de soins que les *malades standard*. Les soins seraient alors différents. De plus, aucune auto-guérison n'est effectuée par l'individu et le score infectieux augmente de façon journalière. En pratique, on considère que sa production d'anticorps n'est pas suffisante face à celle des antigènes. On restreint donc la conception à une simple production d'antigènes, même si on s'écarte un peu d'un modèle médical réaliste. La différence entre cette aggravation infectieuse et les soins disponibles caractérise en quelque sorte la mortalité de chaque infection. L'individu est poussé pour sa survie à se soigner rapidement : au-delà d'un nouveau seuil, l'individu est *décédé*. Un individu en état *malade grave* n'a pas besoin d'être identifié pour être connu : il n'y a pas de distinction entre l'état réel et connu.

Les réflexes infectieux de ces individus deviennent la priorité au quotidien de ces personnes. Leurs instincts de survie n'est pas soumis à la politique territoriale. D'ailleurs, cette politique doit prendre en compte les ressources de soins limitées du territoire pour éviter d'abandonner ses individus en état *grave*.

3.2.6. Guérit & Vacciné

Pour une première conception, on ne s'intéressera pas aux réinfections après guérison ou après vaccination des individus. Un individu en état *vacciné* n'a pas besoin d'être identifié pour être connu : il n'y a pas de distinction entre l'état réel et connu. La question se pose toutefois pour l'état *guérit*. Dans une version simpliste, on considère que lorsque l'état réel *porteur* est terminé, l'individu passe en état *guérit* réel et connu. Il s'agit simplement d'une question d'ordre pragmatique. Dans la digitalisation, laisser un temps de préservation, qu'on pourra comparer à un temps de quarantaine, n'a que très peu d'impact sur l'étude de la propagation épidémique. La question pourra davantage se poser sur le critère économique. Et il n'est pas intéressant de raisonner sur une décision politique dans ce sens dans une conception initiale.

3.2.7. Décédé

Un individu en état *décédé* n'a pas besoin d'être identifié pour être connu... Dans cette conception infectieuse, les décès des individus ne suivent pas un taux de mortalité prédéfini par une donnée observée dans la réalité. La mortalité des individus dépend de la proportion d'individus *malades graves*, de l'écart entre l'aggravation infectieuse et les soins disponibles de l'infection, et également de la réussite des individus à "survivre" avec ces soins disponibles. Il n'est pas aisé de faire correspondre directement un taux de mortalité observé dans la réalité avec la digitalisation. C'est probablement l'un des points les plus délicats de la conception infectieuse : la mortalité d'une infection dépend sensiblement de la dynamique des individus *malades*. De ce fait, la mise au point réaliste de la mortalité de l'infection est à approfondir.

3.3. Soins, Vaccins et Tests

Dans une première conception, on peut simplifier les paramètres liés aux soins, aux vaccinations et aux tests. Toutefois, il est important de rendre la digitalisation réaliste et de spécifier les éléments en détail :

- Préciser, par une variable fixe ou aléatoire, le délai en jour dans la digitalisation avant que l'élément soit disponible pour les individus.
- Contraindre l'individu à se déplacer dans une catégorie d'espace en particulier pour percevoir cet élément. Auquel cas on pourrait ajouter le temps moyen de visite, et attribuer un rôle nécessaire à ce déplacement.
- Ajouter un délai avant l'apparition des effets de l'élément, que ce soit le temps d'effet d'un soin ou le temps d'attente des résultats d'un test. Le cas de la vaccination est plus intéressant à traiter : chaque dose d'un vaccin a un effet sur l'organisme, plus ou moins rapide. Un de ses effets pourrait être de diminuer drastiquement les paramètres gaussiens lors de l'affectation du score infectieux. Comme nous l'indiquent les personnels médicaux, on peut véhiculer un agent infectieux même vacciné ; on s'approchera alors d'une infection *asymptomatique*. Mais pour rappel, on ne s'intéresse pas au cas de réinfection après guérison ou vaccination dans l'immédiat.
- Indiquer la proportion des effets positifs ou négatifs de chaque type de soins ; en référence à une diminution ou augmentation du score infectieux, dont la valeur devra être précisée. Par simplicité, on se propose de renseigner un montant fixe sur l'augmentation ou la diminution du score. Ces deux montants pouvant bien évidemment être différents.
- Indiquer une matrice de confusion pour les tests infectieux, qu'on pourra simplifier dans un premier temps. La question sur la durée de validité des tests serait également intéressante à traiter. D'autant plus qu'elle peut être incluse dans les décisions territoriales, et donc évoluer au cours de la digitalisation.
- Réfléchir à la conception quantitative des éléments disponibles, sur le temps et la géographie.

3.4. Détails intéressants

3.4.1. Affectation du score infectieux

Lors de l'initialisation du score infectieux des *malades*, il est intéressant de spécifier plusieurs éléments. Il est nécessaire de vérifier la validité de l'utilisation de la loi normale : on ne souhaite pas obtenir de valeur négative pour le score infectieux. Le souci se situe dans l'affectation des seuils avec l'objectif de retrouver des statistiques de distribution similaires aux données observées.

Comment faire lorsque les paramètres gaussiens produisent une proportion non négligeable de valeur négative ?

Deux options permettent de corriger ce problème :

- Réinitialiser la variable autant de fois jusqu'à obtenir une valeur valide.
- Prendre l'opposé de la valeur négative et effectuer en quelque sorte un effet de miroir sur la distribution normale par rapport à la valeur nulle.

Dans les deux cas, il est nécessaire de réétudier la distribution effective de loi de probabilité. Il faut donc ajouter cette problématique dans la recherche des paramètres gaussiens optimaux à partir de données épidémiques observées.

Même si la loi gaussienne semble être tout indiquée d'un point pragmatique, on pourrait laisser la porte ouverte à l'essai d'autres lois de distribution. D'autant plus qu'on pourra toujours adapter convenablement les seuils de façon adéquate.

3.4.2. Flexibilité de la gravité infectieuse

Un autre point intéressant à noter est qu'il est toujours possible de paramétrer les seuils pour enlever le cas d'une infection *asymptomatique*. Par exemple, on peut fixer le seuil entre les *malades asymptomatiques* et *standard* par la valeur nulle. A l'inverse, il est possible de reproduire l'absence de cas *graves* en définissant un seuil supérieur trop élevé. Même si le cas est plus anecdotique, on peut également retirer le cas des *malades standard* en choisissant la même valeur pour les deux seuils séparant les états infectieux. En résumé, cette conception permet de moduler assez aisément les différentes gravités infectieuses.

Dans cette conception, une fois le score infectieux initialisé, on peut uniquement observer une transition de l'état *standard* à *grave* au cours de l'infection d'un individu. Dans tous les cas, l'infection se termine une fois son score infectieux nulle, qu'il soit en état *asymptomatique*, *standard* ou *grave*.

3.4.3. Complexité du cas asymptomatique

Il existe une certaine difficulté dans la digitalisation d'une épidémie en ce qui concerne les cas asymptomatiques. Comment connaître les données réelles de ces malades ? Il serait intéressant d'y réfléchir pour la poursuite du projet. Dans tous les cas, il est nécessaire de les représenter pour avoir une conception réaliste : une épidémie est difficilement contrôlable sans s'intéresser aux individus transmettant l'agent pathogène et ne présentant aucun symptôme.

3.5. Gestion mémoire des états individuelles

Dans la poursuite d'une implémentation désolidarisée, il est plus organisé et intuitif de regrouper la gestion des états infectieux des individus dans la couche infectieuse. On retrouverait alors une organisation avec une table de hachage et des clés composés des identifiants individuels et des différentes infections de la base de données. Pour chaque individu, il est nécessaire de distinguer deux états distincts : celui réel et celui connu. L'état connu est plus une question d'étiquette simple. Celui réel est plus complexe à modéliser. Il est nécessaire d'implémenter et d'automatiser la mise à jour de l'état infectieux, en prenant en compte les spécificités de chaque état. Il convient donc de dissocier la gestion de l'état réel lorsqu'il se situe en *susceptible* de celui *malade porteur*, tout en automatisant la transitions entre les états.

4. Exploitation des digitalisations

4.1. Digitalisation sur le temps

L'enjeu de ce projet repose sur la digitalisation de la circulation d'agents infectieux dans la mobilité d'une société. Pour cela, il convient d'initialiser une situation initiale de la société et de la faire évoluer jour par jour jusqu'à un certain point final à définir. L'état initial peut être obtenu en initialisant l'environnement avec les différents éléments cités précédemment et contaminer un certain nombre d'individus, de façon déterministe ou aléatoire. De l'autre côté, on peut arrêter la digitalisation sous différents critères :

- On peut définir un nombre prédéfini de jour à digitaliser et évaluer l'état final de la société. C'est un des points importants dans l'objectif de l'optimisation combinatoire : restreindre l'espace de recherche sur le temps. D'un point de vue algorithmique, il semble de toute façon évident de fixer une limite au-delà de laquelle la digitalisation s'arrête.
- Dans un optique d'optimisation, on peut établir des conditions d'arrêts selon la situation de la société. Pour faire un parallèle avec un jeu vidéo, cela pourrait représenter des conditions de victoire de la gestion. A l'inverse, on peut également fixer des conditions de défaite où la gestion n'est pas suffisamment optimale pour maintenir un état minimum de la société. Parmi les conditions de victoire, on peut identifier des bonnes évaluations sanitaires de la société, où il ne reste que peu de contamination des agents infectieux. D'un point de vue utopique, on pourrait se dire que la victoire serait d'éradiquer complètement les agents infectieux de la société. Cela semble toutefois peu imaginable compte tenu de l'influence extérieur de la région dont notre gestion subit les conditions sanitaires. Du côté des conditions de défaite, en cas de grave crise économique de la société, on pourra arrêter la digitalisation. De même, on peut ajouter des conditions sanitaires trop critiques pour la société, que ce soit en termes de mortalité, de contamination ou de ressources disponibles.

Une unité d'observation digitale représente l'évolution sanitaire et économique de la société et de ses individus pendant un laps de temps défini. On définit cette période de temps en nombre fixe de jour, qui reste identique pendant toute la digitalisation. C'est à la fin de chacun de ses diagnostics qu'il est approprié de modifier ou non la stratégie politique. D'un point de vue réaliste, cela semble d'autant plus lucide qu'une politique ne peut définir une stratégie au jour le jour :

il faut prendre en compte l'adaptation des individus. Le projet se doit d'être pragmatique pour être intéressant.

Pour diagnostiquer au mieux le statut de la société, il est judicieux de ne pas se limiter à une évaluation sanitaire et économique finale à la fin de chaque observation. On pourra étendre les données de chaque observation avec les évaluations sanitaires et économiques quotidiennes. Ces données sont en particulier pertinentes dans le cas où les jours respectent une logique calendaire et où l'on retrouve une tendance sur la dynamique de la digitalisation. Chaque observation aurait ainsi un ensemble de données journalières à analyser pour pouvoir déterminer les décisions adéquates à prendre pour la prochaine période d'observation. Ces diagnostics pourraient être approfondis dans l'optimisation combinatoire avec l'apprentissage par renforcement.

En pratique, l'implémentation de cette digitalisation n'est pas si simple à effectuer. Il est nécessaire de regrouper tous les éléments de la conception cités précédemment dans une implémentation simple, lisible et efficace. Il est nécessaire d'enregistrer les résultats des digitalisations au fil des jours. La gestion mémoire des données pourra être enrichie dans les implémentations statiques. En associant le numéro du jour aux tables de hachage, on peut ainsi analyser les résultats des digitalisations dynamiques, associées à chaque individu et espace. L'intérêt est d'observer les temps d'occupations dans les espaces et d'évaluer les contaminations ainsi que les stabilités économiques.

Dans la table de hachage des espaces, on trouve une liste d'identifiants d'individus triés en fonction de leur temps d'arrivée et associés à chaque espace et jour de la digitalisation : on y retrouve alors l'identifiant des individus ayant occupé l'espace, le rôle/importance de ce déplacement ainsi que leur temps de sortie. Du côté de la table de hachage des individus, on retrouve les scénarios mis à jour en retranchant les durées des trajets digitalisés : on retrouve alors une liste triée avec le temps des arrivées, l'identifiant des espaces et le temps des sorties.

En parallèle de ces données, la digitalisation enregistre également les évaluations sanitaires des individus dans la couche infectieuse elle-même. La mise-à-jour des états infectieux des individus est d'ailleurs quasi autonome dans cette même couche. L'idée est de décentraliser la gestion mémoire des données en direction des couches appropriées. On retrouve donc le même principe avec la gestion des données concernant les décisions territoriales.

En règle générale, il sera toujours intéressant de conserver les données des digitalisations, ne serait-ce que pour pouvoir analyser les fragilités de la

conception. On pourra toujours se poser la question de la pertinence de conserver des données. La question est plutôt : avons-nous vraiment la prétention de croire que la conception est suffisante et ne nécessite pas d'être perfectionnée ? Or, plus on peut retracer le processus et plus on peut raisonner.

Déjà, une optique non négligeable de la digitalisation est la représentation statistique et graphique des résultats. Et par "résultats", il est implicite d'en extraire les résultats intermédiaires de la digitalisation tel que la dynamique de la société.

4.2. Evaluation infectieuse

4.2.1. Contamination

L'évaluation infectieuse est tout d'abord calculée à l'échelle de l'individu, au sein de la couche d'implémentation infectieuse.

Au cours de la digitalisation, un individu *susceptible* accumule des particules infectieuses lorsqu'il est en contact avec des individus *porteurs* ou lorsqu'il est dans un espace chargé en particules. Au terme de chaque jour, pour chaque individu *susceptible*, on peut retracer une suite d'espace avec des données sur :

- Les durées d'occupation : les temps précis d'arrivée, les temps précis de départ.
- Les identifiants des individus partageant les mêmes espaces pendant les durées d'occupation ; on aura alors fait un scanner pour identifier les *malades porteurs* parmi eux.
- L'aptitude à faciliter la proximité et la contamination de la catégorie de chaque espace.
- Les quantités de particules présentes dans chaque espace à l'arrivée de l'individu, selon le paramètre de conservation des particules de chaque infection dans un espace.
- Les surfaces de chaque espace visité ; ce paramètre permettra de nuancer directement la contamination de l'individu dans un espace chargé en particules en restant pragmatique.

Il est nécessaire d'approfondir la réflexion sur la contamination à l'extérieur d'un territoire avant d'en parler. D'autres données sont également intéressantes à intégrer dans l'évaluation infectieux journalier d'un individu :

- La contamination dans les transports en commun : sous la forme d'un score simpliste selon le type de transport et la fréquence de leur utilisation par des individus *porteurs*. On pourrait voir pour détailler cette considération d'un point de vue plus individuel.
- Le nombre de particules accumulées dans chaque individu sur les derniers jours selon le paramètre de conservation de chaque infection.
- La propension de chaque individu, en particulier interne, à se déplacer à l'extérieur de la région. On peut observer deux phénomènes : plus un individu circule fréquemment dans le territoire et mieux on peut estimer une considération aléatoire à l'extérieure ; et moins un individu circule fréquemment dans le territoire, plus il est aberrant de trouver une forme de logique dans la contamination extérieure.

L'évaluation individuelle fait donc la somme pondérée de ces variables selon une formule générique pour obtenir un score de contamination individuel, pour chaque infection de la base de données. Lorsqu'au moins un de ces scores de contamination dépasse le seuil de contamination d'une des infections, qui peut dépendre de l'âge de l'individu, celui-ci devient *porteur* de cette infection et change d'état pour être en *phase d'incubation*. Pour rappel, il devient alors "immunisé" pour les autres agents infectieux de l'environnement tant qu'il est *porteur*. De même, dans une version simpliste, on ne s'intéresse pas au cas où un individu peut être réinfecté lorsqu'il en l'état *vacciné* ou *guéri* : l'individu devient également "immunisé" à cette infection. Bien entendu, même si la trace objet d'un individu *décédé* peut persister, il est inutile de traiter leurs cas.

4.2.2. Evolution des états

Lorsque la fonction d'évaluation individuelle indique une contamination ou lorsque l'individu change d'état infectieux au cours de la simulation, les réflexes infectieux de la personne changent et modifient sa mobilité. Il est nécessaire de mettre à jour ses propensions en termes de soins, de déplacements non nécessaires, etc. De manière générale, cette phase est automatiquement déclenchée par la couche d'implémentation infectieuse et transmise indirectement à la couche de digitalisation dynamique. Il est convenu d'appliquer une opération généraliste et automatique pour prendre en compte ces mises à jour, au cas par cas pour chaque individu.

4.2.3. Scores d'évaluations territoriales

Pour résumer les évolutions infectieuses au terme de chaque jour, il est nécessaire d'implémenter des fonctions spécifiques où l'on place son observation sur la société. Ces fonctions dépendent de plusieurs variables à pondérer selon certains poids. Pour chaque infection, on peut distinguer les variables suivantes :

- La quantité ou la proportion de la population *susceptible* réels et connus
- La quantité ou la proportion de la population *porteur* réels et connus
- La quantité ou la proportion de la population *guérit* ou *vacciné*
- La quantité ou la proportion de la population *malade grave*
- La quantité ou la proportion de la population *décédés*
- La variation immédiate du nombre de contamination réels et connus
- Un score sur la marge disponible des soins pour les *malades standard*

- Un score sur la marge disponible des soins pour les *malades graves*
- etc.

Ces différents scores d'évaluation permettront de décrire la situation d'une société, sous une multitude d'aspects. On ne se limiterait pas à un simple score censé représenter un pseudo-équilibre, immuable quelque soit les conditions. On partirait alors sur réelle optimisation multicritère.

4.3. Evaluation économique

4.3.1. Entreprises

L'évaluation économique des espaces peut être effectuée à partir de l'affluence des individus et caractéristique propre à chaque catégorie d'espaces. On retrouve une fonction d'évaluation générale où des variables doivent être insérés en paramètres en fonction de la catégorie de l'espace. On peut ainsi individualiser le fonctionnement économique de chaque espace pour avoir des optimisations plus concrètes et applicables. Par exemple, la fonction d'évaluation d'un bureau sera différente d'un restaurant. Et bien évidemment, certaines catégories n'ont même pas de fonction économique, comme les parcs publics par exemple. Dans la notion de fonction économique, on parle bien ici de l'opportunité de l'espace à rendre productif le travail d'un individu dans son environnement. De toute façon, il serait tout autant intéressant d'étudier correctement la question de l'économie de nos jours. Chaque espace est donc visité par un certain nombre d'individus ayant chacun un rôle défini. Certains peuvent travailler dans cet espace et avoir une dépendance financière vis-à-vis de lui. Tandis que d'autres peuvent l'utiliser dans une autre dépendance personnelle, ou sinon comme loisir. L'évaluation économique d'un espace dépend donc de données sur :

- La catégorie de l'espace et ses caractéristiques économiques.
- La durée totale des individus ayant travaillé dans l'espace.
- Le nombre d'individus employés dans l'espace.
- L'aire de l'espace.

Selon les différents types de fonctionnalité économiques, on pourra ajouter d'autres données sur :

- Le nombre de personnes ayant utilisé/visité l'espace.
- La durée totale des individus ayant travaillé en télétravail. La question concernant le facteur de production économique associé au télétravail est toutefois à discuter parce qu'elle est hautement subjective.

On pourra citer quelques fonctionnalités économiques différentes tel que :

- Les bureaux dont on pourra représenter la dépendance économique uniquement avec l'occupation des espaces par les employés. Il y aurait alors possibilité d'aménager la gestion des compagnies en permettant le télétravail de leurs employés depuis chez eux. Il y a une corrélation directe et plutôt simple entre la durée totale de travail des employés et l'évaluation économique de l'espace.

- Les commerces qui ont une dépendance économique très sensible à la présence des clients. Plus il y a d'individus qui visitent ces espaces et plus leur évaluation économique augmente. Bien évidemment, cette évaluation est également très corrélée avec la présence d'employés qui vendent leurs produits. On pourra éventuellement s'interroger sur l'efficacité de disposer de plusieurs employés en même temps. Mais on s'engage peut-être trop loin dans la modélisation. Dans une version ultérieure, à l'instar du télétravail, il pourrait être intéressant d'ajouter une option de livraison où les commerces envoient leur produits directement aux individus. On ne s'intéressera pas à priori aux modalités de livraison.

4.3.2. Individus

Il est essentiel de s'intéresser également à l'économie du point de vue de l'individu. Il ne faut pas oublier l'objectif de ce projet et ne pas trop entrer dans une problématique individuelle où l'on cherche à optimiser les revenus de chacun. L'intérêt ici est d'évaluer la différence entre l'état initial de la société et de chaque individu par rapport aux états de restrictions. C'est-à-dire qu'on ne cherche pas à mettre un score sur les revenus de chaque individu mais plutôt d'observer la perte en temps de travail de chacun en cas d'épidémie. C'est un point de vue encore une fois discutable puisqu'il faut pouvoir reproduire la distribution du travail des individus. A l'échelle d'une société, il semble trop difficile de reproduire la réalité. Donc l'intérêt de l'évaluation individuelle économique est de mettre un score en prenant du recul à l'échelle d'une population.

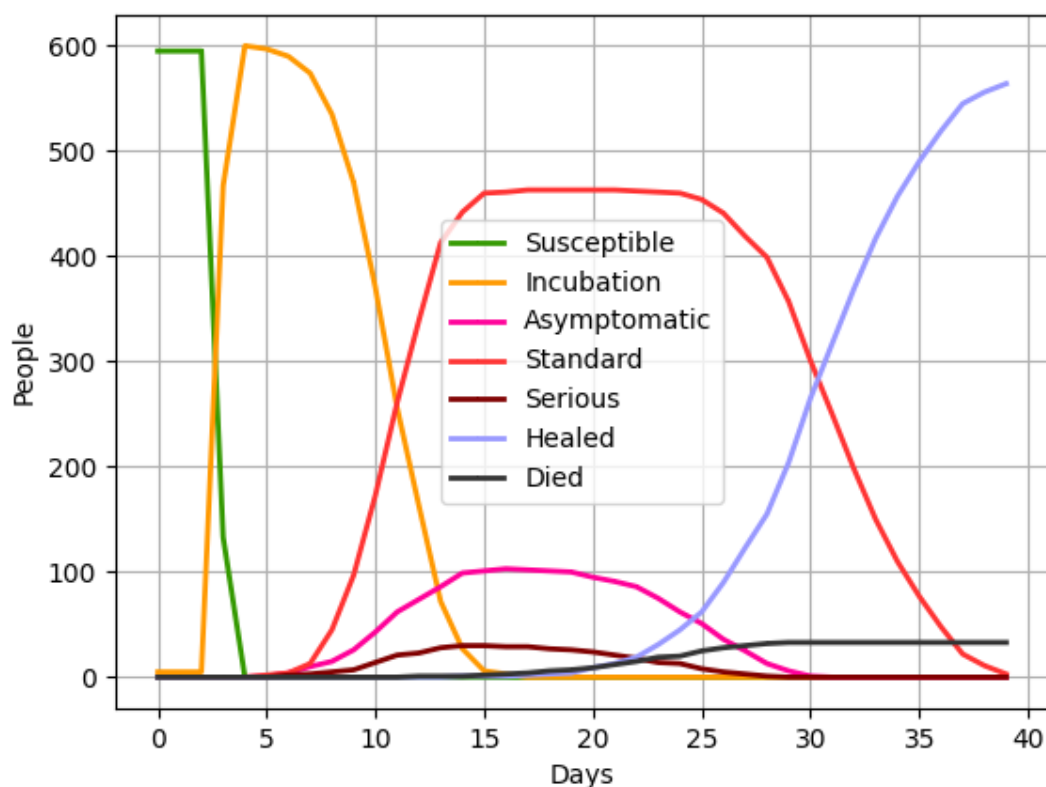
A noter qu'il y a des éléments très importants à observer avec cette évaluation individuelle. Il sera judicieux par exemple de s'intéresser aux cas des enfants et du personnel médical pour apporter des analyses essentielles à la gestion épidémique, notamment sur la stabilité du système éducatif.

4.3.3. Scores d'évaluations territoriales

A l'instar de l'évaluation globale sanitaire, il faut pouvoir renseigner des scores économiques illustratifs à chaque état de la société, pour utiliser au mieux l'apprentissage par renforcement. Avec l'économie, on cherche avant tout à stabiliser la production plutôt qu'à l'optimiser. C'est pourquoi deux éléments prioritaires doivent intervenir dans l'évaluation globale : la production moyenne économique de la société et la dispersion des productions entre les différents espaces et individus.

5. Exemple d'application illustratif

Dans un souci de pertinence, il n'est pas intéressant de se préoccuper de l'exemple très simple du code implémenté. Pour contextualiser le rendu pratique, il est possible à l'heure actuelle d'implémenter une digitalisation statique en utilisant n'importe quel fichier *openstreetmap* et en spécifiant quelques paramètres arbitrairement pour la démographie du territoire. La digitalisation dynamique est très simpliste et effectuée totalement aléatoirement. L'implémentation infectieuse permet une gestion automatique des transitions infectieuses. Le rendu concret correspond donc à la digitalisation du campus universitaire de Belval, à Esch-sur-Alzette, sur un nombre de jours prédéterminé, et avec une initialisation infectieuse à définir arbitrairement. La base de la digitalisation est vraiment à consolider, que ce soit celle statique ou dynamique. Ce n'est que le début du travail. Mes dernières expérimentations étaient de jouer avec les paramètres d'une infection, notamment le seuil de contamination, et d'observer empiriquement les évolutions infectieuses de la population sur un graphique.



Un exemple illustratif des courbes infectieuses avec l'implémentation actuelle.

Rapidement, on peut avoir un aperçu de la conception sur la digitalisation d'une épidémie avec le graphique ci-dessus. Sans rentrer dans les détails, ce test a été réalisé avec les données cartographiques du campus Belval à Esch-sur-Alzette. Six cents individus ont été implémentés, sur une durée de quarante jours. Chaque jour leurs dynamiques sont totalement aléatoires et couvrent la totalité du réseau. On peut donc voir ça comme un graphe totalement connexe où tous les individus se croisent nécessairement. Il n'y a aucune considération extérieure ici. Le test s'est effectué avec cinq individus en *phase d'incubation* dès le premier jour.

On observe une contamination très peu réaliste. Même si tous les individus se côtoient quotidiennement sans exception, il semble peu rationnel de voir l'ensemble de la population se contaminer aussi rapidement ; d'où l'intérêt d'implémenter une forme d'élasticité dans le seuil de contamination. En effet, le profil dynamique est quasiment identique pour tous les individus : on retrouve donc des scores de contaminations trop similaires pour observer une dispersion des infections. Après, on peut retrouver un graphique assez connu avec l'actualité. On retrouve une augmentation exponentielle des cas *malades* ainsi qu'un "plateau". Dans la réalité, on n'identifie pas les *malades asymptomatiques* et le réseau de proximité entre les individus est nettement moins connexe. On pourra bien entendu revoir les proportions entre les gravités infectieuses ; il n'a pas été essayé de reproduire des données réelles de la Covid-19. De plus, aucun soin n'a été implémenté, d'où la fatalité mortelle des individus en état *grave* et l'augmentation du nombre d'individus *décédés*. Les vaccinations n'ont pas non plus été implémentées ici.

La réalisation pratique et tout ce qui entoure sa production pourront être traités lors d'une soutenance. On pourra alors aborder la gestion structurelle des gestions mémoires, les décisions pour la digitalisation statique ou des résolutions de problèmes dans la dynamique de la société. Ce sera bien plus explicite et stimulant que d'évoquer toutes les réflexions théoriques.

6. Complexités de l'environnement

6.1. Overfitting de la réalité

A trop vouloir modéliser, on peut perdre en logique globale et en robustesse. D'autant plus que plus on cherche à modéliser en détail, plus on pourrait être amené à s'intéresser aux données personnelles des individus ; ce qui n'est pas le souhait ici.

On pourra éventuellement s'intéresser à des représentations plus fines de la société même si on parle alors d'un niveau de détail très poussé. C'est le cas par exemple des contaminations lors des transports. Cela serait donc directement en lien avec les capacités de la modélisation *SumoMobility*. Par exemple, lorsqu'une infection à transmission aérienne émerge dans la société, aucun individu n'est préparé à se protéger des contaminations et à éviter la proximité. La proximité fait partie intégrante du fonctionnement collectif mais l'usage de masques sanitaires n'est pas omniprésent, en particulier au début de la propagation. La question à se poser est donc de savoir s'il est vraiment nécessaire de modéliser les contaminations à l'air ambiant. Toutefois, on pourra faire un parallèle avec un overfitting pour faire le lien entre un réseau de neurones et la digitalisation dynamique. *SumoMobility* est conçu pour effectuer des déplacements pedestres d'un point A à point B sans s'arrêter. A la rigueur, on pourrait trouver un repère en observant les collisions des individus sur une même voie piétonne ; il serait alors nécessaire de lancer une exécution interactive, qui pour rappel peut sensiblement provoquer des erreurs et des incertitudes aux niveaux des scénarios entre les espaces.

D'un point de vue réaliste, on pourrait se questionner aux contaminations au sein des voitures personnelles lorsque le conducteur est accompagné. Il semble aberrant d'approcher un tel niveau de détail. Ce serait un détail non essentiel, qui diminuerait finalement la robustesse de la modélisation. D'autant plus que les personnes qui accompagnent le conducteur sont alors très souvent liées par des liens familiaux ou se côtoient déjà régulièrement dans les espaces que l'on considère.

6.2. Evolution et concurrence des infections

Dans l'environnement, la coexistence de plusieurs infections a lieu selon deux évènements différents :

- La mutation aléatoire d'un agent infectieux qui existe déjà dans l'environnement ; on caractérisera alors chaque infection par des paramètres liés la fréquence de mutation lors d'une nouvelle contamination, et par son ampleur sur ses caractéristiques originales. Par simplicité, l'apparition d'une mutation aurait lieu au moment où un individu dépasse un seuil de contamination et qu'il entre en *phase d'incubation*. Selon une loi de Bernoulli, il peut devenir alors l'hôte d'un nouvel agent infectieux avec des caractéristiques voisines de l'infection qu'il l'a contaminée. Parmi ces nouvelles caractéristiques, il est possible d'observer un seuil de contamination différent de celle originale, qui serait alors appliqué au prochain individu que celui-ci contamine. Il est bien entendu envisageable de retirer la possibilité d'une mutation.
- L'introduction manuelle et programmable d'un nouvel agent infectieux, avec des caractéristiques totalement indépendantes des infections actuelles. En dérivant un peu le concept, on pourra alors programmer manuellement le variant d'une infection préexistante, en maîtrisant donc les nouvelles caractéristiques modifiées. Cette introduction peut être programmée à un jour particulier dans la simulation, ou alors lorsque la société atteint un certain état.

A noter qu'il suffit de réduire le seuil de contamination d'une infection pour rendre l'agent pathogène plus contagieux, réduire le seuil du niveau de son état *grave* pour aggraver sa virulence et d'augmenter le temps de persistance des particules dans l'organisme ou dans les espaces pour augmenter sa résistance. On peut donc assez facilement faire évoluer chaque agent infectieux. C'est ce qui apporte une souplesse à l'implémentation infectieuse.

6.3. Diversité de la société

L'environnement doit permettre d'inclure des individus aux comportements peu ordinaires. L'intérêt est d'amplifier le mouvement brownien de la propagation épidémique dans la société et de rendre la digitalisation de la société très imprévisible. Par exemple, on peut représenter les visiteurs dans les prisons afin de mieux initier la propagation dans les établissements pénitentiaires. Bien que l'évaluation individuelle peut être compliquée, on peut également représenter les livreurs qui traversent beaucoup d'espace dans une journée en y restant très peu de temps. Dans un type de quotidien similaire, on peut aussi représenter les services d'urgences tels que les gendarmes, les pompiers ou le personnel médical d'urgence. Sans oublier par exemple les rassemblements entre les amis ou la famille. En règle générale, il faut pouvoir intégrer une grande diversité dans le scénario des individus, en particulier pour les événements insolites. On peut citer l'exemple des lieux diplomatiques, des lieux funéraires, des crèches, des garages automobiles, etc. L'avantage de la base de données des espaces est que l'on peut choisir parmi l'ensemble des espaces répertoriés, sans négliger certains d'entre eux. Il peut donc être intéressant d'introduire des scénarios totalement aléatoires, pouvant même aller à parcourir plusieurs logements dans la même journée. Un peu à l'image des agents immobiliers. Ce n'est certes pas rationnel mais cela peut contribuer au mouvement brownien des agents pathogènes, et potentiellement apporter une meilleure élasticité des dynamiques.

Le projet peut donc s'inspirer au mieux du quotidien des individus, tout en modélisant les comportements non ordinaires par des scénarios totalement aléatoires. On peut dissocier deux représentations différentes :

- Soit chaque individu aurait un certain ratio de scénarios ordinaires et de scénario totalement aléatoire.
- Soit on modéliserait rationnellement une certaine population dont les scénarios seraient totalement aléatoires.

Dans ces deux options, le facteur imprévisible et aléatoire est exécuté de façon significativement différente. Sur le long terme, la digitalisation peut approcher la réalité tout en retirant la complexité infinie de modéliser la diversité des quotidiens. En revanche, sur une courte période d'observation, où l'on souhaite choisir une décision en fonction de l'état de la société, cela peut suffisamment influencer le programme par un paramètre très aléatoire.

6.4. Evolution des paramètres

Dans la digitalisation statique, on pourra s'interroger sur l'évolution des individus. D'un point de vue pragmatique, il ne semble pas intéressant dans une version primaire de modéliser les naissances et les décès non infectieux de l'environnement. Ce serait en revanche une piste intéressante pour un projet annexe démographique. Ici l'intérêt serait plus de faire évoluer le socle individuel de chacun. Il s'agirait de représenter les probabilités de déménagement et les statistiques d'embauche.

Avec l'exécution de la digitalisation dynamique, on peut identifier des problématiques sur le plan économique et sanitaire.

Sur le plan des espaces, il serait intéressant de retirer les lieux économiques en faillite due à des décisions trop restrictives. Il faudrait alors pouvoir produire une évaluation critique à partir de laquelle l'espace serait retiré dans l'élaboration des scénarios.

Beaucoup de paramètres sont importants à prendre en compte sur le plan infectieux. Tout d'abord, l'évolution des ressources disponibles, qui peut représenter un important espace de définition. De plus, avec les soins, tests et vaccination, on choisira le point de vue rationnel de fixer leurs paramètres sur la durée de la digitalisation. Mais il est nécessaire de réfléchir aux paramètres de ces éléments lorsqu'on en ajoute au cours de l'exécution : c'est le cas possible lors de l'apparition d'une nouvelle infection par mutation. Toute la réflexion à avoir doit se faire sur la programmation manuelle des événements aléatoires ou stochastiques. Et dans les deux cas, en quoi cela pourrait approcher la réalité ? Comment retrouver une forme de prédiction ? Comment traiter les données épidémiques historiques ?

Davantage de problèmes devraient sensiblement apparaître dans la suite des travaux de l'implémentation.

6.5. Incidence des territoires extérieures

Le débat à plus ou moins était évoqué le long de ce rapport. Les thématiques à s'intéresser sont :

- Les contaminations à l'extérieur pour les individus internes.
- Les ressources sanitaires disponibles à l'extérieur.
- Les demandes en ressources sanitaires depuis l'extérieur.
- Les décisions sur les frontières.
- Les propensions extérieures dans les dynamiques individuelles.
- Les paramètres aléatoires des états d'infections réels des individus externes.
- L'évaluation des états connus des individus externes.
- La quantité d'individus externes à implémenter ; autrement dit la quantité de travailleurs dans le territoire qui proviennent de l'extérieur, en particulier en termes d'ajout par rapport aux travailleurs internes. Dans cette même optique, il faut également considérer les individus internes qui travaillent à l'extérieur.
- etc.

Dans l'immédiat, il est nécessaire d'approfondir les réflexions.

6.6. Contraintes avancées des scénarios

La problématique de la conception des scénarios est complexe. Sans parler de la diversité des paramètres, il y a les contraintes temporelles, les estimations des trajets, les structures types, etc. Cette complexité doit pouvoir retracer des données observées en termes d'affluence dans les espaces, respecter des logiques socio-dynamiques, etc. C'est déjà quelque chose de sympa à traiter.

Maintenant, pour ajouter du réalisme et une justesse des décisions/optimisations, on pourra ajouter d'autres contraintes, nettement moins amusantes à première vue. On ne les détaillera pas plus puisque les complexités sont implicites :

- Une quantité limitée d'individus présents simultanément dans un espace.
- Le besoin de se diriger dans un espace et de devoir respecter un créneau disponible spécifique, en lien avec les limitations de places.
- Les horaires d'ouverture spécifiques de certaines catégories d'espaces.
- Les préférences des individus sur la sélection d'un espace dans une catégorie définie, à l'échelle collective pour éviter de se rapprocher de données personnelles.

6.7. Respect des gestes barrières

Comment représenter le respect des gestes barrières ? Y-a-t-il vraiment une typologie d'individus qui ne les respectent pas dans la vie réelle ?

Et surtout, est-ce que cela représente réellement un intérêt en termes de décisions territoriales ?

L'implémentation demanderait plus de temps de réflexion. Mais dans le cas où il semblerait judicieux de le considérer, on pourrait apporter un facteur aléatoire, par jour et par individu, qui influencerait les évaluations infectieuses. Elle augmenterait le paramètre d'émission des *malades* et diminuerait le seuil de contamination des infections des *susceptibles*. Le respect des gestes barrières fonctionne dans les deux sens ! Un point intéressant est la typologie des individus. Est-ce que ce sont les mêmes individus qui ont tendance à ne pas respecter ces gestes ou devons-nous plutôt représenter une diversité aléatoire pour approcher la réalité ?

6.8. Tester, alerter et protéger

Au terme de chaque jour, on peut identifier toutes les personnes ayant partagé un espace en même temps. Toutefois, comme les clients d'une boutique, il n'est pas pertinent, ni réaliste d'espérer remonter la liste exhaustive des individus. Pour faire donc le lien entre l'individu en *phase d'incubation* et les cas contacts, on pourra trouver une stratégie rapprochant la réalité : on alerterait les individus identifiés dans le socle du nouvel infecté, et on retrouverait les individus ayant partagé un temps suffisamment long en sa compagnie. Il serait nécessaire de préciser plusieurs paramètres :

- Le nombre de jours pendant lesquels on doit rechercher les cas contacts, non associés au socle individuel.
- Le seuil de durée où l'on considère les individus à proximité comme cas contact.
- Le répartition des contacts pendant le période de recherche ; si l'individu à proximité a été en compagnie avec le nouvel infecté en une fois ou en plusieurs fois dans plusieurs espaces.
- La décision à appliquer sur les cas contacts : une contrainte sur les déplacements, une obligation de tests, etc. A noter qu'on pourra ajouter ce statut en tant qu'un état infectieux à part entière dans une version ultérieure. On attribuerait alors des décisions combinatoires aux individus Alertés.

7. Homologation avec des données réelles

L'idée est d'utiliser des données saines vis-à-vis de l'humain. Il est tout à fait possible de modéliser, reproduire et anticiper la réalité, uniquement à l'aide de données environnementales. L'une des contraintes de ce veut donc de ne pas utiliser des données personnelles tel que des traçages GPS. On peut trouver un nombre conséquent de données saines dans un territoire :

- La répartition des âges de la population locale.
- La répartition de la taille des ménages.
- Certaines affluences courantes des espaces, en fonction des paramètres calendaires.
- Le nombre de voiture par ménage.
- La répartition des travailleurs dans les différents secteurs professionnels.
- Les données cartographies dans leur globalité ; cf. *openstreetmap*.
- Les données concernant les travailleurs frontaliers.
- etc.

Ces données se veulent globales et nécessitent moins d'espace mémoire. Elles peuvent couvrir une grande partie de l'environnement.

Tout l'intérêt est de déterminer les boîtes noires entre elles !

Finalement, il n'y a pas de vérifications comme on pourrait le retrouver dans un réseau de neurones. Il n'y a pas de données tests, ni de validation. Mais plus on possède de données saines, et plus il est intuitif de trouver un lien. Dans cette recherche, le plus simple est de regarder les causes et les conséquences de l'environnement réel ; on peut identifier des processus. On peut alors lancer des simulations avec certaines conditions initiales et comparer avec des données observées dans la réalité. Il est ensuite possible de revoir petit à petit les paramètres de la digitalisation pour se rapprocher d'une réalité ; tout en s'éloignant des données personnelles et en gardant une éthique humaine.

8. Décisions combinatoires

Il n'a pas encore été évoqué précédemment que plusieurs décisions peuvent être prises pour une même période. Les impacts de décisions peuvent être significativement différents et il est donc possible d'additionner certaines d'entre elles. On pourra citer par exemple une décision portant sur la validité des tests infectieux, avec un couvre-feu à partir de 18h.

Il pourrait être intéressant de considérer le non-respect de ces règles avec un taux de respect dans une version ultérieure : à chaque création de scénario, l'individu pourrait éviter ces contraintes selon une probabilité; c'est-à-dire que ce non-respect serait appliqué indépendamment du jour et de l'individu, selon la catégorie de l'individu. Bien entendu, le champ d'action de ce non-respect dépendrait de la gravité/ampleur de la décision. Mais il semble aberrant de caractériser un sous-champ d'action, qui s'appliquerait spécifiquement sur un créneau temporaire d'un scénario.

Dans l'immédiat, l'implémentation n'est pas suffisamment constructive pour davantage s'intéresser aux décisions. Ce serait produire des résultats aberrants, qu'on chercherait à optimiser...

8.1. Contraintes sur les individus

Pour appliquer une décision spécifique aux individus, il faut s'intéresser aux caractéristiques extérieures de l'individu. L'idée est de faire le parallèle avec la réalité. On ne peut donc pas utiliser les classes socio-dynamiques implémentées dans l'environnement. On pourra alors séparer les individus par leurs âges, leurs états infectieux connus et leur organisation dynamique. En effet, on pourra restreindre spécifiquement les individus dont le travail est un commerce ; cela pourrait alors impacter plusieurs classes socio-dynamiques dans l'implémentation par exemple.

8.2. Contraintes sur les espaces

Avec les espaces, il semble toutefois obligatoire d'utiliser les catégories de la base de données pour leurs appliquer des décisions spécifiques. On pourra alors influencer cette sélection par l'aire de chaque espace pour avoir une meilleure classification par rapport à la réalité. C'est le cas par exemple lorsque l'on souhaite appliquer des décisions sur des grands centres commerciaux et laisser les petits commerces libres. C'est à partir de cette limitation de distinction qu'on pourra agencer les catégories. Par exemple, puisqu'à l'origine les clés du fichier *openstreetmap* sont plus dans le détail, on peut redissocier la catégorie des restaurations par les restaurations intérieurs et extérieurs.

8.3. Contraintes sur les scénarios

Les contraintes sur les scénarios sont plus complexes puisqu'elles même dépendant de la digitalisation statique. On pourrait les résumer à des contraintes temporelles telles qu'un couvre-feu ou un confinement ainsi qu'à des contraintes de sélection d'espace. En effet, cela rejoint le cas réel des limitations de distance. Pour rappel, la base de données des espaces peut filtrer elle-même les lieux avec une localisation de départ et plusieurs critères.

8.4. Contraintes sur les états infectieux

Les contraintes sur les infections semblent évidentes dans une optique de gestion épidémique. De manière générale, elles portent sur les états connus, et non pas réels, ainsi que sur les modalités d'accès des soins, tests et vaccins. On pourra ajouter d'autres décisions telles que la validité des tests, dans leur ensemble pour simplifier. L'une des difficultés est dans la prédétermination des soins, tests et vaccins : on ne peut pas prédire ces éléments à l'apparition d'une infection. Peut-être qu'en ayant un meilleur regard sur les processus médicaux, on pourrait avoir une approximation sur leurs délais d'apparitions et leurs caractéristiques ; on peut retrouver des logiques entre elles. La question se pose alors sur le caractère aléatoire.

9. Parallélisation des calculs

9.1. Phase d'initialisation

L'initialisation est certainement la phase la plus longue à cause de la lecture complète du fichier *openstreetmap*. Cette étape est visiblement difficile à paralléliser. S'ensuit l'attribution de la route la plus proche pour chaque espace. Cette étape est nettement plus longue que la lecture du fichier d'import. Toutefois, il est possible de la paralléliser assez aisément : il suffit simplement de répartir la recherche local de chaque espace sur plusieurs centres de calcul.

Ensuite, l'initialisation suit avec la création des individus, des infections et des états infectieux, et éventuellement des scénarios lorsqu'on souhaite les mémoriser. Ces étapes sont bien évidemment parallèles. Il faudra veiller à avoir un “maître” pour respecter les demandes prédéfinies dans l'initialisation.

9.2. Phase itérative

La phase itérative correspond à la suite des digitalisations sur les journées. Elle commence par la création/adaptation des scénarios, qui peut être parallélisée comme dans la phase d'initiation. On utilise alors *SumoMobility* pour traduire ces scénarios sous son format avec *duarouteur*. Cette étape peut être parallélisée par le logiciel lui-même en spécifiant un nombre de threads ; à noter que nous n'avons pas accès à la parallélisation puisqu'elle s'effectue en interne. Il y a ensuite une étape non négligeable qui est celle de retrier les itinéraires en fonction du temps de départ.

Maintenant, on peut lancer l'exécution de *SumoMobility* pour digitaliser la dynamique de la société. Pour rappel, on peut paralléliser en interne cette exécution sous plusieurs threads mais on n'en connaît pas vraiment les détails. Une fois la digitalisation réalisée, on s'occupe alors de l'analyse des résultats. On doit d'abord lire un fichier puis enregistrer les données. Les traitements suivants peuvent quasiment tous être parallélisés puisque les calculs se rattachent à chaque espace ou chaque individu, etc. Parmi ces interventions, on trouve notamment les évaluations économiques et infectieuses des individus/espaces.

10. Perspectives de développement

10.1. Modélisation locale dans un espace

La modélisation à l'intérieur des espaces peut être intéressante à implémenter. Elle n'apporterait toutefois pas de grande évolution dans la digitalisation.

Du point de vue pratique, la plupart des simulations épidémiques modélisent des balles qui rebondissent dans un rectangle. On pourra faire le parallèle avec une circulation chaotique des individus dans un même espace. On pourrait même ajouter des formes infranchissables dans le rectangle pour représenter l'architecture structurelle dans l'espace. Par exemple, on pourrait concevoir les rayons parallèles d'un supermarché. Mais encore une fois, il est très probable de tomber dans un overfitting de la digitalisation, qui pourrait affaiblir la robustesse de la modélisation épidémique en scénarios.

En réalité, l'intérêt dynamique d'une telle considération ne permet pas d'apporter davantage d'assurance sur la pertinence des contaminations dans un espace. D'autant plus qu'aucune donnée ne pourrait réellement soutenir la dynamique ; et il s'agirait probablement de données personnelles, qui serait antipathique aux yeux des individus.

En revanche, d'un point de vue recherche, il serait très intéressant d'étudier la capacité d'un espace à favoriser une proximité/contamination de ses occupants en fonction de son architecture. En particulier, dans notre projet de digitalisation, ce serait plus qu'intéressant pour définir au mieux les facteurs de contamination des différentes catégories d'espaces.

10.2. Exploration des décisions potentielles

Même si les espaces de recherches exploseraient, on pourrait également augmenter les choix combinatoires en remplaçant les décisions territoriales par les contraintes elles-mêmes de la digitalisation.

10.3. Considération multi-territoriale

On se doit d'être bienveillant. La considération ici ne sera jamais d'être dans une concurrence négative avec le territoire voisin.

L'idée est d'optimiser l'ensemble des territoires dans une considération globale. La différence majeure par rapport à la modélisation originale est la séparation/duplication du territoire en plusieurs régions ayant chacune des décisions indépendantes. Le terme "indépendantes" est toutefois à regarder de près. On peut retrouver une forme de théorie des jeux ici, à l'exception qu'on ne veut pas de concurrence négative. Du point de vue l'optimisation bienveillante, seule la concurrence positive, où l'un comme l'autre se développe, est à considérer. On notera que la considération des mouvements interterritoriaux serait à modifier.

10.4. Modification des données cartographiques

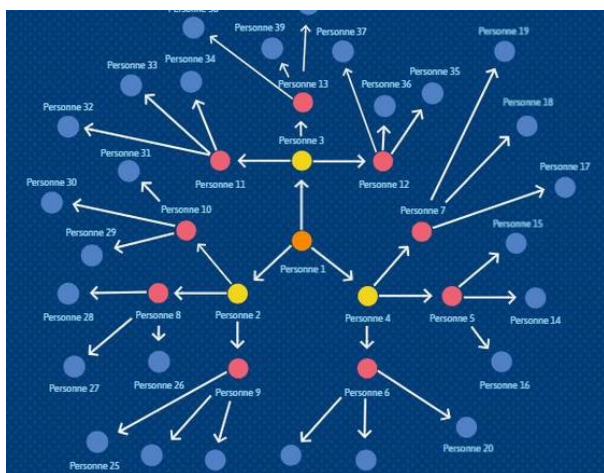
Dans une version ultérieure, il pourrait être intéressant d'ajouter manuellement des espaces pour étendre l'espace de recherche des solutions d'une gestion épidémique. Par exemple, on pourrait s'intéresser à construire un nouvel hôpital, avec une certaine aire et une localisation stratégique, pour optimiser davantage la gestion des malades infectieux. On pourrait alors jouer sur sa localisation et son aire pour observer son potentiel impact dans la digitalisation. Pour cela, on pourra utiliser un package Python nommé *Folium* qui permettrait de positionner les coordonnées stratégiques correctes du nouvel espace en observant la carte de la zone étudiée. Il resterait à définir sa catégorie et son aire. Il serait alors nécessaire de déterminer la route la plus proche, même si on pourra discuter aux raccords entre la localisation de l'espace par rapport à celles des routes du réseau.

On peut trouver une très grande variété de possibilités avec cette nouvelle considération. L'espace de recherche pourrait exploser si l'on ajoute des décisions stratégiques dans la digitalisation statique.

10.5. Ajout du critère psychologique

Une des pistes d'évolution est également l'ajout de la dimension psychologique à la gestion de l'épidémie. On le voit bien avec la réalité du Covid-19 ; on observe une augmentation des dépressions et des suicides à la suite des grandes restrictions. Alors bien entendu, on pourra s'interroger sur l'origine réelle des souffrances psychologiques. Proviennent-elles directement des décisions territoriales ? Dans tous les cas, il est évident que l'instauration de grandes restrictions telles que les confinements ont impacté les états psychologiques des individus. On pourrait donc essayer de trouver une forme d'évaluation psychologique sur la suite des décisions fixées. En effet, il semble y avoir une plus grande cohérence à analyser les décisions dans leur succession plutôt que sur un temps donné. On pourra également ajouter le facteur psychologique de certaines catégories d'espaces dans la dynamique des individus.

10.6. Analyse approfondie des résultats



Afin d'avoir de plus de détails sur la conception et perfectionner la modélisation, il peut être plus qu'intéressant de déterminer des scores épidémiologistes. Par exemple, on pourrait s'intéresser à calculer les scores R_0 des infections ; c'est-à-dire le taux de reproduction du virus dans la société. Cela donnerait plus d'indications pour mieux comparer et paramétrer les infections dans l'implémentation.

De façon générale, il serait intéressant de recueillir davantage de données sur la digitalisation épidémique pour pouvoir tendre la digitalisation vers des paramètres infectieux observés. Ce serait un peu comme la rétropropagation du gradient dans l'apprentissage d'un réseau de neurones, plus manuellement ici.

S'interroger sur l'intelligence

Le vrai savoir est la connaissance profonde de soi et l'empathie. Ça devrait être la base même de l'individu avant de s'intéresser aux domaines scientifiques, littéraires, etc. C'est ce qui distingue l'humain de la machine. Après, il est concevable que c'est bien souvent la souffrance qui nous amène à mieux nous connaître, ou du moins la difficulté. C'est d'ailleurs pour ça qu'on doit parfois apprendre à écouter les plus anciens, qui ont déjà souffert. C'est une des contradictions de la vie : comment faire progresser la connaissance de soi de chacun en maintenant l'empathie ? Sommes-nous tous égaux dans notre capacité à l'introspection personnelle ? Certains pourront évoquer le service militaire... Dans l'intérêt individuel, il serait mieux de simplement la promouvoir correctement auprès des individus pour qu'elle soit volontaire. Le savoir est avant tout volontaire. On pourrait s'intéresser à la retransmission de la sagesse par les anciens ; ce qui apporterait de la dynamique et une motivation aux retraités. Le système éducatif a beaucoup trop tendance à raccourcir ou endiguer le développement personnel de la jeunesse. L'éducation devrait même être revue sous de très nombreux aspects et encadrer les enfants avec plus de bienveillance et moins de façon industrielle. A plein de raison, il serait nécessaire de considérer l'ensemble de notre environnement pour mieux nous connaître, jusqu'aux mouvements planétaires...

Les mentalités des individus sont trop individualistes dans ce monde. Ils se limitent à une expression extérieure, en limitant leur regard sur leur entourage très local. On s'intéresse bien plus à notre façon d'être perçu plutôt que d'être nous-même. C'est d'ailleurs pour ça que les erreurs humaines ont tendance à se répéter, et que le vide en chacun de nous persiste. On a tendance alors à élargir notre entourage, tout en conservant les mêmes mentalités ; d'où l'émergence des réseaux sociaux et des données personnelles. Est-ce qu'il a seulement un intérêt à la surmédiatisation des personnes qui veulent juste s'afficher, promouvoir la débilité et l'irresponsabilité ? C'est toujours étonnant de voir une plus grande popularité de ce genre de médias que pour des médias qui nous développent davantage. De nos jours, même le journalisme n'a qu'un faible intérêt. A partir de quand la population cherchera au bon endroit pour combler le vide personnel de chacun ? Le monde en est matérialiste et froid.

Et si l'on changeait les mentalités ? Cela devrait nécessairement passer par une prise de conscience collective. Sinon, on créerait des décalages psychologiques qui conduisent aux malheurs individuels.

Si on se concentrait tous sur notre expression intérieure ? On serait à même de pouvoir prendre du recul sur le monde dans sa globalité. On serait engagé à réagir aux réels problèmes de l'humanité. On pourrait davantage promouvoir le respect et la compassion auprès de tous. Davantage se considérer d'égale à égale. On enlèverait la peur aux individus de se limiter, qui vient souvent de l'éducation d'ailleurs. On observerait beaucoup moins de tristesse et de regrets dans les yeux des mourants.

On serait alors plus enclin à réellement accepter l'humanité et amoindrir l'orientation matérielle et individualiste du monde. On pourrait s'interroger sur les vrais problèmes d'ordre humain, tel que la souffrance psychologique, l'euthanasie ou le "coût financier d'une vie humaine". De nos jours, l'empathie est trop souvent mise en valeur uniquement lorsque la souffrance vient à nous. Ce n'est que lorsqu'on subit une catastrophe naturelle qu'on se demande où sont les autres. Après, il existe la peine "légitime" qui, tôt ou tard, rappelle à l'ordre les gens malhonnêtes et malveillants...

Toutefois, quoi que l'on fasse, l'entropie est destinée à augmenter avec le temps. Et paradoxalement, l'origine de cette entropie est sensiblement liée à la recherche de l'idée du bonheur propre à chacun ; le terme "idée" est volontaire, et fait référence à la distinction avec le faux bonheur. Mais on sort du contexte de l'intelligence artificielle.

La plus grande force de l'entropie sont les faiblesses humaines.

Le danger : donner du pouvoir à l'expansion de l'entropie.

C'est bien connu, la ressource la plus chère de demain sera les données. L'outil sera l'intelligence artificielle. On peut distinguer deux processus dans le traitement des données :

- Soit par le jugement de l'individu à travers ses données personnelles. On représente l'individu avec des chiffres, et on tente de prédire son comportement factuel ou de le classer. On restreint l'individu à sa nature de machine. Mais aucune intelligence artificielle ne pourra comprendre l'âme humaine. C'est pourquoi la pertinence et le traitement des données personnelles est hautement controversable.
- Soit par l'analyse de l'environnement par des données globales. On modélise l'environnement et on travaille sur des phénomènes humains dans leur globalité. Cela demande un effort de conception pour relier les données entre elles, à la manière d'une boîte noire. Mais ce traitement serait en respect de la condition humaine.

Par exemple, il est probable que peu d'européens voudraient être soumis à la politique chinoise qui tend vers une surveillance de sa population par leurs données personnelles. On peut aussi citer l'affaire *Cambridge Analytica* qui a influencé le libre arbitre d'individus dans l'élection du président d'un des plus puissants pays du monde. Leur utilisation est très répandue à l'échelle d'entreprises commerciales. L'intérêt est-il donc purement économique ? Voir une question de contrôle des individus ?

Il est utopique, et même dangereux de croire qu'une intelligence artificielle pourra tout résoudre par elle-même, sans l'introspection et l'empathie humaine. Le savoir humain est nécessaire.

Il est obligatoire de respecter le vécu de chacun, et souvent de se mettre à la place de l'autre. Mais il semble déjà irrationnel de croire que tout individu peut s'approprier la personnalité et tolérer les choix de tous. La justice de notre monde n'est pas véritablement juste et omnisciente. Un peu à l'image des religions, on peut toutefois retrouver au moins une forme de guide, qui oriente les individus vers des décisions rationnelles et équilibrées. Même si cela nous écarte trop souvent de notre introspection personnelle...

On pointe trop souvent des idées pour expliquer et donner une excuse à la malveillance. On peut parler de surpopulation, de sélection naturelle, de discrimination, de haine culturelle ou religieuse, etc. Mais ne serait-ce pas un manque d'intelligence ? Une incapacité à résoudre les problèmes et un côté fataliste ?

Le monde est trop complexe pour faire respecter une justice divine. Alors pourquoi ne pas anticiper les catastrophes dans leur globalité ? Environnementales, épidémiques, etc.

Un exemple simple : prenons le cas du géant Amazon, qui s'est enrichi pendant la pandémie du Covid-19. L'entreprise ne semble pas véhiculer une bonne bienveillance de l'individu, que ce soit au niveau employeur ou de leurs fonds économiques. Un de leur projet serait de faire livrer les colis par des drones. L'idée est novatrice, il faut le souligner. Mais à plein de raison, il semble y avoir un danger d'un point de vue humanitaire. La question n'est pas de rejeter l'idée. Mais pourquoi ne pas investir cette technologie novatrice dans l'humanitaire d'abord ?! On pourrait par exemple l'utiliser pour mieux gérer les incendies forestiers. Ne serait-ce que pour semer un ensemble de produits ignifuges, et si possible respectueux de l'environnement, aux contours d'un feu pour stopper sa propagation. On voit trop souvent des foyers et des individus souffrir par ces

catastrophes. On pourrait alors s'intéresser à la communication, la logistique, la résistance des matériaux, etc. Et beaucoup de ces technologies seraient identiques par rapport au projet original.

La limite : l'argent. L'argent a une force d'inertie très grande. L'argent provient trop souvent de l'argent lui-même. Il n'a que rarement d'intérêt avec les projets empathiques qui le consume. Mais il est nécessaire de trouver un équilibre qui est parfois difficile à déterminer, avoir ou à maintenir. D'autant plus que cet équilibre pourra toujours être critiqué par l'extérieur. Et la plus grande difficulté est avant tout d'acquérir le socle financier pour initialiser les projets, même à l'échelle d'une vie humaine.

Une projection intéressante

La fin de ce projet serait d'en faire un jeu scientifique. Il devrait être mis à jour régulièrement et être développé au fil du temps. Les environnements seraient de plus en plus perfectionnés/réalistes au fil des mises à jour. Au lieu de confronter les territoires les uns aux autres, on confronterait l'intelligence des joueurs sur un même territoire avec des conditions identiques. Qui optimisera le mieux l'harmonie globale ? On explorerait toutes les dynamiques réalistes possibles de notre monde. On pourrait par exemple implémenter le jeu dans un serveur centralisé et laisser l'accès aux joueurs directement depuis internet. Le jeu serait accessible à tous sans discrimination, et dans la plupart des parties du monde. Il faudrait adapter les performances de calcul pour la fluidité du jeu et pour attirer les joueurs. Mais il sera alors possible d'enregistrer directement les résultats expérimentés dans une base de données. Cet enregistrement pourrait même permettre d'éviter de recalculer des évaluations courantes de l'environnement, et de renvoyer simplement les résultats enregistrés. Plus on participerait et plus les optimisations seraient robustes. On combinerait alors l'exploration des solutions de l'environnement, l'exploitation locale et l'apprentissage par renforcement de nos décisions sur la gestion du monde. Bien entendu, on pourra préférer un intérêt local dans un cas pratique. Ce serait davantage une optimisation locale du monde, et l'objectif serait alors de la rendre mondiale.

On pourrait ajouter des couches d'environnement supplémentaires ? Un environnement climatologique, un environnement sismologique, un environnement éthologique, et tout autre environnement scientifique du monde, en passant par l'étude des volcans, des dangers nucléaires, de la propreté des océans, etc.

Ce serait d'ailleurs une forme d'éducation excellente pour approfondir notre compréhension des sciences ou développer des passions/intérêts auprès de tous, du moment qu'elle est stimulante. Sans oublier que cela pourrait inciter beaucoup de personnes à s'orienter dans l'informatique, les statistiques, l'intelligence artificielle, l'optimisation, etc. On favoriserait une optique bienveillante envers le monde. Les jeunes curieux qui souhaitent apprendre comme les grands experts scientifiques seraient satisfaits de tous les côtés. On pourrait alors rassembler la population autour du débat humanitaire et développer l'intérêt concernant l'anticipation et l'optimisation de la civilisation, en s'orientant dans la bienveillance humaine et le respect de principes moraux. On utilisera l'intelligence collective pour l'harmonie globale. Cela pourrait même inciter les populations des régions les plus pauvres à contribuer aux données locales saines : épidémiques,

démographiques, cartographiques, etc. Et surtout, on pourrait rendre l'intelligence artificielle accessible à tous, en l'orientant dans une optique d'harmonie globale, et non pas qu'à une riche élite qui pourrait être amenée à augmenter l'entropie.

Avec une vraie transparence et une ouverture d'esprit, on pourrait chercher à optimiser :

- La gestion de l'énergie verte.
- La gestion des ressources de la planète.
- La gestion des crises humanitaires.
- La protection des écosystèmes et des êtres vivants.
- L'anticipation des catastrophes naturelles.
- L'anticipation des évolutions démographiques.
- L'anticipation du changement climatique.
- etc.

On pourrait trouver des solutions optimales aux problèmes en apparence impossibles de la civilisation.

ON POURRAIT RÉVOLUTIONNER LE MONDE.



Remerciements

Equipe pédagogique de Polytech Lille.

Je souhaite remercier les personnes qui ont su enseigner avec beaucoup de respect, de bienveillance et une certaine sympathie. Celles intéressées par l'éducation et qui ont bien accompagné les élèves. En particulier les enseignants dans le domaine de l'informatique, de l'optimisation et de l'anglais.

El-Ghazali Talbi.

Un grand merci à lui pour m'avoir inspiré à approfondir l'optimisation et ses applications dans la réalité. Avec un camarade, il nous a donné l'opportunité de nous épanouir dans l'optimisation avec un projet de fin d'études puis des stages de fin de formation.

Emmanuel Kieffer.

Un grand merci à lui pour avoir été un tuteur bienveillant. Il m'a inspiré et soutenu à travailler ce sujet. Il n'a pas freiné ou contraint mes pensées, m'a laissé l'autonomie et la liberté dont j'ai besoin. J'ai pu équilibrer mon bien être mental et me passionner pour ce projet.