



**HAL**  
open science

# Approximation of stochastic models for epidemics on large multi-level graphs

Madeleine Kubasch

► **To cite this version:**

Madeleine Kubasch. Approximation of stochastic models for epidemics on large multi-level graphs. Probability [math.PR]. Institut Polytechnique de Paris, 2024. English. NNT : 2024IPPAX031 . tel-04717689

**HAL Id: tel-04717689**

**<https://theses.hal.science/tel-04717689v1>**

Submitted on 2 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT  
POLYTECHNIQUE  
DE PARIS

NNT : 2024IPPAX031

Thèse de doctorat



# Approximation of stochastic models for epidemics on large multi-level graphs

Thèse de doctorat de l'Institut Polytechnique de Paris  
préparée à École polytechnique

École doctorale n°574 Ecole Doctorale de Mathématiques Hadamard (EDMH)  
Spécialité de doctorat : Mathématiques appliquées

Thèse présentée et soutenue à Palaiseau, le 10 juillet 2024, par

**MADELEINE KUBASCH**

Composition du Jury :

Anne de Bouard Directrice de recherche, École polytechnique (CMAP)	Présidente
Tom Britton Professor, Stockholm University (Department of Mathematics)	Rapporteur
Laurent Massoulié Directeur de Recherche, INRIA (DYOGENE)	Rapporteur
Simon Cauchemez Directeur de recherche, Institut Pasteur (Unité de Modélisation Mathématique des Maladies Infectieuses)	Examineur
Clémentine Prieur Professeur, Université Grenoble Alpes (Laboratoire Jean Kuntzmann)	Examinatrice
Viet Chi Tran Professeur, Université Gustave Eiffel (LAMA)	Examineur
Vincent Bansaye Professeur, École polytechnique (CMAP)	Directeur de thèse



*À Eliza.*





# Remerciements

Tout au long de mon parcours, j'ai eu la chance de rencontrer des personnes formidables, qui m'ont inspirée et ont guidé ma trajectoire, tant professionnelle que personnelle. Ces quelques paragraphes sont l'occasion de les remercier, et de leur rendre hommage.

Malheureusement, la rédaction de ces remerciements est aussi un moment doux-amer, car ma thèse a été marquée par un évènement très douloureux: ma directrice de thèse, Elisabeta Vergu, est décédée en mai 2023. Eliza, merci d'avoir été là, jusqu'au bout. Je n'ai plus eu l'occasion de te dire que je me considère chanceuse d'avoir été ton étudiante. J'espère que ces mots te parviendront, où que tu sois.

Cette aventure commencée à trois, nous l'avons donc finie à deux. Vincent, je n'aurais pu imaginer de meilleur guide pour affronter les hauts et les bas de ces dernières années. Merci pour votre optimisme mathématique, mais surtout pour votre soutien infaillible et pour avoir la porte toujours ouverte. Vous avez su être là, à chaque instant. J'ai eu deux directeurs de thèse exceptionnels, et j'espère un jour pouvoir transmettre votre curiosité scientifique, votre énergie et votre bienveillance à de futurs étudiants.

Many thanks to Laurent Massoulié and Tom Britton, who have taken the time to be my thesis referees, as well as all the examiners: Simon Cauchemez, Anne de Bouard, Clémentine Prieur and Chi Tran.

Further, I am very grateful to Frank Ball, Bertrand Cloez, Chi Tran, Marie Doumic, Josselin Garnier, and Clémentine Prieur for taking the time to discuss my research over the years. I would also like to thank Simon Cauchemez for giving me the opportunity to spend a day at his lab. The scientific discussions I have had with you and your teams have been very inspiring, both for the present work and hopefully the years to come.

François, j'ai beaucoup aimé travailler avec toi. Merci pour tes conseils et le temps passé à répondre à mes questions, y compris après ton départ de MalAGE.

En terme d'évènements scientifiques inspirants, les rencontres de la chaire MMB et l'école d'été associée ont été parmi les plus belles opportunités de ma thèse. Je suis véritablement ravie de continuer à évoluer au sein de cette communauté dans les deux années qui suivent.

Je tiens à remercier aussi toute l'équipe administrative du CMAP, et en particulier Nasséra, Alex et Roberta, pour votre soutien et vos conseils au cours des trois dernières années.

C'est à MalAGE que je suis arrivée pour mon tout premier jour de stage, quelques jours avant mes débuts au CMAP. Il y a là au 210 un coin café très accueillant, où se multiplient les discussions: merci donc à tous les membres du 210 de rendre cet endroit aussi vivant et agréable.

Merci à mes grands frères et sœur de thèse Lina, Henri et Romain pour l'accueil sympathique que vous m'avez réservé à l'époque, les balades de Vauboyen, et les discussions philosophiques au coin café. Par la suite, nous avons continué de former une belle équipe de doctorants ! Marie, merci pour tes encouragements sincères et ta bonne humeur. Marion et Tom, nous l'avons commencée en même temps cette épopée, et pouvoir franchir les étapes

plus ou moins en même temps a été un soutien précieux. Pour la génération  $n + 1$ , constituée d'Andréa, Antoine, Katia et Arthur, merci pour votre énergie, les sorties au Massy Pal' et autres bons plans. Il reste les plus jeunes: Julie, Anastasia, et Eleonora, gardez cette dynamique, vous verrez, la thèse c'est une belle aventure ! Il y a eu aussi beaucoup de stagiaires, trop nombreux pour tous être cités, mais vous vous reconnaîtrez ! En particulier, les aventures bretonnes et normandes étaient vraiment de très belles escapades. Aussi, j'aimerais remercier nos post-doctorants Thibault, Auguste, Aurélien et Jeanne pour leur encouragements et les discussions sur l'après-thèse.

En bon accord avec les thématiques de recherche de l'INRAE, il y a également un petit potager: merci Alain d'avoir accompagné mes premiers pas entre haricots et pommes de terre, et pour un beau séjour au Bon Repos qu'on espère renouveler bientôt !

Enfin, je tiens à remercier Estelle: merci pour l'engagement sincère que tu as pour tous les doctorants du laboratoire.

Du côté du CMAP, je m'excuse d'avance: vous êtes tellement nombreux que je ne pourrai pas tous vous citer ! Tout d'abord, un petit mot pour les doctorants "maths-bio". Josué, Claire, Ignacio, Maxime, Céline, Ana, Jules, Anouar, Alexandre et Luce, je vous considère (plus ou moins à juste titre) tous comme mes frères et sœurs de thèse - nous avons beaucoup partagé, entre les rencontres de la chaire, Aussois, ou encore Saint-Flour. Merci pour votre formidable compagnie, la bonne humeur et le soutien mutuel qui règne. Ignacio, on s'est particulièrement épaulés ces dernières années: j'ai passé de très beaux moments en ta compagnie, et j'ai hâte que nos chemins se croisent à nouveau.

Plus généralement, le coin café du CMAP n'a rien à envier à celui de MalAGE, et j'y ai passé de très beaux moments avec nombreux d'entre vous: Manon, Guillaume, Antoine VB, Armand, Richard, Grégoire, Leila, Christoph, Vanessa, Antoine, Maxime, Margaux, Clément, Solange, Angèle, Laetitia... Une pensée particulière aux explorateurs de l'Île Saint-Oléron, occasion de rencontre entre doctorants des deux laboratoires. Je tiens également à mentionner séparément Constantin et Jessie, qui ont été mes compagnons en tant que représentants des doctorants: on a formé de beaux duos !

Merci à toute l'équipe du mentorat Femmes & Sciences pour l'énergie que vous investissez dans l'organisation de tables rondes, formations, retours d'expériences. Merci à toi Sylvie pour toutes nos discussions, qui m'ont apporté bien plus que tu ne le pensais.

Avant la thèse, j'ai eu la chance déjà d'avoir rencontré des personnes qui m'ont encouragée et qui m'ont offert leur soutien sincère. Cela concerne en particulier mes anciens encadrants de stage. Julia, c'est dans ton équipe que j'ai eu ma première expérience de recherche, et c'est un monde excitant et merveilleux qui s'est ouvert à moi. Dylan, you have been a wonderful mentor, when I really needed one. Enfin, mon stage avec Renaud, Simon, Luis, Ayman et Nicolas chez Greenshield a été l'occasion de découvrir la modélisation et l'épidémiologie.

Benoit et Stéphane, vous avez été là lors du passage de la comète: merci, du fond du cœur.

Toutes ces aventures, je les ai partagées avec de formidables amis. Il y a eu de belles après-midi entre jeux de société et expositions, en compagnie souvent des Piwouis et Pimousses: Anna, Léo, Arthur, Louise, Tanguy, Gaël, Clément, Iris, Alan... Les soirées chaleureuses avec Abel et Pauline (meilleure conséquence du cours d'Analyse fonctionnelle qu'on aurait pu

imaginer !). Ishak, c'est toujours un plaisir de te croiser. Les nombreuses et diverses sorties avec mes *nakama*, la fluffesque Sophie, Anaïs, Stéphanie, Jessie et Camille, notre C-Rex. Les soirées films avec Marine et Edouard, qui finissent souvent en papotage car on a trop de choses à se raconter. Et bien sûr, Hermine et Cyrus: bien que loin, vous êtes toujours près du cœur.

Merci également aux Gourmet, et en particulier à Michèle et Michel, pour nous avoir véritablement "adoptés" à notre venue en France.

Depuis le début, j'ai pu compter sur le soutien infailible et l'amour de ma famille. Vous m'avez aidée à surmonter toutes les épreuves qui ont pu se présenter en cours de chemin. Si ces 9 dernières années (et plus !) ont été une belle aventure, c'est en grande partie grâce à vous. Cela vaut aussi pour toi, Juliette: j'ai déjà hâte de te voir en août. Merci de toujours être à mes côtés, je suis fière de faire partie des Fous ! Hanno, danke für all die Kerzen, die du über die Jahre für mich hast brennen lassen - sie haben mir immer Glück gebracht. Ich weiss, Hans-Jürgen würde sich so über alles freuen. Er ist im Herzen immer bei uns.

Merci également à ceux qui de loin m'ont envoyé beaucoup d'encouragements ces dernières années: mon homonyme Madeleine, Louis, Diane, Judith, Aristide, l'oncle Émile et Virginie, Sonya et Ayité. Il y a certains d'entre vous que je n'ai pas encore pu rencontrer, sachez que j'ai vraiment hâte de venir vous voir !

Enfin, Henri. Tu as été la plus belle des rencontres. Tu sais déjà tout ce que j'ai envie de te dire, tout ce que tu représentes pour moi. Je suis et je serai là, à tes côtés. Underneath the mango tree...



# Contents

<b>1</b>	<b>Introduction</b>	<b>13</b>
1.1	Contact heterogeneities in epidemic models	13
1.1.1	Some classical epidemic models	14
1.1.2	Modeling contacts among individuals	14
	Uniform mixing	15
	Contact matrixes	16
	Network models	18
	Multilayer contact models	25
1.2	Reduced models in epidemiology	30
1.2.1	A motivating example: the uniformly mixing <i>SIR</i> model	30
1.2.2	Some heuristics for reduction of epidemics on networks	32
	Pairwise models	32
	Effective degree models	33
	Edge-based compartmental models (EBCM)	34
1.2.3	An overview of reduction results for epidemics on random graphs	36
1.3	Main contributions of the thesis	39
1.3.1	Chapter 2 - The epidemic footprint of contact structures	41
1.3.2	Chapter 3 - Large population limit for a multilayer <i>SIR</i> model including households and workplaces	42
1.3.3	Chapter 4 - Sensitivity of the reduction accuracy to network and epidemic parameters	46
1.3.4	Chapter 5 - On spinal constructions for interacting populations	48
1.4	Contributions principales de la thèse	52
1.4.1	Chapitre 2 - L'empreinte épidémique des structures de contact	52
1.4.2	Chapitre 3 - Limite grande population d'un modèle <i>SIR</i> multicouche incluant foyers et lieux de travail	54
1.4.3	Chapitre 4 - Étude de la sensibilité de la précision de réduction aux paramètres du réseau et de l'épidémie	58
1.4.4	Chapitre 5 - Sur les constructions spinales pour populations en interaction	60
<b>2</b>	<b>The epidemic footprint of contact structures</b>	<b>65</b>
2.1	Introduction	66
2.2	Model with two levels of mixing: description, simulation approach, key parameters, simulation scenarios	70
2.2.1	General model description	70
2.2.2	Structure size distributions	70
2.2.3	Numerical simulation scenarios: structure size distributions and epidemiological parameters	71
2.2.4	Simulations of the population structure and epidemic process	73
2.2.5	Simulation of teleworking strategies	74

2.3	The impact of the size distribution of closed structures and assessment of teleworking strategies . . . . .	75
2.3.1	Outbreak criterion, $R_I$ and type of infection . . . . .	75
2.3.2	The effect of structure size distribution on epidemic outcomes . . . . .	77
2.3.3	Teleworking strategies . . . . .	79
2.3.4	Robustness to the form of the infection term . . . . .	82
2.4	Reduction to compartmental ODEs based on the initial growth rate . . . . .	83
2.4.1	Laplace transform of the infection rate in a uniformly mixing population . . . . .	85
2.4.2	Characterization of the initial growth rate . . . . .	88
2.4.3	ODE reduction of the multilevel model based on the initial growth rate . . . . .	90
2.4.4	Assessment of the reduction robustness . . . . .	91
2.5	Discussion . . . . .	92
	Appendix . . . . .	98
2.A	Generation of structure size distributions . . . . .	98
2.B	Parameter values for scenarios . . . . .	98
2.C	Numerical computations of epidemic parameters and outcomes . . . . .	99
2.D	Proof of Proposition 2.4.1 . . . . .	100
2.E	Computation of the exponential growth rate for the <i>SEIR</i> model with two levels of mixing . . . . .	103
2.F	Numerical aspects for the model reductions of Section 2.4.4 . . . . .	104
<b>3</b>	<b>Large population limit for a multilayer <i>SIR</i> model including households and workplaces</b> . . . . .	<b>107</b>
3.1	Introduction . . . . .	108
3.2	Presentation of the model . . . . .	111
3.2.1	General presentation of the model . . . . .	111
3.2.2	The epidemic model at the level of households and workplaces . . . . .	112
3.3	Main results . . . . .	117
3.3.1	Large population approximation of $(\zeta^K)_{K \geq 1}$ . . . . .	117
3.3.2	Numerical assessment of the limiting dynamical system . . . . .	120
	Implementation of the dynamical system and illustration of Theorem 3.3.3 . . . . .	120
	Comparison to edge-based compartmental models . . . . .	122
3.4	Proofs . . . . .	123
3.4.1	Proof of Theorem 3.3.2 . . . . .	123
	Uniqueness and continuity of the solution of Equation (3.6). . . . .	123
	Tightness of $(\zeta^K)_{K \geq 1}$ in $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), w))^2$ . . . . .	126
	Identification of the limiting values of $(\zeta^K)_{K \geq 1}$ . . . . .	132
	Proof of Theorem 3.3.2 . . . . .	137
3.4.2	Proof of Theorem 3.3.3 . . . . .	138
	Preliminary study of the dynamical system . . . . .	138
	Some properties of the limiting measure $\eta$ . . . . .	139
	Proof of Theorem 3.3.3 . . . . .	141
	Discussion . . . . .	147
	Appendix . . . . .	148
3.A	Absolute continuity of $\eta$ and related PDE system . . . . .	148

3.B	Implementation of the large population limit . . . . .	154
3.B.1	Automatic implementation of the dynamical system . . . . .	154
3.B.2	Computational performance . . . . .	154
3.C	Edge-based compartmental model . . . . .	157
3.C.1	Presentation of the EBCM . . . . .	157
3.C.2	Computational performance . . . . .	160
3.D	Proof of Proposition 3.4.10 . . . . .	161
<b>4</b>	<b>Sensitivity of the reduction accuracy to network and epidemic parameters</b>	<b>165</b>
4.1	Introduction . . . . .	165
4.2	Sensitivity analysis of the precision of the uniformly mixing reduced model . . . . .	167
4.2.1	General approach and quantities of interest . . . . .	167
4.2.2	Experimental design . . . . .	168
4.2.3	Results . . . . .	169
4.3	The epidemic impact of network model perturbations . . . . .	170
4.3.1	A generalization of the household-workplace model . . . . .	170
4.3.2	Simulation study design . . . . .	172
	Network parameters . . . . .	172
	Epidemic scenarios for numerical explorations of the generalized household-workplace model . . . . .	173
4.3.3	Results . . . . .	173
4.4	Discussion . . . . .	176
<b>5</b>	<b>On spinal constructions for interacting populations</b>	<b>179</b>
5.1	Introduction . . . . .	179
5.2	The population process and its $\psi$ -spine construction . . . . .	181
5.2.1	The population process . . . . .	181
5.2.2	Reminders on the $\psi$ -spine construction for interacting populations . . . . .	182
	The $\psi$ -spine for finite-size populations . . . . .	183
	Large population approximation of the $\psi$ -spine . . . . .	184
5.3	The $\psi$ -sampled lineage . . . . .	185
5.3.1	A many-to-one formula . . . . .	185
5.3.2	Proofs . . . . .	187
	Existence and uniqueness of the $\psi$ -auxiliary process . . . . .	187
	Proof of the many-to-one formula . . . . .	191
5.4	Quantification of the $\psi$ -spine large population approximation . . . . .	194
5.4.1	Coupling the $\psi$ -spine with its large population limit . . . . .	195
5.4.2	Quantifying the approximation error . . . . .	196
5.5	Discussion . . . . .	198
	<b>Bibliography</b>	<b>201</b>





# Introduction

---

## Contents

---

<b>1.1</b>	<b>Contact heterogeneities in epidemic models</b>	<b>13</b>
1.1.1	Some classical epidemic models	14
1.1.2	Modeling contacts among individuals	14
<b>1.2</b>	<b>Reduced models in epidemiology</b>	<b>30</b>
1.2.1	A motivating example: the uniformly mixing <i>SIR</i> model	30
1.2.2	Some heuristics for reduction of epidemics on networks	32
1.2.3	An overview of reduction results for epidemics on random graphs	36
<b>1.3</b>	<b>Main contributions of the thesis</b>	<b>39</b>
1.3.1	Chapter 2 - The epidemic footprint of contact structures	41
1.3.2	Chapter 3 - Large population limit for a multilayer <i>SIR</i> model including households and workplaces	42
1.3.3	Chapter 4 - Sensitivity of the reduction accuracy to network and epidemic parameters	46
1.3.4	Chapter 5 - On spinal constructions for interacting populations	48
<b>1.4</b>	<b>Contributions principales de la thèse</b>	<b>52</b>
1.4.1	Chapitre 2 - L’empreinte épidémique des structures de contact	52
1.4.2	Chapitre 3 - Limite grande population d’un modèle <i>SIR</i> multi-couche incluant foyers et lieux de travail	54
1.4.3	Chapitre 4 - Étude de la sensibilité de la précision de réduction aux paramètres du réseau et de l’épidémie	58
1.4.4	Chapitre 5 - Sur les constructions spinales pour populations en interaction	60

---

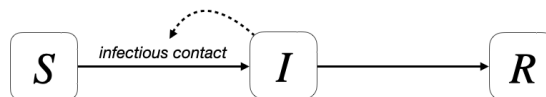
## 1.1 Contact heterogeneities in epidemic models

The spread of an epidemic naturally is influenced by the way individuals encounter one another, and as a consequence, mathematical models in epidemiology try to take into account this organization of contacts. Depending on the disease under consideration, different aspects of contact heterogeneities may be relevant, leading to more-or-less fine-grained descriptions of connections between individuals. In this section, we will first give a brief, general description of some classical epidemic models, such as the *SIR* model. Second, we will see how these models incorporate information on contacts within the population and discuss the pros and cons of each modeling choice.

### 1.1.1 Some classical epidemic models

Many epidemic models rely on the concept that one may summarize the possible evolution of any individual in the population using, basically, a flow chart. The general idea is that each individual belongs to one, and only one, sanitary state among a given set of possible such states. The individual will remain in that state for some (random) amount of time, before switching to another state, according to predefined transition rules which may depend on characteristics of the individual himself, or of other members of the population.

Maybe the most famous example is the *SIR* model, introduced in the seminal paper by Kermack and McKendrick, 1997. Here, each individual can either be *susceptible* (S), if he has never encountered the disease and may be contaminated in the future; *infected* (I); or *recovered* (R), in which case the individual is supposed to be immune against infection. Transitions among those states occur as follows. A susceptible gets infected at an *infectious contact*, *i.e.* an encounter with an infected individual. He subsequently stays infected for a period of random length, which is usually supposed to be distributed according to a given probability  $\nu$  on  $\mathbb{R}_+$ , independently for each individual. Once this *infectious period* is over, the individual recovers and remains in state R. Figure 1.1 synthesizes this model description.



**Figure 1.1:** The SIR compartmental model.

In particular, this approach has the advantage of being very flexible. Indeed, one may take into account a larger number of states. For instance, there may be an *incubation period* during which an individual is already contaminated but not yet infectious, which can be represented by an additional *exposed* (E) state: this corresponds to the *SEIR* model. Similarly, in order to model waning immunity after recovery, one may allow recovered individuals to return to the susceptible state, amounting to the *SIRS* model, and so on. This procedure can lead to quite sophisticated models, as illustrated for instance by models of the recent COVID-19 epidemic (Romano et al., 2020; Roux et al., 2023). Also, it is possible to incorporate more details which further influence the evolution of the sanitary state of an individual. For instance, one may want to consider age structured models, as an individual's age may influence the risk of being infected or developing severe forms of a disease upon encountering an infected (*e.g.* Davies et al., 2020 in the context of COVID-19). We refer to Chapter 2 of Brauer et al., 2008 for a more extensive overview.

Among the many possible refinements, one possibility consists in focusing on how infectious contacts occur, *i.e.* how often, and under which conditions, susceptible and infected individuals encounter one another. Let us introduce some major modeling choices for this step.

### 1.1.2 Modeling contacts among individuals

Incorporating information on how contacts are organized within a population is a challenging task, as there are many aspects which interplay to yield contact patterns. For instance, the frequency at which two individuals meet may depend simultaneously on:

- where they live (spatial structure);
- which social groups they identify themselves to, such as age class or sexual orientation;
- the context in which they may get together: at home, at work, at the grocery store...;
- the time period considered: there may be fluctuations depending on the individuals' weekly schedule, changes of habits during holidays, and so.

As a consequence, a natural question arising is: how much of this information should one keep when constructing an epidemic model? A major part of the answer is that it depends on the issues the model tries to address. Thus, several ways of taking into account different parts of the population's contact structure have been developed. Here, we will present some of the most common model choices, for each of which we will illustrate both their pertinence and their limitations.

### Uniform mixing

Let us start with the simplest form of contact structure within a population, namely the one where no contact heterogeneities are considered: a population is said to be *uniformly mixing* if each pair of individuals meet at the same frequency. For example, in the case of the *SIR* model, this translates into each susceptible being contaminated at rate  $\beta I$ , where  $\beta$  is the contact rate and  $I$  the current number of infected within the population.

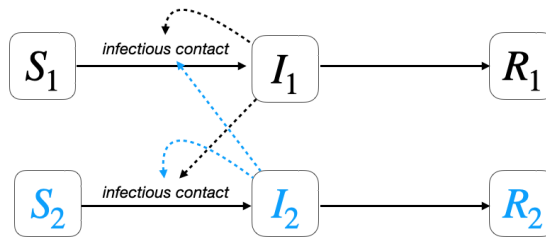
Despite its apparent simplicity, uniform mixing actually is able to capture some main features of real-life epidemics, especially if considering refinements of the *SIR* model, for example taking into account demographic effects or seasonal variability in infectivity (e.g. Brauer et al., 2008, Chapter 1 and references therein). In addition, it has the advantage of being well studied: many results have first been established in this setting, yielding a detailed understanding of the epidemic. As a consequence, uniform mixing can serve as a toy case, where many questions may be addressed more easily than when taking into account contact heterogeneities. In particular, it serves as a good basis to explore the effect of factors which are not related directly to contact structure, such as varying susceptibility (Hyman and Li, 2005), and whose impact one may want to study without addressing simultaneously contact heterogeneities.

Nevertheless, it is clear that uniform mixing is an oversimplified representation of human contacts, which is very far from real-life settings. On the one hand, for some diseases, assuming uniform mixing may almost be absurd given the way contaminations occur. For example, in the case of sexually transmitted diseases, whom makes contact with whom may depend on the age and gender of the partners involved (Tomori et al., 2022). On the other hand, even when the means of contamination by itself does not introduce specific contact patterns (e.g. air-borne diseases), it is well known that some crucial epidemic aspects are not well represented through uniform mixing. Consider for instance the *herd immunity threshold*, which is defined as the proportion of the population which needs to be immune against the disease in order to prevent an epidemic outbreak. Notably, estimating the herd immunity threshold is of practical interest, as illustrated in the recent COVID-19 epidemic (Fontanet and Cauchemez, 2020). However, considering uniform mixing yields an overestimation of this threshold, compared to more realistic settings where contact heterogeneities are taken into account. Intuitively, this is due to the fact that in the uniformly mixing population,

all individuals exert the same infectious pressure once they are infected. However, when the number of contacts is not identical for all individuals, those having the most social interactions contribute the most to the epidemic spread. As a consequence, herd immunity may be achieved at a lower threshold by preferentially targeting those individuals (Britton et al., 2020). Taken together, these arguments illustrate the need to go beyond uniform mixing.

### Contact matrixes

In order to take into account contact heterogeneities, a first step is to consider refinements of the uniformly mixing point of view. More precisely, it is possible to distinguish different groups of individuals within the population (Figure 1.2). In that case, the frequency at which two individuals encounter one another may depend on the groups to which they belong. This leads to the definition of a *contact matrix*  $C$ , which is such that infected of type  $i$  contaminate susceptibles of type  $j$  at rate  $\beta C_{i,j}$ . In words, the coefficient  $C_{i,j}$  represents the average rate at which individuals of type  $i$  meet individuals of type  $j$ , and  $\beta$  represents the infectivity of the disease under consideration. Notice that  $C$  is generally supposed to be symmetric. Of course, other parameters such as the distribution of infectious periods may also be group-specific.



**Figure 1.2:** Schematic representation of an SIR compartmental model distinguishing two groups of individuals.

This modeling choice has several advantages. First, it may serve as a proxy for more complex contact heterogeneities, while being easily interpretable as the choice of groups usually is informed from a phenomenological point of view. A classical example is to consider age classes, as children may have frequent contacts among each other in schools compared to contacts with elderly people, and so on (Ram and Schaposnik, 2021). Similarly, in the case of sexually transmitted diseases such as gonorrhea (Lajmanovich and Yorke, 1976) or HIV (Jacquez et al., 1988), one may want to distinguish how sexually active individuals are, or their gender as the disease may evolve differently in males or females. Finally, these models are also well suited to model transmission of diseases among different species. This is of particular interest for zoonotic diseases, such as monkeypox (Peter et al., 2022) or bovine tuberculosis (Brooks-Pollock and Wood, 2015).

Additionally, it is possible to estimate contact matrixes using large-scale surveys. This allows to gain a deeper insight into how contacts occur within a given population, and the obtained contact matrixes can be used to parametrize epidemic models. To our knowledge, the first such survey occurred during the POLYMOD study in 2008 (Mossong et al., 2008), which was motivated especially by airborne diseases such as influenza and SARS. In this



**Figure 1.3:** Contact matrixes computed from the UK POLYMOD data, grouped by age class. All contexts (A), workplace contacts only (B) or household contacts only (C).

study, over 7000 participants from 8 European countries kept track of their daily contacts, including contact length, age and gender of the person they would interact with, *et cetera*. Subsequently, age-structured contact matrixes for each of the participating countries can be computed from this data. More recently, the COVID-19 epidemic has incited a similar study, called CoMix (Wong et al., 2023), which has allowed to infer contact matrixes from 21 European countries as well as their evolution over time, covering several months up to two years. Further, in both studies, contacts were characterized by social circumstances as well, such as contacts within households, workplaces or schools (Fig. 1.3). The collected data thus allow to have a rather detailed description of contact patterns in Europe.

Nevertheless, this representation of contacts still assumes a well-mixing population: if the contact rates between two groups are non zero, then each individual of the first group encounters any individual of the second group at the same frequency. Hence, contact matrixes remain insufficient to capture some characteristics of human contacts, such as the fact that people usually have a limited number of close contacts. In addition, it further follows that this organization of contacts can directly be represented as a finite dimensional compartmental model.

Notably, both uniform mixing and contact matrix models correspond to *global* mixing models. These represent populations in which each individual has an infinite number of possible contacts, leading each contact event to involve a different partner. Not only is this assumption unrealistic, it also intuitively is expected to have significant impact on epidemic spread: as each contact involves a different partner, infected individuals have a stronger infection potential than if their number of partners is limited. Indeed, in the latter case, infected individuals make repeated contacts with a given partner. Thus, once this partner is infected, all subsequent contacts with him cannot lead to an infection event. This setting corresponds to *local* mixing, and can be modeled through the use of contact networks.

## Network models

Arguably the most appropriate representation of contacts consists in a contact network. The general idea is that such a network is composed of nodes, also referred to as vertices, which represent individuals. Two given nodes are said to be neighbors if they are connected by an edge, which means that the corresponding individuals may get in contact with one another.

Let us take the example of an *SIR* model to illustrate how an epidemic unfolds on the network. As previously, each node is either in the susceptible, infected or recovered state. During its infectious period, an infected can contaminate only its susceptible neighbors, *i.e.* the epidemic spreads exclusively along the edges. Once the infectious period is over, the node recovers and is henceforth immune to the disease. As a consequence, contact networks are a powerful modeling tool, as they can be extremely precise by depicting all possible contacts of each individual within the population.

Before going into more detail on network models in epidemiology, let us introduce some general notions related to networks and refer to Newman, 2003 and Durrett, 2006 for a more complete overview. First, throughout the thesis, the words *network* and *graph* will be used interchangeably. As mentioned above, a graph corresponds to a couple  $\{V, E\}$  of vertices  $V$ , which is a finite or countably infinite set, and edges  $E$ , where for any  $v, w \in V$ , the edge  $(v, w)$  belongs to  $E$  if and only if  $v$  and  $w$  are neighbors.

This set of information can equivalently be encoded using an *adjacency matrix*  $A$  which is defined as follows. Suppose that  $V$  is finite, and consider a bijection between  $V$  and  $\llbracket 1, \text{Card}(V) \rrbracket$ . With slight abuse of notations, we thus let  $V = \llbracket 1, \text{Card}(V) \rrbracket$ . Then  $A$  is a square matrix of dimension  $\text{Card}(V)^2$ , such that  $A_{i,j} = \mathbf{1}_{\{(i,j) \in E\}}$ . This construction extends to the case where  $V$  is in bijection with  $\mathbb{N}$  in a straight-forward manner. Notice that one distinguishes directed and undirected graphs. In the former case, the edges  $(i, j)$  and  $(j, i)$  are considered separately: an edge may lead from  $i$  to  $j$ , without the possibility of going directly from  $j$  to  $i$ , which implies that the contact matrix  $A$  is not necessarily symmetric. For undirected graphs, however,  $A$  is symmetric as  $i$  is linked to  $j$  if and only if  $j$  is linked to  $i$ . Finally, a network may be weighted, in which case each edge  $(i, j)$  is characterized by its weight  $W_{i,j} \in \mathbb{R}$ , and one usually sets  $A_{i,j} = W_{i,j}$ .

There are several quantities which are of interest when trying to describe the network structure. Consider a graph  $G$ . First, one may want to if all nodes have roughly the same number of neighbors, or whether there is a lot of heterogeneity. The number of neighbors of a node  $v$  is also called its degree  $d(v)$ , and this information is captured by the *degree distribution*  $\mu_G$  defined as follows:

$$\mu_G = \sum_{v \in V} \delta_{d(v)}.$$

In particular, if the variance of  $\mu_G$  is low, all nodes have approximately the same number of contacts. A particular example of this case are  $d$ -regular graphs, where each node has exactly  $d$  neighbors, leading to  $\mu_G = \mathbf{1}_{\{d\}}$ . Conversely, if the variance of the degree distribution is high, some nodes have significantly more contacts than others: in this case, there hence are contact heterogeneities.

Another question would be whether two neighbors of a node are likely to be connected between themselves, or in other words, if "the friend of my friend is likely to be my friend"? This leads to the definition of two *clustering coefficients*. The first one corresponds to

computing the proportion of triangles among connected triples, which are sets of vertices  $(u, v, w)$  such that at least two out of the three edges  $(u, v)$ ,  $(u, w)$  and  $(v, w)$  exist (Barrat and Weigt, 2000):

$$C_1 = \frac{3 \times (\text{number of triangles})}{\text{number of connected triples}}. \quad (1.1)$$

Alternatively, it is possible to compute this coefficient locally, meaning that for each node  $v$ ,  $C_1(v)$  corresponds to the ratio of the number of triangles including  $v$  divided by the number of connected triples including  $v$ . The clustering coefficient  $C_2$  is obtained by taking the average local clustering coefficient over all nodes (Watts and Strogatz, 1998):

$$C_2 = \frac{1}{\text{Card}(V)} \sum_{v \in V} C_1(v).$$

The difference between  $C_1$  and  $C_2$  stems from the fact that  $C_1$  takes the ratio of the average number of triangles per node over the average number of triples per node, while  $C_2$  consists in the average of the ratio for each node. As emphasized in Barrat and Weigt, 2000 and Newman, 2003, this leads to both clustering coefficients having a qualitatively similar behavior, while quantitatively their values can differ.

Last but not least, one may also wonder how quickly one may reach a node  $w$  starting from another node  $v$ , following the edges of the graph. First of all, that may not be possible for all pairs of nodes. If such a path going from  $v$  to  $w$  does exist for all pairs  $(v, w)$  of nodes, then the graph  $G$  is said to be *connected*. Otherwise, the set of nodes which can be reached starting from a given node  $v$  is referred to as  $v$ 's component, and the largest component of a graph is sometimes called its "giant component". Notice that for directed graphs, one usually distinguishes an out-component (nodes which can be reached from  $v$ ) and an in-component (nodes from which  $v$  may be attained).

Hence, consider an undirected graph  $G = (V, E)$ . Let  $v \in V$  and consider a node  $w$  belonging to its component, in which case we write  $v \leftrightarrow w$ . Then the shortest path connecting  $v$  to  $w$  is called their geodesic path, and let  $\ell_{vw}$  be its length, *i.e.* the number of edges involved in the path, or the sum of their weights for weighted networks.  $\ell_{vw}$  is also referred to as the geodesic distance between nodes  $v$  and  $w$ . Notice that, by convention,  $v \leftrightarrow v$  and  $\ell_{vv} = 0$ . The average geodesic distance  $\ell$  of the graph then is defined as:

$$\ell = \frac{2}{n(n+1)} \sum_{(v,w):v \leftrightarrow w} \ell_{vw}.$$

Analogously, the *diameter* of the graph is defined as the maximum distance separating two connected nodes:

$$\text{diam}(G) = \max\{\ell_{(v,w)} : v \leftrightarrow w\}.$$

Both the average geodesic distance and diameter thus quantify how quickly one may move between two connected nodes of the graph.

A natural interrogation which arises is: what are the characteristics of real-life networks involved in epidemic dynamics? Many real-life networks, including examples of social networks, are known to present some common properties (Newman, 2003). Indeed, they usually are *small world*, meaning that their diameter is (very) small. Also, they are clustered, as one would expect for small world networks. Finally, they often are *scale-free*, meaning that the



degree distribution follows a power-law of parameter  $\alpha \in (2, 3)$ : the probability  $p_k$  of having  $k$  neighbors is equivalent to  $Ck^{-\alpha}$  for large values of  $k$ ,  $C$  being a normalizing constant. In this case, the degree distribution admits a finite average, whereas its second moment is infinite, as the degree distribution is very heterogeneous: most nodes have few neighbors, while few nodes have many neighbors.

In our case, it is of particular interest to have a look at empirical studies which have tried to assess social contact networks which are relevant for disease spread. Several types of data may be used. A first source of data stems from contact tracing, which can serve as a means of epidemic control: individuals who are diagnosed with the disease are asked to indicate their recent contacts, who in turn may be tested and report their contacts if they are contaminated. While this data is a partial observation of the complete network (only contacts involving at least one infected are reported), it still allows to have a look at the contact graph underlying the spread of the epidemic at hand. Figure 1.4a corresponds to a sexual contact network which has been obtained this way, in the context of a contact tracing strategy adopted to face the HIV/AIDS epidemic on Cuba between 1986 and 2006 (Cléménçon et al., 2015).

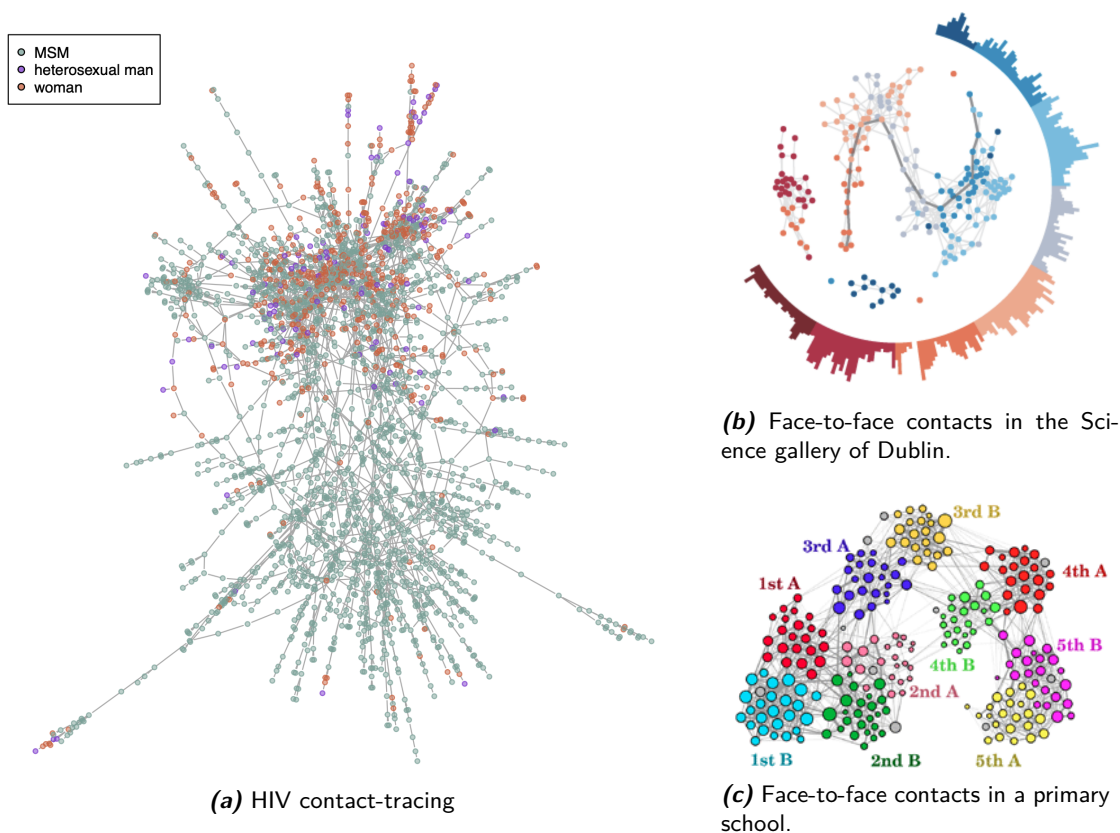
Another data source comes from experiments which explore face-to-face contacts, typically by using sensors which automatically register whom has a close contact with whom, and how long each contact lasts. This is particularly suiting for airborne diseases, such as influenza or COVID-19, which could be transmitted in these circumstances, even without having a physical contact (touching, hugging...). Figures 1.4b and 1.4c are example of the obtained networks in two different settings, namely at a museum (Isella et al., 2011) and primary school (Stehlé et al., 2011), respectively.

Taken together, these networks illustrate the aforementioned properties of real-life networks. Indeed, in all cases, the degree distribution is heterogeneous, with a scale-free and small world setting established for the sexual network in Figure 1.4a (Cléménçon et al., 2015). Similarly, all three of them present clustering, the most striking example being maybe Figure 1.4c: the authors note that within classes, there appears to almost be uniform mixing (Stehlé et al., 2011). Of course, all of these properties may not always be satisfied: for instance, the museum contact network of Figure 1.4b does not exhibit the small world behavior (Isella et al., 2011).

Finally, let us emphasize here that real-life contact networks actually exhibit further structure. On the one hand, they can be *dynamic*: this means that the edges are not constant over time, but may be removed or added. This can be observed on several time scales: for example, in the school setting of Stehlé et al., 2011, students interact differently with one another during lessons (contacts mainly within each class) or during breaks (mixing among classes). The network presented in Figure 1.4c actually corresponds to the cumulative network, where all contacts are accumulated over a day, while a video of the dynamical network is available on the [SocioPatterns](#) website. Another example of dynamics which arise from an epidemic context is that individuals may modify their contacts depending on their own health status (Van Kerckhove et al., 2013), or on their current risk perception. On the other hand, networks can also be spatially organized. Again, several scales are concerned: for instance, people may commute within their home town on a daily basis (Aguilar et al., 2022), or travel over long distances, participating in the global spread of diseases (Riquelme et al., 2021).

Overall, designing and carrying out studies which aim at measuring real-life contact

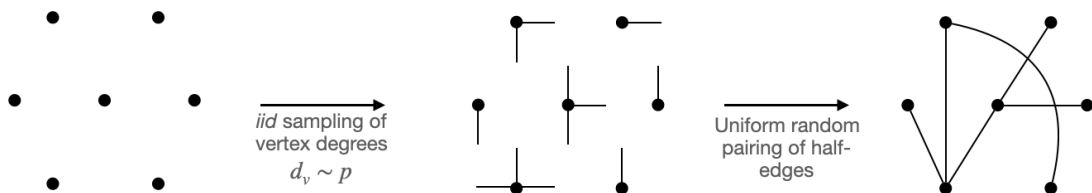
networks is not an easy task, as emphasized in Eames et al., 2015. Difficulties range from questions as to what defines a contact, over the difficulties arising from data collection such as unreported links in surveys, to interrogations about the possibilities of deducing conclusions about large contact networks from small-scale studies. As a consequence, the authors conclude that mathematical models should be used to point out which information about a contact network is the most important to assess, if one wants to predict the epidemic spread on this graph. Interestingly, Kiss et al., 2024 go one step further and explore the question whether it is possible to use epidemic data in order to infer information on the underlying contact network.



**Figure 1.4:** Examples of real-life networks which are of interest in an epidemic setting. (a) Largest connected component of the Cuban sexual contact network, obtained from HIV contact-tracing data. Nodes are colored by gender and/or sexual orientation, as indicated in the legend. This figure is reproduced from Cl  men  on et al., 2015. Courtesy of the authors. (b) Aggregated face-to-face contact network inferred from sensor data, during one day of the experiment Infectious Sociopatterns of the Dublin Science gallery (Isella et al., 2011). Nodes are colored by the time at which individuals entered the museum, in the order shown by the contour. The barplot on the contour further indicates the number of contacts over each 2-minute interval. Finally, the diameter of the graph is emphasized by the bold line. Available on [SocioPatterns](#). (c) Aggregated network of face-to-face contacts in a primary school, as measured by sensor data (Steh  l   et al., 2011). Colors correspond to different classes, as indicated on the figure. Available on [SocioPatterns](#).

Given the complexity of real-life networks and the challenges in measuring them, a lot of attention has been drawn to studying epidemic spread on mathematical models of random graphs (Danon et al., 2011; Kiss et al., 2017). In this paragraph, we will introduce some of those models, as well as their properties in the light of what we know about empirical contact networks. Notice that many more random network models have been developed, which will not be covered below, and refer the interested reader to Newman, 2003; Durrett, 2006; Hofstad, 2016 for more detail.

Historically, the first random graph model which has been extensively studied has been introduced in Solomonoff and Rapoport, 1951 and Erdős and Rényi, 1959. The general idea goes as follows. A so-called Erdős-Rényi graph  $G(n, p)$  is a graph with  $n$  vertices, where each pair of vertices is connected by an edge with probability  $p$ . Hence, by construction, the degree distribution at fixed population size  $n$  is given by a binomial law  $\mathcal{B}(n-1, p)$ . For this model, many explicit results have been established, especially in the large graph limit  $n \rightarrow \infty$ . Indeed, considering a large population is of interest in many applications, including epidemics as the populations in which diseases spread are often very large. For the Erdős-Rényi model, a common setting is to study  $G(n, p/n)$  as  $n$  goes to infinity, which corresponds to keeping the average node degree fixed. In this case, the degree distribution converges in probability to a Poisson law of parameter  $p$ , and the associated graph is referred to as Poisson graph. Further, the graphs become *locally tree-like*, which means that the neighborhood of a vertex typically resembles a tree: there are no small circles, implying in particular that clustering is low (Hofstad, 2024, Section 2.4.5). Other known results include the probability of the graph being connected and the size distribution of its connected components. In particular, there is a phase-transition depending on  $p$  where either there are relatively small components only, or where there exists a giant connected component, meaning that a positive fraction of the nodes is connected. We refer to Chapter IV.A of Newman, 2003, Chapter 2 of Durrett, 2006 and references therein for an extensive overview.



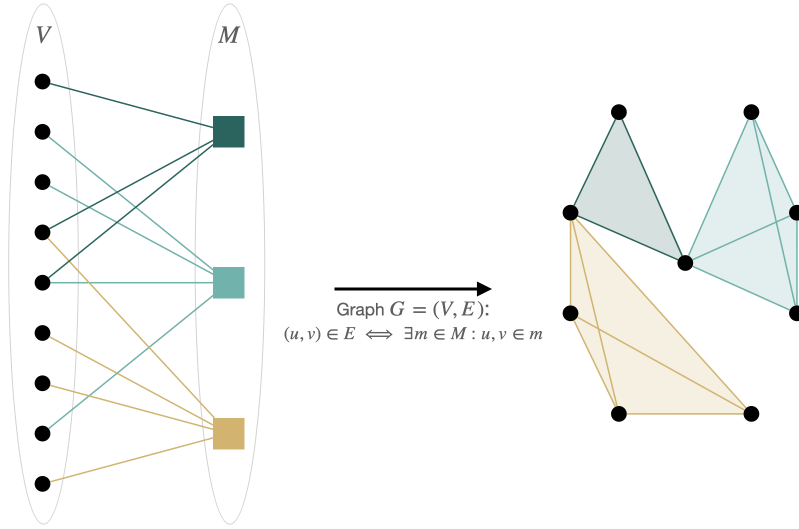
**Figure 1.5:** Main steps for obtaining one realization of the configuration model with target degree distribution  $p$ .

However, Erdős-Rényi graphs have the drawback of not being very flexible: for instance, as mentioned above, the degree distribution is necessarily binomial or asymptotically Poisson, which further does not correspond to a realistic setting. This raises interest in a more general random graph model. Here, we will focus on the configuration model (CM), which was first developed by Bollobás, 1980 with the aim of constructing random graphs with a given degree for each node. A common choice of the degree sequence is to consider *i.i.d* realizations of a target distribution  $p = (p_k)_{k \geq 1}$ , in which case a realization  $G = (V, E)$  of the configuration model is obtained as follows (Figure 1.5). For each node  $v \in V$ , draw its degree  $d_v$  from  $p$  independently from all other nodes, and attach  $d_v$  half-edges to node  $v$ . The edges of the

graph are then obtained by successively choosing uniformly at random, with removal, pairs of half-edges which are assembled to form an edge. This procedure stops as soon as either there are no-more half-edges left, or if exactly one half-edge remains which is removed as no edge can be formed. Notice that in principle, multi-edges and self-loops occur here. In the large graph limit, however, they become negligible, with precise control on the number of occurrences of those phenomena under mainly a second moment condition on  $p$  (Angel et al., 2019). In particular, the obtained degree distribution converges to the target distribution  $p$ . Notably, conditioned on being simple (no self-loops or multi-edges), each realization  $G = (V, E)$  of the configuration model corresponds to a uniform choice among all possible graphs having the same degree sequence  $(d_v : v \in V)$ . Again, this model is very well studied, and many of its properties have been investigated, including the size distributions of clusters, the conditions under which a giant connected component exists, and the average path length (Molloy and Reed, 1995; Molloy and Reed, 1998; Newman et al., 2001). An overview can be found in Chapter 3 of Durrett, 2006 and Chapter 7 of Hofstad, 2016. As previously, the configuration model generates asymptotically locally tree-like networks (Hofstad, 2024, Section 4.2). While this may not be a realistic setting, the configuration model nonetheless can to some extent produce valid approximations of social networks (Newman et al., 2002), some of which arise in epidemiology (Dhanjal et al., 2011). Finally, we will see in upcoming Section 1.2.2 that this property facilitates the mathematical analysis of epidemic dynamics spreading on such a network.

As mentioned previously, none of the described random graph models exhibit clustering, at least not in the large graph limit. However, for epidemic models, taking into account clustering is actually important, for at least two reasons. First, clustering is known to affect epidemic outcomes, such as the maximum number of infected over time (epidemic peak size), the time at which this peak occurs, or the total number of infected (epidemic final size) (House and Keeling, 2011; Volz et al., 2011). Second, in the case of human contact networks, clustering emerges for example by people sharing a common household, where contacts are likely to be close leading to possibly high contamination risks. As in this case distinguishing within-household from other contacts is natural, this setting will be explored in greater depth in Section 1.1.2 below. However, it already motivates our interest in clustered network models. Here, we will present two such models, which actually rely on the presence of highly connected subgraphs, and which have been used in epidemic models. For an extensive review of random graph models with clustering, we refer to Section 9.4 of Hofstad, 2024 and references therein.

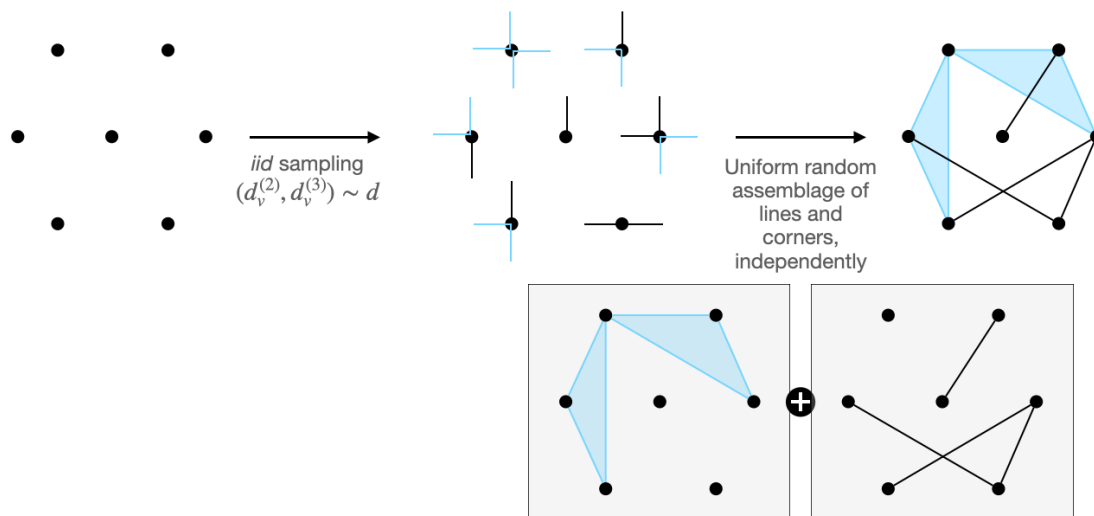
Let us start by introducing random intersection graphs, which first originated in Karoński et al., 1999. Consider a set  $V = \{1, \dots, n\}$  of vertices, as well as a set  $M = \{1, \dots, m_n\}$  of so-called groups or *cliques*. Then each node  $v \in V$  belongs to a group  $k \in M$  with fixed probability  $p_n > 0$ , independently for all groups and from other vertices. Subsequently, the graph  $G(n, m_n, p_n)$  is obtained by joining two vertices by an edge if and only if they are members of at least one common group. Again, this model is well studied. For instance, if  $m_n$  is of order  $n^\alpha$  for  $\alpha > 0$  and under adequate assumptions on the sequences  $(p_n)_{n \geq 1}$ , the large graph limit of the degree distribution has been established (Stark, 2004), which in particular is Poisson if and only if  $\alpha > 1$ , recalling Erdős-Rényi graphs. Also, it seems worth emphasizing that if  $\alpha = 1$ , the model becomes locally tree-like at the level of cliques. This means that if instead of focusing on individual vertices, one actually considers the network



**Figure 1.6:** Construction of a random intersection graph  $G(n, m, p)$  with  $n = 9$  vertices and  $m = 3$  groups. First, each node is connected independently to each group with fixed probability  $p$ . The graph  $G = (V, E)$  is obtained by adding an edge  $(u, v)$  to  $E$  if and only if vertices  $u$  and  $v$  share a group. Colors indicate the different groups.

$G_M$  formed by the groups  $k \in M$  such that two groups are neighbors if and only if they share a common vertex, then  $G_M$  is asymptotically locally tree-like: there are no short loops at the level of clusters (Kurauskas, 2022). Notice that the authors use this result to derive an explicit formula for the asymptotic clustering coefficient. In addition, many variations of this model have been developed, for instance in order to obtain power-law distributed degree distributions when  $n$  goes to infinity (Deijfen and Kets, 2009). Section 9.4 of Hofstad, 2024 yields a larger presentation of the results established for random intersection graphs.

Another way of ensuring the presence of highly connected sub-groups can be obtained by modifications of the configuration model. One possibility of achieving this is the configuration model with clustering, which will also be referred to as *clique configuration model* throughout the thesis. Consider a set of  $n$  vertices. Let  $d$  be a probability distribution taking values in  $(\mathbb{N} \cup \{0\})^2$ . Then independently for each node  $v$ , draw  $(d_v^{(2)}, d_v^{(3)}) \sim d$ , and attach  $d_v^{(2)}$  half-lines and  $d_v^{(3)}$  corners to node  $v$ . Next, as for the configuration model, choose uniformly at random with removal pairs of half-lines and connect them to form a line. Then, choose uniformly at random with removal triples of corners, and connect the three corresponding nodes by edges, forming a triangle. The procedure stops when at most one half-line and at most two corners remain, which are dropped to end the graph construction. Figure 1.7 illustrates the construction of such a clique configuration model. Notice that this procedure can be generalized to include totally connected subgraphs, or cliques, of any size  $k \geq 2$ . The clique configuration model has first been proposed in Newman, 2009, where some properties such as the size of the giant component are derived heuristically. Additionally, the obtained graphs are again supposed to be locally tree-like at the level of cliques (Karrer and Newman, 2010b). However, to our knowledge, no rigorous mathematical analysis has been performed



**Figure 1.7:** Main steps for obtaining one realization of the clique configuration model, where the target distribution of lines and triangles per node is given by  $d$ .

so far in this setting.

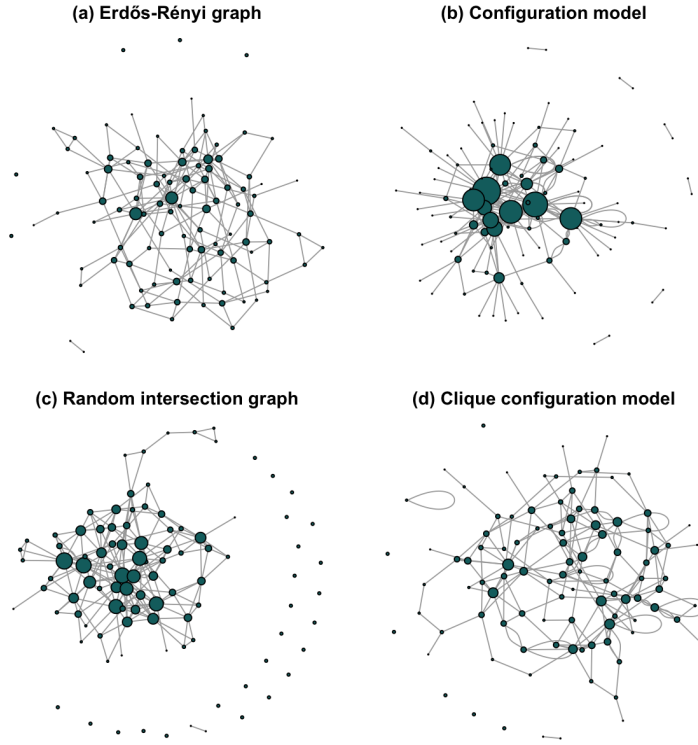
The models introduced in this section (Figure 1.8) allow for a detailed description of the contact networks underlying epidemic spread. It is worth emphasizing here that random graph models can be designed in order to incorporate further aspects which are of interest in an epidemic context, such as spatial structure (Barthélemy, 2011) or evolution of contacts over time (dynamic networks) (Bansal et al., 2010). Another pertinent refinement is to distinguish different types of contacts, occurring at different rates, allowing to take into account that the rate at which contacts occur, or the associated risk of infection, may depend on the social context (at home, with friends, at the grocery store...). This will be the focus of the following subsection.

### Multilayer contact models

In addition to the elements explored above, human contacts are structured by different social contexts in which they can occur: a first glance at this is provided by Figure 1.3, as the contact matrixes significantly differ according to the context under consideration. Further, an empirical study on social contacts in Sweden has shown that several characteristics of contacts, such as their duration, their frequency, the number of people involved, and the probability of the contact being physical, strongly depends on the meeting place (Strömngren et al., 2017). More precisely, based on the data collected, the authors identify five main clusters of meeting places defined by sharing similar traits: the family venue, and vehicle; the fixed activity site (including workplaces and schools); the social network, represented by friends and relatives; and the general population, or trading plaza (stores, public transports, etc.).

From a modeling point of view, taking into account such different types of contacts is especially motivated by non-pharmaceutical interventions such as teleworking or school

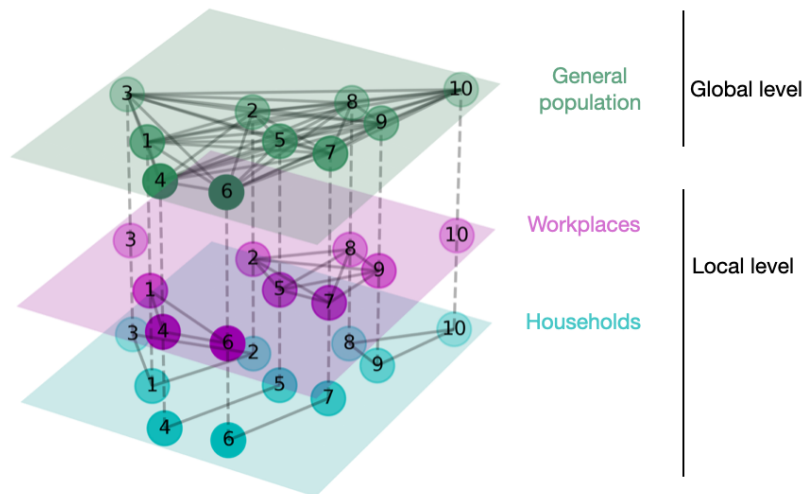




**Figure 1.8:** Examples of realizations of all four major random network models introduced in Section 1.1.2. In all cases, there are 100 nodes and in the large graph limit, the average degree is close to 4. (a) Erdős-Rényi graph  $G(100, 0.04)$ . (b) Configuration model with degree distribution  $p = (p_k, k \geq 0)$  such that  $p_k$  is proportional to  $k^{-1.5}$  for  $k \in \llbracket 1, 27 \rrbracket$ , and equal to zero otherwise (truncated power-law distribution). (c) Random intersection graph  $G(100, 60, 0.0275)$ . (d) Clique configuration model with cliques of sizes  $\llbracket 2, 4 \rrbracket$ . The distribution of number of cliques of size  $j \in \llbracket 2, 4 \rrbracket$  to which a node belongs is given by  $p^{(j)}$ , with  $p^{(2)} \sim \mathcal{B}(4, 0.375)$ ,  $p^{(3)} \sim \mathcal{B}(3, 1/6)$  and  $p^{(4)} \sim \mathcal{B}(2, 0.25)$ .

closures. Indeed, they consist in specifically acting on some type of contacts, while leaving others in place. Recent studies find that the way contacts are represented in epidemic models changes the predicted outcomes of such control measures (Contreras et al., 2022; Di Lauro et al., 2021), as will be discussed in detail below. As a consequence, it seems natural that epidemic models seek to distinguish different social contexts.

This is allowed for by multilayer models, which distinguish different types of contact within the population. The idea is to represent contacts by a multilayer network, where each node belongs to all layers simultaneously. Each layer corresponds to the contacts established within one sort of social context, and thus can be characterized by its network structure and contact rate, for instance. Subsequently, the epidemic spreads on the aggregated contact network, which corresponds to the addition of all three layers. The result of this aggregation can be regarded as a weighted graph, where an edge's weight represents the rate at which a contact occurs along that edge, and with possible multi-edges as some individuals may be neighbors in several layers (e.g. spouses who also work together). This procedure can lead to elaborate contact networks, which mimic realistic settings, including among others



**Figure 1.9:** Multilayer contact network of 10 nodes, obtained by the household-workplace model. Nodes represent individuals, belonging to each layer simultaneously as indicated by their number. Edges represent possible contacts within each layer. The layers and corresponding levels are indicated on the side.

households, workplaces, schools and stores which are grouped into communities in order to take into account the spatial organization of contacts (Backhausz et al., 2022).

A common procedure is to distinguish mainly two or three layers of contact, representing contacts in the general population, and contacts within households and possibly schools or workplaces. To our knowledge, this setting has first been explored by Ball et al., 1997. The authors take into account two layers, the first being uniformly mixing, while the second is characterized by individuals having a limited number of neighbors only. This coined the term of *two levels of mixing*: one *global* level, where uniform mixing implies that contacts are wide-spread in the population, and one *local* level, where contacts are limited by the network structure. In particular, this implies that the associated models are multiscale, with an infinite number of possible partners with whom contact is established at most once at the global scale, as opposed to repeated contacts with a finite number of partners at the local scale. This setting has been explored rather extensively, considering different variations for the local level. In particular, a classical local layer is that of the households model, in which the population is partitioned into households of possibly random size (Ball et al., 1997; House and Keeling, 2008). The main idea behind household models is that contacts are uniformly mixing at high rate within households, whereas contacts are significantly less frequent in the general population. However, in this setting, the epidemic necessarily needs the global level to sustain its spread, as households are otherwise disjoint from one another.

A natural extension is to consider a local level which itself consists in two layers, typically representing households and workplaces. At first, these were assumed to be of fixed size (Ball and Neal, 2002), before considering arbitrary structure size distributions (Pellis et al., 2009). This model will be at the center of attention in Chapters 2 and 3, leading us to introduce it in some detail. As for the household model, the global level corresponds to a uniformly mixing general population of size  $K$ . The local level is split into one layer of households



and one layer of workplaces, which are constructed as follows. We will suppose that there exists a maximal size for households and workplaces  $n_{\max} < \infty$ , and consider two probability distributions  $\pi^H, \pi^W$  with support in  $\llbracket 1, n_{\max} \rrbracket$ . Let  $k$  be the number of individuals who are not yet assigned to a household, so initially  $k = K$ . Pick a household size  $h \sim \pi^H$ , then either  $h \leq k$  and a new household of size  $h$  is formed by choosing uniformly at random  $h$  individuals who are not yet member of a household, or  $h > k$  in which case one last household is assembled containing all  $k$  remaining individuals. Subsequently,  $k$  becomes  $k - h$  and the process stops when  $k = 0$ . The same procedure is repeated independently for workplaces, using  $\pi^W$  instead of  $\pi^H$ . In particular, if  $K$  goes to infinity,  $\pi^H$  and  $\pi^W$  correspond to the household and workplace size distributions, respectively. The obtained multilayer contact network is represented in Figure 1.9. Subsequently, the epidemic spreads at the global level at rate  $\beta_G/K$ . Also, within households and within workplaces, uniform mixing is assumed, and contacts occur at rates  $\lambda_H$  and  $\lambda_W$ , respectively. Once contaminated, individuals remain infected for a random amount of time, which is sampled from the infectious period length distribution  $\nu$  with support in  $(0, +\infty)$ .

Many variations of the household or household-workplace model have been considered in the literature. For instance, one may want to replace the global, uniformly mixing level by another local level, where contacts in the general population are depicted by a configuration model (Ma et al., 2013). It also is possible to consider dynamic networks, allowing for the evolution of the graph over time (Barnard et al., 2018). Indeed, many variations could be considered, as each layer of the graph can be modified to take into account one or another refinement of the contact network.

Nevertheless, considering multilayer contact networks comes at the cost of additional difficulties in understanding epidemic spread, especially when considering a rich local level as in the case of the household-workplace model. Indeed, the model is multi-scale as both a global and local level are considered, and contrary to the case of the household model, the local level now may suffice to sustain the epidemic. As a consequence, unraveling how the model parameters impact epidemic spread is not an easy task. A natural question is to wonder whether it actually worth the effort to consider such models.

As we will see, a natural motivation arises when considering non-pharmaceutical interventions (NPIs) such as teleworking or school closures. Indeed, both empirical and simulation studies show that these are among the most effective NPIs (Mendez-Brito et al., 2021; Backhausz et al., 2022; Simoy and Aparicio, 2021), motivating the desire to understand how these measures act on the epidemic, through both mathematical analysis and simulation studies.

First, it is crucial to identify which level of detail of the contact network needs to be taken into account when assessing such control measures. A recent study has compared different ways of modeling teleworking and school closures, using either a household model or a contact matrix which is split in home, school, workplace and other contacts (Di Lauro et al., 2021). For both models, the authors find that uniformly scaling contacts on the whole contact network yields different results than scaling separately household and general contacts, or each component of the contact matrix, respectively. Hence, distinguishing several types of contact in the population is important when studying such NPIs. Multilayer models seem particularly suiting, as households or workplaces imply clustering, which we already know to significantly impact epidemic dynamics when compared to uniform mixing at large scales which underlies the contact matrix setting.

A related question is to wonder up to which scale one should incorporate detailed information on the contact network: for instance, is considering uniform mixing within households and workplaces enough? This question has been explored by comparing the effectiveness of different interventions when considering either uniform mixing, or empirical contact networks derived from sensor data (Contreras et al., 2022), at the scale of large structures which in most cases have several hundreds of members. The authors' findings suggest that while considering uniform mixing does not change qualitative results, it does have an impact on quantitative aspects which are crucial when studying the cost effectiveness of control measures. On the other hand, in the school setting of Figure 1.4c, the authors mentioned that there seems to be almost uniform mixing within each class (Stehlé et al., 2011). Another study has explored the epidemic impact of different workplace cultures, *i.e.* of how individuals interact at their workplace (Timpka et al., 2016). A significant influence on modeling outcomes is established, raising the question of an effective workplace size depending on how many coworkers typically interact, and how close these contacts are. Hence, it actually is a difficult task to define the precise level at which the advantage of simplifying the model by assuming uniform mixing at small local scales outweighs the implied loss of precision when predicting epidemic outcomes.

Several questions may be addressed regarding NPIs like teleworking or school closures. So far, a commonly investigated subject is the comparison of different NPIs, among which teleworking and/or school closures (Backhausz et al., 2022; Colosi et al., 2022; Favero et al., 2022; Ferguson et al., 2006; Hilton et al., 2022; Simoy and Aparicio, 2021). For teleworking or school closures alone, when to start and for how long to sustain the intervention has also been studied (Bin Nafisah et al., 2018; Britton and Leskelä, 2023), while optimal intensity has been of interest for lockdowns (Britton and Leskelä, 2023; Caulkins et al., 2021; Richard et al., 2021). Notably, even though the modeling choice may impact the quantitative and possibly qualitative outcomes as emphasized by Di Lauro et al., 2021, these issues can be explored without the necessity to consider multilayer models. Indeed, while some of the studies above make use of multilayer models (Backhausz et al., 2022; Ferguson et al., 2006; Hilton et al., 2022; Simoy and Aparicio, 2021), another focuses on social structures alone, without modeling precisely contacts in the general population (Colosi et al., 2022), and some rely on contact matrixes (Richard et al., 2021) or assume uniform mixing (Britton and Leskelä, 2023; Favero et al., 2022; Caulkins et al., 2021).

There are questions, however, which can hardly be considered without making use of multilayer models. For instance, does the way teleworking rates are enforced matter? In other words, how does changing the workplace distribution influence epidemic outcomes? Some answers will be provided in Chapter 2.

Taken together, using multilayer contact networks to better understand how dispatching individuals into small contact structures such as households and workplaces influences epidemic outcomes thus is a pertinent goal. Unfortunately, as mentioned above, analyzing these models is challenging, both in terms of mathematical analysis and numerical exploration, as simulation times grow with the population size. This leads to an interest in developing reduced models.

## 1.2 Reduced models in epidemiology

In epidemiology, many key results are derived using reduced models, in the sense of models which yield good approximations of the epidemic dynamics at hand, but which are more prone to mathematical analysis and/or numerical exploration than the original model. Indeed, this already holds true for the classical *SIR* model: instead of studying the stochastic *SIR* model in a finite population, resorting to branching approximations and the Kermack-McKendrick equations arising in the large population limit is a fruitful approach. This will be illustrated in upcoming Section 1.2.1.

Indeed, model reduction can have several benefits. By essence, it consists in trying to keep as much information as necessary to capture the dynamics of interest, while removing as much detail as possible. As a consequence, it can make emerge key parameters, such as the reproduction number or epidemic growth rate. Further, the reduced models obtained can serve as a basis to develop more involved models.

Thus, the heart of the present thesis will be to establish such reduced models in order to study epidemic models of the household-workplace type. More precisely, we will focus on large population approximations of the epidemic dynamics. This motivates a closer look at the results previously obtained for epidemics spreading on related networks in Section 1.2.2.

### 1.2.1 A motivating example: the uniformly mixing *SIR* model

The uniformly mixing stochastic *SIR* model can be defined as follows. Let  $\lambda$  and  $\gamma$  be the contact rate and infection rate, respectively. Consider a population of  $N$  individuals, and at time  $t \geq 0$ , let  $S(t)$  and  $I(t)$  be the number of susceptible and infected individuals in the population, from which it follows that the number of recovered is given by  $N - S(t) - I(t)$ . Then the couple  $(S(t), I(t))_{t \geq 0}$  defines a Markov process taking values in  $\{(S, I) : 0 \leq S, I \leq N, S + I \leq N\}$ , whose transitions are given by:

$$\begin{aligned} (S, I) &\rightarrow (S - 1, I + 1) && \text{at rate } \frac{\lambda}{N}SI, \\ (S, I) &\rightarrow (S, I - 1) && \text{at rate } \gamma I. \end{aligned}$$

Despite its apparent simplicity, the analysis of this stochastic model is actually not immediate and many key results are derived by the following scheme: the stochastic *SIR* model is shown to be well approximated by another, handier model, which subsequently is used to answer the question of interest.

A first example consists in studying the beginning of the epidemic: given the introduction of one individual into a large susceptible population, is it possible that a significant proportion of the population will be contaminated throughout the epidemic? The answer is provided by the reproduction number,  $R_0 = \lambda/\gamma$ , which is such that a major outbreak may occur if and only if  $R_0 > 1$ . In order to establish this result, the epidemic dynamics are reduced to a discrete-time branching process by neglecting the depletion in susceptible individuals: infection events are regarded as births, removals as deaths, and the branching property is ensured by assuming that  $S/N \approx 1$ . Similarly, the epidemic growth rate  $r = \lambda - \gamma$  can be derived as the Malthusian parameter of a continuous-time branching process, which approximates the epidemic dynamics.

In order to transfer these results from the branching process approximation to the stochastic *SIR* model, it remains to give a rigorous meaning to the idea that the former is "well

approximating” the latter. This is achieved by coupling arguments, which show that the branching process can be constructed on the same probability space as to coincide with the epidemic process of the stochastic *SIR* model, until the infected population becomes of order  $K/(\log K)^a$  (Bansaye et al., 2023b). We refer for instance to Section 3.3 of Andersson and Britton, 2000b for an introduction to coupling methods for epidemic models, and Barbour and Reinert, 2013 for a detailed arguments. Notably, branching approximations actually also serve to derive reproduction numbers and epidemic growth rates for more involved models, including the *SIR* model on configuration graphs (Britton et al., 2007; Barbour and Reinert, 2013), random intersection graphs (Ball et al., 2014; Fransson, 2022) and models with two levels of mixing (Britton and Pardoux, 2019a; Pellis et al., 2011).

Finally, let us mention here that it is possible to derive more information than the reproduction number and epidemic growth rate from branching approximations. For instance, they may be used to capture the tree of infections at the beginning of the epidemic. As demonstrated in Britton and Scalia Tomba, 2019, such models thus allow to point out and rectify biases associated to statistical inference from contact tracing data.

A second, classical example of model reduction consists in studying the large population limit of the epidemic dynamics. Indeed, suppose that  $(S(0), I(0))/N$  converges in probability to constants  $(s_0, i_0)$ . Then the renormalized process  $(S/N, I/N)$  converges in probability, uniformly on finite time intervals, to the unique solution  $(s, i)$  of the Kermack-McKendrick ODE system

$$\begin{aligned} s'(t) &= -\lambda si, & i'(t) &= \lambda si - \gamma i, \\ s(0) &= s_0, & i(0) &= i_0. \end{aligned}$$

This result actually follows from general convergence results for density-dependent Markov jump processes. One may further establish Gaussian fluctuations of the stochastic process around its deterministic limit, hence quantifying the asymptotic approximation error. We refer to Section 5 of Andersson and Britton, 2000b for detail. In particular, this rigorously exhibits the connection between the stochastic and deterministic *SIR* models, the latter of which allows to derive important results such as the final epidemic size in the large population limit (Brauer et al., 2008, Section 2.1.2). Similar results can be obtained in more general settings, for instance in a non-Markovian model where infectious periods are not necessarily exponentially distributed (Kurtz, 1981).

For network models, however, capturing the deterministic dynamics emerging in the large population limit is not straightforward. Let us first mention that closed, dynamical systems describing the exact evolution of the sanitary state of each individual node over time are achievable for finite-size graphs, by focusing either on the network level (lumping of the arising master equations) or node level (generalized closure) (Kiss et al., 2017, Sections 2.4 and 3). However, the obtained systems of equations become intractable as the number of nodes grows large, due to the number of equations scaling at least linearly with the network size. As a consequence, alternative approaches have been developed, which lead to closed, finite dynamical systems which either exactly describe or yield useful approximations of the large population dynamics. Here, we will start by introducing some intuition behind these large population approximations, before giving an overview of their application to different network models.

## 1.2.2 Some heuristics for reduction of epidemics on networks

In order to describe epidemic dynamics on networks, several approaches have been developed. Here, we will focus on three of the most common points of view: pairwise models, effective degree models and edge based compartmental models. We will consider the case of an *SIR* model with infection rate  $\lambda$  and recovery rate  $\gamma$ . For a more extensive review, we refer to Sections 4.1, 4.2, 5.6 and 6.5 of Kiss et al., 2017, which partly inspire the following paragraphs.

### Pairwise models

Instead of considering the epidemic state of each individual node, let us focus on the average number of nodes  $[X]$  (or their proportion as  $N \rightarrow \infty$ ) in a given epidemic state  $X \in \{S, I, R\}$ . Similarly, for  $X, Y, Z \in \{S, I, R\}$ ,  $[XY]$  designates the average number of nodes  $u$  and  $v$  in state  $X$  and  $Y$  respectively which are connected by an edge,  $[XYZ]$  the average number of triples of nodes  $u, v, w$  respectively in states  $X, Y, Z$  such that the edges  $(u, v)$  and  $(v, w)$  exist, and so on.

In this case, one may notice that the number of susceptibles decreases each time an infection occurs, which happens at rate  $\lambda[SI]$  as the disease can only spread along the edges of the graph. In turn, an *SI* edge becomes of type *SR* if the infected recovers, at rate  $\gamma$ , or of type *II* if the susceptible is infected. Such an infection happens at rate  $\lambda$  either along that given edge, or along any other edge connecting the susceptible to another infected. Similarly, if one of two susceptible neighbors  $u$  and  $v$  is infected by one of his infectious neighbors, a new *SI* pair is created. In this case,  $[SS]$  decreases by two as both oriented edges  $(u, v)$  and  $(v, u)$  originally were counted in  $[SS]$ . By continuing this reasoning for other combinations of epidemic states, we obtain the following ODE system:

$$\begin{aligned} [S]' &= -\lambda[SI], \\ [I]' &= \lambda[SI] - \gamma[I], \\ [SS]' &= -2\lambda[SSI], \\ [SI]' &= \gamma[SI] + \lambda([SSI] - [ISI] - [SI]), \\ [II]' &= -2\gamma[II] + 2\lambda([SI] + [ISI]), \dots \end{aligned}$$

This system is technically speaking exact, but unclosed: in order to describe the dynamics of triples, one needs to consider quadruples, and so on. Pairwise models thus propose to close the dynamical system by expressing the triples as functions of pairs, under simplifying assumptions. Originally, the first pairwise model has been developed for  $n$ -regular graphs containing  $N$  nodes, where each node has exactly  $n$  neighbors (Keeling et al., 1997). Here, we will make the same assumption for ease of presentation. In this case, the following approximation can be derived for the dynamics of triples:

$$[XYZ] \approx \frac{n-1}{n} \frac{[XY][YZ]}{[Y]}.$$

Let us decompose this expression. Consider three nodes  $u, v, w$  respectively in states  $X, Y, Z$ , such that the edges  $(u, v)$  and  $(v, w)$  exist. If  $v$  is chosen uniformly among nodes in state  $Y$ , the probability that  $v$  has a  $X$ -type neighbor is given by  $[XY]/(n[Y])$  as  $n[Y]$  edges start at nodes of type  $Y$ . Further, there are  $[Y]$  possible choices for  $v$  and  $n(n-1)$

possible choices for neighbors  $u$  and  $w$ . If we ignore any correlation among the epidemic states of  $u$  and  $w$  which is not due to their common neighbor  $v$ , we obtain that  $[XYZ] \approx [Y]n(n-1)[XY][YZ]/(n[Y])^2$ . Of course, there are cases when  $u$  and  $w$  are correlated not only through  $v$ , for instance if the three nodes form a triangle. However, on unclustered networks as in the large graph limit of the  $n$ -regular graph, this is rare, justifying the above approximation.

We finally obtain the following pairwise model:

$$\begin{aligned} [S]' &= -\lambda[SI] \\ [I]' &= \lambda[SI] - \gamma[I] \\ [SS]' &= -2\lambda \frac{n-1}{n} \frac{[SS][SI]}{[S]} \\ [SI]' &= -\gamma[SI] + \lambda \left( \frac{n-1}{n} \frac{[SS][SI]}{[S]} - \frac{n-1}{n} \frac{[SI]^2}{[S]} - [SI] \right). \end{aligned}$$

In the case where the graph is not regular, Eames and Keeling, 2002 have derived a pairwise model which takes into account the nodes' degree and has  $O(M^2)$  equations, where  $M$  is the number of possible degrees in the network. This will serve as a basis for comparison with the upcoming effective degree and edge based models.

### Effective degree models

Another family of reduced models focuses on the proportion  $S_{s,i}$  (resp.  $I_{s,i}$ ) of nodes which are susceptible (resp. infected) and have  $s$  susceptible and  $i$  infected neighbors. Notably, the number of recovered neighbors does not need to be included in order to describe the epidemic dynamics: this coins the term "effective degree", which only takes into account a node's edges leading to susceptible and infected individuals.

Let us consider for example the dynamics of  $S_{s,i}$ , which may be affected by several possible events. First, a node of this kind can be contaminated by one of its infected neighbors: this occurs at rate  $\lambda i$ , and the node becomes of type  $I_{s,i}$ . Second, one of its infected neighbors may recover, yielding a node of type  $S_{s,i-1}$  at rate  $\gamma i$ . Third, a node of type  $S_{s,i+1}$  becomes of type  $S_{s,i}$  following the removal of one of its infected neighbors, at rate  $\gamma(i+1)$ . Finally, a susceptible neighbor may be infected by another node, in which case our node becomes of type  $S_{s-1,i+1}$ . Similarly, a node of type  $S_{s+1,i-1}$  may become of type  $S_{s,i}$  upon infection of a susceptible neighbor. Hence, as for the above pairwise model, we need to take into account the neighbors of our initial node's neighbors, *i.e.* triples, whose dynamics depend on quadruples and so forth: the dynamical system needs to be closed.

Here, we will focus on the approach of Lindquist et al., 2011, who consider unclustered, simple graphs (*i.e.* no multiple edges or self loops). In addition, they suppose the maximal degree  $M$  to be finite, and that there is no degree correlation among neighbors. Typically, this corresponds to the large graph limit of configuration networks with *i.i.d* degrees sampled from a bounded distribution.

In this setting, the closure proceeds as follows. Susceptible nodes see their neighborhood change due to the infection of a susceptible neighbor, and such an infection occurs at total rate  $\sum_{n=1}^M \sum_{k+l=n} k\lambda l S_{k,\ell}$ . Let  $p$  be the probability that a node whose neighborhood has changed following the infection event is in state  $S_{s,i}$ . Then one may notice that  $p$  is

proportional to  $s$ , hence  $p = sS_{s,i} / \sum_{n=1}^M \sum_{k+l=n} \ell S_{k,\ell}$ . Letting

$$\Lambda_S = \frac{\sum_{n=1}^M \sum_{k+l=n} k \lambda \ell S_{k,\ell}}{\sum_{n=1}^M \sum_{k+l=n} \ell S_{k,\ell}},$$

a node in state  $S_{s,i}$  thus becomes  $S_{s-1,i+1}$  at rate  $\Lambda_S s S_{s,i}$ . Notice that we again assume the absence of correlation among the health states of a given node's neighbors, aside from their connection to that node. This illustrates the importance of the assumptions made on the graph.

Reasoning similarly for infected nodes  $I_{s,i}$ , and letting

$$\Lambda_I = \frac{\sum_{n=1}^M \sum_{k+l=n} \lambda k^2 S_{k,\ell}}{\sum_{n=1}^M \sum_{k+l=n} \ell S_{k,\ell}},$$

we finally obtain the effective degree model: for any non-negative integers  $s, i$  such that  $s + i \leq M$ ,

$$\begin{aligned} S'_{s,i} &= -\lambda i S_{s,i} + \gamma((i+1)S_{s,i+1} - iS_{s,i}) + \Lambda_S((s+1)S_{s+1,i-1} - sS_{s,i}) \\ I'_{s,i} &= \lambda i S_{s,i} + \gamma(i+1)(I_{s,i+1} - I_{s,i}) + \Lambda_I((s+1)I_{s+1,i-1} - sI_{s,i}). \end{aligned}$$

The proportions of susceptible and infected nodes are finally computed using  $S = \sum_{n=1}^M \sum_{s+i=n} S_{s,i}$  and  $I = \sum_{n=1}^M \sum_{s+i=n} I_{s,i}$ . In particular, this dynamical system contains  $M(M+3)$  equations. If the degree distribution charges all elements in  $\{1, \dots, M\}$ , the pairwise and effective degree models thus have the same order of equations.

### Edge-based compartmental models (EBCM)

Last but not least, let us explain how to derive an edge-based compartmental model (EBCM). Again, we will consider the case of the large graph limit of a configuration model with *i.i.d.* degrees, which we know to be an almost surely simple and locally tree-like graph. While originally developed by Volz, 2008, we will here follow the lines of Miller, 2011. The author proposed an alternative derivation of the EBCM, leading to a more concise formulation consisting in only two differential equation, instead of four. Throughout the following, let  $p = (p_k, k \geq 0)$  be the degree distribution, and  $\Psi$  its probability generating function: for  $s \in [0, 1]$ ,

$$\Psi(s) = \sum_{k \geq 0} p_k s^k.$$

The starting point of the EBCM is the observation that a node  $u$  is susceptible at time  $t$  if and only if it has not been contaminated by any of its neighbors. In addition, as the graph is locally tree-like, there is no short path connecting to neighbors  $v$  and  $w$  of  $u$  other than the one passing through  $u$ . Hence, if  $u$  is susceptible, the sanitary state of all its neighbors can be considered to be independent from one another. Thus, if  $\theta(t)$  is the probability that at time  $t$ , the disease has not been transmitted along a random edge, the proportion  $S(t)$  of susceptible nodes at time  $t$  is given by

$$S(t) = \sum_{k \geq 0} p_k \theta(t)^k = \Psi(\theta(t)). \quad (1.2)$$



As further, by definition of the *SIR* model,

$$I'(t) = -S'(t) - \gamma I(t), \quad (1.3)$$

it is enough to derive the dynamics of  $\theta$  in order to capture the epidemic spread. Let  $\phi(t)$  be the probability that at time  $t$ , an edge connects a susceptible and an infected individual, but that the disease has not been transmitted along this edge. Then by definition,

$$\theta'(t) = -\lambda\phi(t). \quad (1.4)$$

Let us thus focus on the dynamics of  $\phi$ . On the one hand,  $\phi$  decreases either if the infection is transmitted along such an edge, at rate  $\lambda$ , or if the infected extremity recovers, at rate  $\gamma$ . Conversely,  $\phi$  increases whenever a susceptible neighbor  $v$  of a susceptible node  $u$  is infected by one of its neighbors. If  $h(t)$  designates the probability that  $v$  is still susceptible at time  $t$ , such an infection occurs at rate  $-h'(t)$ . As we have followed an edge to reach  $v$ , by definition of the configuration model, the probability that  $v$  is of degree  $k$  is proportional to  $kp_k$ . Further,  $v$  is susceptible if and only if it has not been contaminated by any of its neighbors: if  $v$  is of degree  $k$ , this occurs with probability  $\theta^{k-1}(t)$  since we know for sure that one of the neighbors,  $u$ , is susceptible. Thus

$$h(t) = \frac{\sum_{k \geq 1} kp_k \theta(t)^{k-1}}{\sum_{k \geq 1} kp_k} = \frac{\Psi'(\theta(t))}{\Psi(1)}.$$

From Equation (1.4), it follows that

$$\phi'(t) = -h'(t) - \phi(t)(\lambda + \gamma) = \phi(t) \left( \lambda \frac{\Psi''(\theta(t))}{\Psi(1)} - \lambda - \gamma \right).$$

Thus, using the fact that  $\phi(-\infty) = 0$  and  $\theta(-\infty) = 1$ ,

$$\phi(t) = -\frac{\Psi'(\theta(t))}{\Psi(1)} + \theta(t) + \frac{\gamma}{\lambda}(1 - \theta(t)).$$

Finally,

$$\theta'(t) = \lambda\theta(t) + \lambda \frac{\Psi'(\theta(t))}{\Psi(1)} + \gamma(1 - \theta(t)). \quad (1.5)$$

The EBCM is then given by differential Equations (1.3) and (1.5), complemented by Equation (1.2), making it remarkably parsimonious when compared to the pairwise and effective degree models. In addition, while originally being derived heuristically, the EBCM has been established to be exact in the large population limit of the stochastic *SIR* model on a configuration graph. This has first been achieved under a finite fifth moment assumption on the degree distribution (Decreusefond et al., 2012), which has subsequently been relaxed to a finite second moment condition (Janson et al., 2014).

This ends our brief dive into some of the major families of large population approximations for epidemics on networks. However, let us mention here that while the pairwise, effective degree and edge-based compartmental model formalisms have been studied extensively and extended to various settings, other approaches exist. For instance, methods from related mathematical topics can be adapted to yield new reductions. Indeed, Karrer and Newman,



2010a make use of the message passing algorithms developed originally for belief propagation on graphical models. This yields exact results for non-Markovian *SIR* models on locally-tree like networks, and bounds for the epidemic on general networks. In addition, reduced models naturally are influenced by the characteristics of the contact network under consideration. For example, if individuals are characterized by a trait which influences whom makes contact with whom, the large population limit can make emerge epidemic processes spreading on graphons (Delmas et al., 2023), allowing to retain that information.

This diversity motivates the following section, which aims at summarizing the main results regarding large population approximations for epidemic dynamics on different kinds of contact networks.

### 1.2.3 An overview of reduction results for epidemics on random graphs

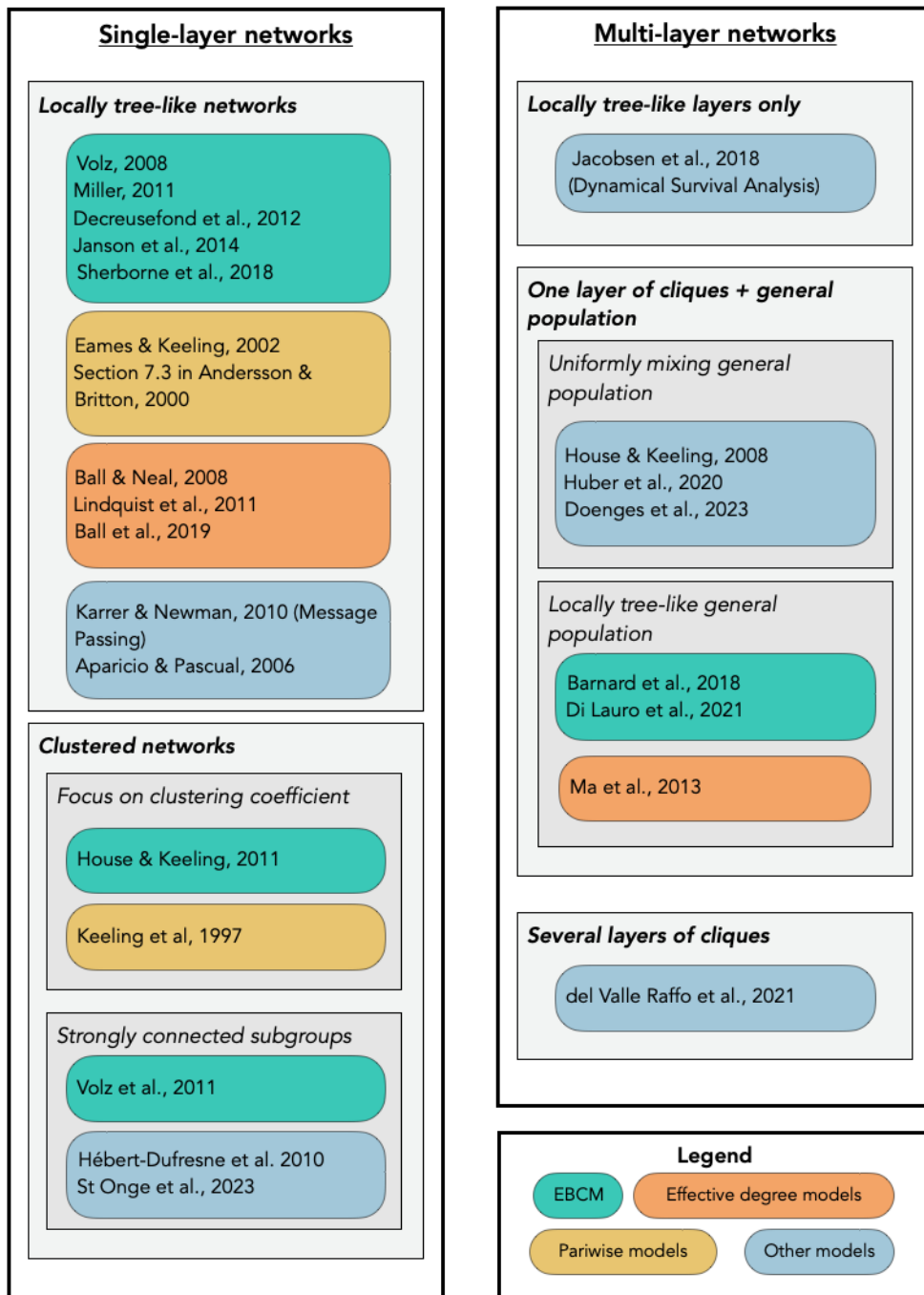
We will now turn to the application of the methods mentioned above, among others, to random contact networks of increasing complexity. We do not claim to give an exhaustive overview, and instead have aimed at including main results, as well as some recent developments for clustered and multi-layer models. The present state of the art is summarized in Figure 1.10.

Given that the epidemic spreads along edges of the graph connecting susceptible and infected nodes, it is key to have a good understanding of their neighborhood, as emphasized by the previously introduced reduction methods. As a consequence, locally tree-like graphs appear to be a good starting-point, and many reduced models have been developed in this setting.

In particular, the *SIR* model on a configuration graph has been the considered for the introduction of edge-based compartmental models in Volz, 2008, which since have been extended to include pertinent variations of the original model, such as non-Markovian settings (Sherborne et al., 2018). Similarly, configuration models have served as a basis for the development of effective degree models (Ball and Neal, 2008; Lindquist et al., 2011), including recently dynamical configuration models (Ball et al., 2019). Further, the pairwise model with closure at the level of triplets has been applied to Poisson graphs (Andersson and Britton, 2000b, Section 7.3), which emerge as the large graph limit of  $G(n, p/n)$  Erdős-Rényi graphs. In general, it is well known that ignoring clustering simplifies the equations arising in pairwise models (Eames and Keeling, 2002). Also, locally tree-like graphs are particularly well-suited for the use of message-passing equations, which are exact in this setting (Karrer and Newman, 2010a).

Given this multitude of large graph approximations, the equivalence between various reduced models has been studied under appropriate assumptions (House and Keeling, 2011; Jacobsen et al., 2018; Kiss et al., 2023; Sherborne et al., 2018; Wilkinson and Sharkey, 2014; Wilkinson et al., 2017). Let us also mention Section 7 of Kiss et al., 2017 for an overview of equivalence results.

Finally, Aparicio and Pascual, 2006 alternatively propose to approximate *SIRS* dynamics on the Poisson graph using a parsimonious, uniformly mixing model. Their result relies on calibrating the model parameters using a reproduction number, and the addition of a *Y* compartment containing infected but no longer infectious individuals. This addition is motivated by the observation that on Poisson contact networks, infected individuals often cause most their secondary infections on a shorter timespan than their actual infectious



**Figure 1.10:** Deterministic reduced models proposed in the large graph approximation, for different types of contact networks.

period. They obtain satisfactory approximations of the epidemic dynamics in simulations.

However, despite locally tree-like network being a favorable setting, taking into account clustering is a natural and crucial extension as real-life networks tend to be clustered which impacts epidemic outcomes.

A first means of achieving this consists in capturing the network's clustering via its clustering coefficient only. This was originally proposed for the pairwise model (Keeling et al., 1997), and has since been adapted to derive a clustered edge-based compartmental model (House and Keeling, 2011). Otherwise, it is possible to focus on a specific type of clustering. Indeed, recall that due to social structures such as households and workplaces, real-life contact networks tend to contain strongly connected subgroups (*cliques*), such as households or workplaces. This leads to focusing on (variants of) the clique configuration model and overlapping groups model, which are designed to incorporate such patterns.

For example, Volz et al., 2011 adapt the EBCM formalism to clique configuration models by considering the network at the level of cliques. Indeed, the graph then appears to be locally tree-like again: the sanitary state of the cliques a susceptible individual belongs to will thus be independent from one another in the large graph limit. In addition, by comparing the epidemic outcomes of their EBCM with those from the EBCM proposed by House and Keeling, 2011, the authors demonstrate that the clustering coefficient alone is insufficient to capture the epidemic impact of cliques. This demonstrates the importance of explicitly modeling the way clustering is achieved within the population.

Let us further mention related settings studied in St-Onge et al., 2023, where contact rates are allowed to differ between cliques as random functions of their size, as well as Hébert-Dufresne et al., 2010, where cliques are not necessarily fully connected. Before proceeding, it is worth emphasizing that all these models taking into account cliques share a major common point: they focus on the epidemic at the level of structures, as they keep track of the proportions of cliques containing a certain number of susceptibles and infected. As a consequence, they lead to high-dimensional reduced models for larger clique sizes. This will also hold true for the forthcoming reduced models for multi-layer settings including cliques, which are derived from EBCM and effective degree model formalisms.

Indeed, as already argued in Section 1.1.2, considering multi-layer networks is very pertinent for studying epidemic spread, leading to innovative approaches being developed in this setting. For instance, the first effective degree model was actually established for the case of a configuration model topped by a uniformly mixing layer (Ball and Neal, 2008), and recently a dynamical survival analysis approach has been developed to study a dynamical multi-layer configuration network (Jacobsen et al., 2018).

Most effort, however, has been spent on household-model like networks. In this case, the contact network contains two layers, namely one layer where each individual belongs to one fully connected clique, *i.e.* a household, and one layer which depicts contacts in the general population. In the pioneering work of House and Keeling, 2008, the general population was assumed to be uniformly mixing. In this case, the reduced model takes the form of a dynamical system whose variables are the proportions of households in a given configuration (number of susceptible, infected and recovered members). This ODE system appears to be the large population limit of its stochastic counterpart, which similarly to the uniformly mixing model can be formalized as a finite type density-dependent Markov jump process.

Let us mention here that recently, more coarse-grained model reductions have been proposed for this household model. They rely on simplifying assumptions such as within-household spread being significantly stronger than mean-field spread (Huber et al., 2020) or aggregating all secondary infections caused within a household following a mean-field infection (Doenges et al., 2023). Further variations of the household model include depicting the general population as a configuration network (EBCM: Di Lauro et al., 2021; effective degree model: Ma et al., 2013), and allowing individuals to belong to several "households" at once (Barnard et al., 2018, EBCM).

Nevertheless, if one is interested in distinguishing different types of cliques as for the household-workplace model, the previously introduced results are not enough. Indeed, the model now needs to take into account two separate layers of cliques. In this setting, prior to the present thesis, the only reduced model consists in extending the aforementioned approach of Aparicio and Pascual, 2006. This has been achieved by del Valle Rafo et al., 2021 for an *SIRS* model spreading within households and workplaces, which is approximated using a uniformly mixing model. In this case, parameters are calibrated using a numerically estimated reproduction number and the network's clustering coefficient. Also, as previously, an additional *Y* compartment is introduced. Again, the obtained reduced model numerically yields satisfying approximations of key epidemic outcomes of the household-workplace epidemic.

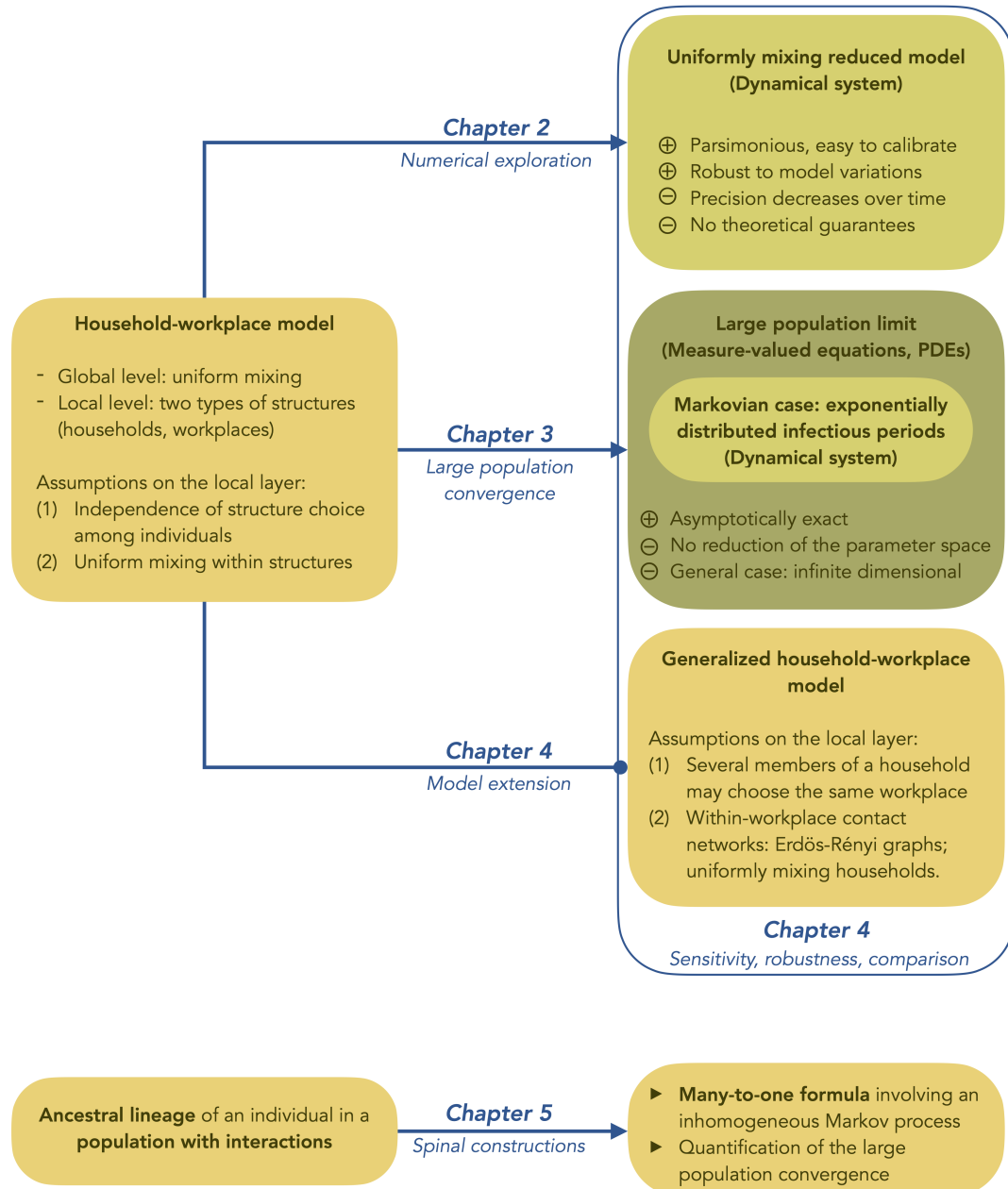
However, there are several drawbacks to this approach. Indeed, model calibration may not be straightforward in a real-life setting, as both epidemic and network parameters need to be inferred. In addition, there are no theoretical guarantees of the approximation error, and the reduced model is not fine-grained enough to fit the epidemic dynamics over time. These challenges will be addressed in this thesis, as will become apparent in the following section.

### 1.3 Main contributions of the thesis

This section is dedicated to introducing and summarizing the results of Chapters 2 to 5. Figure 1.11 yields an overview of the considered models and their connection, indicating the focus of each chapter. Notably, Chapters 2 to 4 aim at proposing and analyzing reduced models which approximate the household-workplace model, introduced in Section 1.1.2. Here, we recall its parameters, as they will be of use throughout the following sections:

- $\pi^H$  and  $\pi^W$  designate the household and workplace size distributions. We assume the maximal household and workplace size  $n_{\max}$  to be finite.
- $\beta_G$ ,  $\lambda_H$  and  $\lambda_W$  are the contact rates for each graph layer: general population, households and workplaces.
- Infectious period lengths are distributed according to a probability measure  $\nu$  on  $\mathbb{R}_+$ , which is assumed to be absolutely continuous with respect to the Lebesgue measure. In the Markovian case,  $\nu$  is an exponential distribution of parameter  $\gamma$  which corresponds to the recovery rate.

We can now turn to the main contributions of this thesis.



**Figure 1.11:** Summary of the different models considered throughout the thesis, as well as their relationships and the chapters they relate to. Stochastic models are shown in yellow, deterministic models in green.

### 1.3.1 Chapter 2 - The epidemic footprint of contact structures

*This chapter is based on the preprint "The epidemiological footprint of contact structures in models with two levels of mixing", which is a joint work with Vincent Bansaye, François Deslandes and Elisabeta Vergu (Bansaye et al., 2023a).*

Generally speaking, contact structures such as households, workplaces or schools are a natural component of human contact networks, inducing these networks to be clustered. Further, it is well known that clustering has a significant impact on epidemic outcomes (House and Keeling, 2011), and that this effect depends on the way clustering is achieved within the network (Volz et al., 2011). Similarly, when trying to analyse control measures such as teleworking or school closures, Di Lauro et al., 2021 show that the way contacts are represented in mathematical models modifies the predicted outcomes of such interventions. In particular, teleworking acts by modifying the workplace size distribution. Taken together, this raises the question: what is the impact on epidemic outcomes of the size distribution of strongly connected subgroups, such as workplaces?

Given the impact of the contact modeling choice on epidemic and intervention outcomes, we have addressed this question by studying the household-workplace model, since it explicitly distinguishes contacts within households and workplaces. As will become apparent throughout the chapter, the epidemic impact of such small contact structures is significant yet subtle, with no straightforward analytical expression. Thus, we propose two indicators for this epidemic impact which as we will see are related to other crucial issues: the design of efficient teleworking strategies, and model reduction.

In order to achieve this, we have undergone a numerical exploration of the household-workplace model. Using stochastic simulations, we explore several epidemic parameter sets, which are inspired by COVID-19, influenza and varicella-like settings. These scenarios are combined with a variety of workplace size distributions, among which a reference distribution based on workplace sizes in Ile-de-France (INSEE, 2018). On the contrary, the household size distribution is kept fixed as it is generally not acted upon by control policies, and corresponds to the one observed in France (INSEE, 2018).

We start by comparing two teleworking strategies. The first strategy consists in a naive approach, where the number of workers allowed on site is proportional to the workplace size (*linear* strategy), whereas it is proportional to the square root of the workplace size for the second strategy (*sublinear* strategy). We observe that the sublinear strategy outperforms the linear strategy: at fixed teleworking rate, the observed reduction in epidemic peak and final size is stronger when the sublinear strategy is used instead of the linear one. Notice that, if the teleworking rate is fixed, the average number of employees per workplace who are not teleworking is fixed as well, as the number of workplaces is constant by essence.

Thus, we are interested in proposing an indicator for the epidemic impact of workplace size distribution  $\pi^W$ , at fixed mean. Through simulations, we show that in this case, the epidemic growth rate, peak size and final size are linearly correlated to the variance of  $\pi^W$ , which thus corresponds to our first indicator. In particular, the sublinear strategy induces by construction a stronger reduction in the variance of the number of employees on site than the linear strategy. This illustrates the pertinence of such indicators, which can be put in practice for the design of efficient control strategies.

However, one drawback of such numerical explorations is that simulations of the stochastic household-workplace model are time-consuming, especially when considering large populations. In addition, the household-workplace model can be difficult to calibrate in practice, as it requests good knowledge of both network parameters and epidemic parameters. Next, we thus aim at proposing a parsimonious reduced model which approximates the epidemic dynamics of the household-workplace model. More precisely, we suggest to make use of a deterministic, uniformly mixing *SIR* model, which depends on two parameters only: a recovery rate  $\gamma$  and a reduced contact rate  $\beta$ . We assume the recovery rate  $\gamma$  to be known, which is generally the case through medical expertise, and it thus only remains to fix the contact rate.

Two natural ways for doing so would be to choose  $\beta$  as to ensure that the reduced and original models either share the same reproduction number, or the same epidemic growth rate. On the one hand, contrary to the uniformly mixing case, there exist several reproduction numbers for the household-workplace model (Ball et al., 2016). On the other hand, the epidemic growth rate is well characterized for our model as the unique solution to an implicit equation (Pellis et al., 2011), for which we derive partially more explicit formulas. As a consequence, we choose to calibrate the reduced model using the epidemic growth rate  $r$ , leading to the following system of equations describing the dynamics of the proportion of susceptibles  $s$  and infected  $i$ :

$$s'(t) = -(\gamma + r)si, \quad i'(t) = (\gamma + r)si - \gamma i.$$

In order to assess the performance of this reduced model, we have computed through simulations the approximation error for the epidemic peak and final size, for a variety of epidemic scenarios and workplace size distributions. In most cases, the error is less than 5%, which is satisfactory. In addition, we notice that the model better approximates the peak size than the final size, which is coherent as it has been calibrated on the initial exponential growth phase. We further have assessed the robustness of this approach to different model variations, namely considering either an *SEIR* model, Gamma-distributed infectious periods (*i.e.* a non-Markovian model) or a sublinear infection rate within structures. In all cases, the approximation performs reasonably well.

This model reduction has several strengths. It is elegantly parsimonious, and expected to be relatively easy to calibrate from data, as it only relies on the recovery and epidemic growth rates. In particular, this is advantageous compared to del Valle Rafo et al., 2021. Further, these results establish that the epidemic growth rate is a good indicator for the epidemic impact of the structure size distributions  $\pi^H$  and  $\pi^W$ .

However, there also are some limitations. Indeed, as mentioned previously, the precision of the reduced model decreases over time, and there are neither theoretical guarantees on the goodness of fit, nor analytical quantifications of the expected approximation error. This motivates the development of another reduced model, which complements the present approach. This is at the heart of the next chapter.

### 1.3.2 Chapter 3 - Large population limit for a multilayer *SIR* model including households and workplaces

*This chapter is based on the preprint "Large population limit for a multilayer SIR model including households and workplaces" (Kubasch, 2023).*



We study the large population limit of the household-workplace model, allowing us to derive a reduced model which yields the exact epidemic dynamics in infinitely large populations. We start by formalizing the epidemic in a population of finite size  $K$  as a measure-valued Markov process, which corresponds to a favorable mathematical setting for studying the large population limit through the associated martingale problem (e.g. Fournier and Méléard, 2004). In our setting, this is actually not straightforward. Indeed, infected individuals correlate the epidemic state of the household and workplace they belong to. This is illustrated by removal events, during which the infected needs to recover in both structures simultaneously. As a consequence, the intuitive description of households and workplaces by the number of susceptible, infected and recovered members is insufficient to correctly represent the epidemic dynamics.

In order to tackle this difficulty, we propose to incorporate into the description of structures the remaining duration of the infectious period length of each individual who has been in contact with the disease, similarly to branching approximations (Ball et al., 2014). In particular, we thus recover dynamics which recall age-structured population models (Tran, 2006), further allowing us to take into account a wide class of infectious period length distributions within a Markovian setting.

Notice that there actually are two sources of randomness: the first corresponds to the realization of the contact network, and the second to the spread of the epidemic process. Our random graph model, whose construction is detailed in Chapter 3, ensures that each individual chooses a household and workplace independently from one another, and from other individuals. In addition, the obtained sequence of contact networks for populations of size  $K \geq 1$  converges almost surely to an infinite version of the household-workplace network, where the household and workplace size distributions are exactly given by  $\pi^H$  and  $\pi^W$ . Here, we consider a fixed realization  $(\mathbf{G}^K)_{K \geq 1}$  of this sequence for which this convergence holds, and we thus do not emphasize further the randomness of the contact network.

Given that realization of the contact network for population size  $K$ , we can now express the epidemic process as an agent based model, whose agents are households and workplaces. Each structure is characterized by its size  $n$ , the number of susceptible members  $s$ , and a vector  $\tau \in \mathbb{R}^{n_{\max}}$  keeping track of infected and recovered members. More precisely, for  $k \leq n - s$ ,  $\tau_k$  is the remaining infectious period of the  $k$ -th member who contracted the disease: if  $\tau_k > 0$ , the individual is still infectious for  $\tau_k$  time units, and recovered otherwise since  $\lfloor \tau_k \rfloor$  time units. For  $k > n - s$ ,  $\tau_k = 0$  by default.

Let us detail the evolution of the structure types over time. Upon an infection event, the household state of the newly infected jumps from  $x = (n, s, \tau)$  to

$$j(x, \sigma) = (n, s - 1, \tau + \sigma \mathbf{e}(n - s + 1)),$$

for some  $\sigma > 0$ , where  $(\mathbf{e}(k), k \in \llbracket 1, n_{\max} \rrbracket)$  is the canonical basis of  $\mathbb{R}^{n_{\max}}$ . Indeed, there is one less susceptible, and the newly infected is the  $(n - s + 1)$ -th member of the households who becomes infected. The corresponding component of  $\tau$  is thus initialized at  $\sigma$  sampled from the distribution  $\nu$ . Simultaneously, the newly infected's workplace state changes analogously, using the same realization  $\sigma$  of the infectious period length.

Such an infection event can be due to a contamination within a household ( $X = H$ ) or workplace ( $X = W$ ) in state  $(n, s, \tau)$ , at rate  $\lambda_X s \sum_{k=1}^{n-s} \mathbf{1}_{\{\tau_k > 0\}}$  where we recall that  $\lambda_H$  and  $\lambda_W$  are the one-to-one contact rates within households and workplaces, respectively. The third source of infection corresponds to the general population, which is assumed to



be uniformly mixing with one-to-all contact rate  $\beta_G$ . At this level, infections occur at rate  $\beta_G SI/K$ , where  $S$  and  $I$  designate the number of susceptibles and infected in the population which can be deduced from the household or workplace states.

Finally, within a structure of type  $(n, s, \tau)$ , the remaining infectious period lengths of infected and recovered individuals decrease over time, according to the simple differential equation  $\tau'_k = -\mathbf{1}_{\{k \leq n-s\}}$ . This leads to a closed system of Markovian dynamics which we describe below.

Let  $K_H$  and  $K_W$  designate the number of households and workplaces, and  $x_k^X(t) = (n_k^X, s_k^X(t), \tau_k^X(t))$  for  $k \in \llbracket 1, K_X \rrbracket$  the type of the  $k$ -th household ( $X = H$ ) or workplace ( $X = W$ ) at time  $t$ . Our process of interest  $\zeta^K = (\zeta^{H|K}, \zeta^{W|K})$  then can be described as follows, for  $X \in \{H, W\}$  and  $t \geq 0$ :

$$\zeta_t^{X|K} = \frac{1}{K_X} \sum_{k=1}^{K_X} \delta_{x_k^X(t)}.$$

In words,  $\zeta^{H|K}$  and  $\zeta^{W|K}$  keep track of the distributions of household and workplace types over time. A rigorous definition of the process  $\zeta^K$  can be provided, as it is characterized as the unique solution to a system of stochastic differential equations driven by Poisson point processes (see Proposition 3.2.1).

Designate by  $E$  the structures' state space, *i.e.*

$$E = \{(n, s, \tau) \in \llbracket 1, n_{\max} \rrbracket \times \llbracket 0, n_{\max} \rrbracket \times \mathbb{R}^{n_{\max}} : s \leq n; \forall j > n-s, \tau_j = 0\}.$$

Let  $\mathcal{M}_1(E)$  be the space of probability measures on  $E$ . For  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$ ,  $t \geq 0$  and  $x \in E$ , let  $f_t(x) = f(t, x)$  and

$$f_t^{\mathcal{I}}(x) = \langle \nu, f_t(j(x, \cdot)) \rangle = \int_0^{+\infty} f(t, j(x, \sigma)) \nu(d\sigma).$$

In order to study the large population convergence of  $(\zeta^K)_{K \geq 1}$ , we make use of a tightness-identification-uniqueness argument. For this purpose, the focus lies on the average of functions  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$  with respect to  $\zeta_T^{X|K}$ , namely

$$\langle \zeta_T^{X|K}, f_T \rangle = \int_E f_T(x) \zeta_T^{X|K}(dx) = \sum_{k=1}^{K_X} f_T(x_k^X(T)).$$

Let us now introduce our convergence result. It requires some technical assumptions on the sequence of initial conditions  $(\zeta_0^K)_{K \geq 1}$ , which however are not very restrictive. As a consequence, they are not stated here and we refer to Assumption 3.3.1 for details.

Introduce the differential operator  $\mathcal{A}$  defined by

$$\forall x = (n, s, \tau) \in E, \quad \mathcal{A}f_t(x) = \partial_t f(t, x) - \sum_{k=1}^{n-s} \partial_{\tau_k} f(t, x).$$

Also, for any  $x = (n, s, \tau) \in E$ , let  $\mathbf{n}(x) = n$ ,  $\mathbf{s}(x) = s$  and  $\mathbf{i}(x) = i(\tau) = \sum_{k=1}^{n-s} \mathbf{1}_{\{\tau_k > 0\}}$  be the functions yielding the size, number of susceptible and number of infected members of a structure in state  $x$ . Further, for  $X \in \{H, W\}$ ,  $\bar{X}$  designates the opposite kind of structure, *i.e.*  $\bar{X} = H\mathbf{1}_{\{X=W\}} + W\mathbf{1}_{\{X=H\}}$ .

We are now ready to state our result.

**Theorem 1.3.1.** *Suppose that  $(\zeta_0^K)_{K \geq 1}$  satisfies Assumption 3.3.1 and converges in law to  $\eta_0 \in \mathcal{M}_1(E)^2$ . Then  $(\zeta^K)_{K \geq 1}$  converges in  $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_1(E))^2$  to  $\eta = (\eta^H, \eta^W)$  defined as the unique solution of the following system of Equations (1.6). For any  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$ , for any  $T \geq 0$  and  $X \in \{H, W\}$ ,*

$$\begin{aligned} \langle \eta_T^X, f_T \rangle &= \langle \eta_0^X, f_0 \rangle + \int_0^T \langle \eta_t^X, \mathcal{A}f_t \rangle dt + \lambda_X \int_0^T \langle \eta_t^X, \mathbf{si}(f_t^I - f_t) \rangle dt \\ &+ \lambda_{\bar{X}} \int_0^T \frac{\langle \eta_t^{\bar{X}}, \mathbf{si} \rangle}{\langle \eta_t^{\bar{X}}, \mathbf{s} \rangle} \langle \eta_t^X, \mathbf{s}(f_t^I - f_t) \rangle dt + \beta_G \int_0^T \frac{\langle \eta_t^H, \mathbf{i} \rangle}{\langle \eta_0^H, \mathbf{n} \rangle} \langle \eta_t^X, \mathbf{s}(f_t^I - f_t) \rangle dt. \end{aligned} \quad (1.6)$$

Additionally, it is possible to show that the marginals of  $\eta^X$ , conditioned on the structure's size  $n$  and number of susceptible members  $s$ , are absolutely continuous. Let  $\rho_{X,n,s}$  designate the associated densities, for  $X \in \{H, W\}$ ,  $n \in \llbracket 1, n_{\max} \rrbracket$  and  $s \in \llbracket 1, n \rrbracket$ . Let

$$\Lambda_X(t) = \lambda_{\bar{X}} \frac{\langle \eta_t^{\bar{X}}, \mathbf{si} \rangle}{\langle \eta_t^{\bar{X}}, \mathbf{s} \rangle} + \beta_G \frac{\langle \eta_t^H, \mathbf{i} \rangle}{\langle \eta_0^H, \mathbf{n} \rangle}.$$

Then the set of functions  $(\rho_{X,n,s})$  is a weak solution to the following differential equations: for any  $\tau \in \mathbb{R}^{n-s}$  and  $t \in (0, T)$ ,

$$\begin{aligned} \partial_t \rho_{X,n,s}(t, \tau) - \sum_{k=1}^{n-s} \partial_{\tau_k} \rho_{X,n,s}(t, \tau) &= -s(\lambda_X i(\tau) + \Lambda_X(t)) \rho_{X,n,s}(t, \tau) \\ &+ \mathbf{1}_{\{s+1 \leq n\}} (s+1) (\lambda_X i(\tau_{1,n-s-1}) + \Lambda_X(t)) \rho_{X,n,s+1}(t, \tau_{1,n-s-1}) g_\nu(\tau_{n-s}), \end{aligned}$$

where  $\tau_{1,n-s-1} = (\tau_1, \dots, \tau_{n-s-1})$  and with appropriate initial conditions. This relates the measure-valued Equation (1.6) to a system of non-linear and nonlocal transport equations.

To summarize, Theorem 1.3.1 yields the exact epidemic dynamics in the large population limit. The limiting object  $\eta$  is rich, as it conveys the epidemic state of households and workplaces, with details on the remaining infectious periods of each individual who has been in contact with the disease. As a consequence,  $\eta$  remains of infinite dimension: we would thus like to achieve a stronger, finite dimensional model reduction, by considering a coarser description of the population's epidemic state.

Such a stronger reduced model can be obtained in the Markovian case, where  $\nu$  is exponentially distributed of parameter  $\gamma$ . Let  $s(t)$  and  $i(t)$  be the proportions of susceptible and infectious individuals, respectively, in the population at time  $t$  according to distribution  $\eta_t$ . Further introduce the set

$$\mathbb{S} = \{(n-i, i) : 2 \leq n \leq n_{\max}, 0 \leq i \leq n-1\}.$$

For  $(S, I) \in \mathbb{S}$  and  $X \in \{H, W\}$ , let  $n_{S,I}^X(t)$  be the proportion of households ( $X = H$ ) or workplaces ( $X = W$ ) containing  $S$  susceptible and  $I$  infected individuals at time  $t$ , as given by  $\eta_t^X$ . Let us further introduce the following applications:

$$\tau_G(t) = \beta_G i(t), \text{ and } \tau_X(t) = \frac{\lambda_X}{\sum_{k \geq 1} k \pi_k^X} \sum_{(S,I) \in \mathbb{S}} SI n_{S,I}^X(t) \text{ for } X \in \{H, W\}.$$

Suppose that, at time 0, the remaining infectious period of each infected individual is supposed to be distributed according to  $\nu$ , independently from one another, and that a fraction  $\varepsilon$  of uniformly chosen individuals are infected amidst an otherwise susceptible population. Then the following result holds.

**Theorem 1.3.2.** *The functions  $(s, i, n_{S,I}^X : X \in \{H, W\}, (S, I) \in \mathbb{S})$  are characterized as being the unique solution of the following dynamical system: for any  $t \geq 0$ ,  $X \in \{H, W\}$  and  $(S, I) \in \mathbb{S}$ ,*

$$\begin{aligned} \frac{d}{dt}s(t) &= -(\tau_H(t) + \tau_W(t) + \tau_G(t)s(t)), \\ \frac{d}{dt}i(t) &= -\frac{d}{dt}s(t) - \gamma i(t), \\ \frac{d}{dt}n_{S,I}^X(t) &= -\left(\lambda_X SI + \tau_{\bar{X}}(t)\frac{S}{s(t)} + \tau_G(t)S + \gamma I\right)n_{S,I}^X(t) \\ &\quad + \gamma(I+1)n_{S,I+1}^X(t)\mathbf{1}_{\{S+I < n_{\max}\}} \\ &\quad + \left(\lambda_X(S+1)(I-1) + \tau_{\bar{X}}(t)\frac{S+1}{s(t)} + \tau_G(t)(S+1)\right)n_{S+1,I-1}^X(t)\mathbf{1}_{\{I \geq 1\}}, \end{aligned} \quad (1.7)$$

with appropriate initial conditions, detailed in Theorem 3.3.3.

Notice that dynamical system (1.7) describes a compartmental model, whose variables include the proportions of households and workplaces having a given number of susceptible and infected members. In particular, this implies that in the Markovian case, the correlation of structure states due to common infected members does not need to be taken into account to derive the correct epidemic dynamics. This allows to obtain a finite-dimensional reduced model.

Nevertheless, the dynamical system's dimension is of order  $O(n_{\max}^2)$ . While this is advantageous when compared to a household-workplace EBCM with  $O(n_{\max}^3)$  equations (Section 3.3.2), it raises the question whether numerically solving the dynamical system is pertinent compared to stochastic simulations. Numerical explorations suggest that for large parts of the parameter space, solving dynamical system (1.7) is significantly faster than the simulation of a single stochastic trajectory of the original epidemic process. This indicates that our obtained reduced model can be a pertinent choice for simulation studies.

### 1.3.3 Chapter 4 - Sensitivity of the reduction accuracy to network and epidemic parameters

The aim of this chapter is to investigate and quantify the impact of both network and epidemic parameters on the accuracy of the two reduced models introduced in the previous chapters. The motivations behind this study are twofold. First, it allows to identify the conditions under which using either reduced model yields a good approximation of the epidemic. Second, for each reduction, it points out parameters which strongly influence the approximation quality: the reduction would thus benefit from precise inference of these parameters, highlighting where estimation effort needs to be invested.

Let us recall here that our two model reductions consist in complementary approaches. On the one hand, the uniformly mixing reduced model from Chapter 2 is very parsimonious,

robust to model variations, but there are no theoretical guarantees on its accuracy. On the other hand, the large population limit of Chapter 3 is certain to be asymptotically exact, but only under the assumption that the contact network is exactly given by the household-workplace model. As a consequence, our sensitivity study explores two different angles, depending on the strengths and weaknesses of each reduced model.

First, let us consider the uniformly mixing reduced model of Chapter 2. Given the absence of theoretical guarantees on its accuracy, we previously have explored numerically the parameter space to identify regions where the reduction is pertinent. Here, we complement this approach with a global sensitivity analysis based on Sobol's variance decomposition. This allows to quantify the parameters' influence, and check if these results are consistent with previous conclusions from Chapter 2.

For this purpose, we focus on the household-workplace model, whose parameters are the structure size distributions  $(\pi^H, \pi^W)$ , contact rates  $(\beta_G, \lambda_H, \lambda_W)$  and recovery rate  $\gamma$ , as explained in Section 1.1.2. Here, we explore the impact of the contact rates only, and keep fixed all other parameters. We conceive an experimental design allowing to cover a pertinent domain of the contact rate parameter space, based on the epidemic scenarios of Chapter 2. We then numerically compute Sobol's main and total indices for our quantities of interest, namely the differences of the peak or final size predicted by the uniformly mixing reduced model of Chapter 2, and by the large population limit of Chapter 3.

The obtained results are qualitatively very different for peak and final size. For the peak size, all three contact rates have equally significant impact on the prediction accuracy of the uniformly mixing reduced model, with almost no interaction effects. For the final size, however, the contact rate within the general population is responsible for a large majority of the output variance, with some minor interaction effects. Notably, this seems coherent with the results from Chapter 2, which indicated that the final size prediction performs better when the proportion of infections in the general population is high.

Second, we focus on the reduced model derived in Chapter 3 as the large population limit of the household-workplace model, thus yielding asymptotically exact dynamics. However, this result relies on the choice of the contact network which contains two strong assumptions: (1) the independence of structure choices of all individuals, and (2) uniform mixing within all structures. Notably, both are simplifications of real-life settings: for example, couples may share a workplace, and empirical studies are not consistent with workplaces being uniformly mixing. But, if they are relaxed, does the large population limit of Chapter 3 still give a good approximation of an epidemic spreading on that contact network?

In order to address this question, we propose a generalized household-workplace model which in general does not comply to assumptions (1) and (2). As previously, the (asymptotic) household and workplace size distributions are given by  $\pi^H$  and  $\pi^W$ , respectively. Intuitively, the first difference with the household-workplace model is that within each household, several individuals choose their workplace together, each household member participating in that choice with probability  $q$ . Additionally, for workplaces of size  $n$ , the within-workplace contact network takes the form of an Erdős-Rényi graph  $G(n, \rho(n))$ , where  $n \mapsto \rho(n)$  is a non-increasing function, which attains its minimal value  $p$  at  $n = n_{\max} < \infty$ . In particular, the household-workplace model is recovered if  $p = 1$  and  $q = 0$ . Throughout the following, we vary the within-workplace density  $p$  and structure overlap  $q$ .

Regarding epidemic spread, we consider *SIR* dynamics, as in the previous chapters. The parametrization remains the same, with contact rates  $\beta_G, \lambda_H, \lambda_W$  specific to each layer and recovery rate  $\gamma > 0$ . The only difference lies in the interpretation of the contact rate  $\lambda_W$ , which now represents the rate at which contacts occurs along each edge of the within-workplace network. In particular, if the within-workplace contact network is a complete graph and contains  $s$  susceptible and  $i$  infected members, this still yields infections at rate  $\lambda_W si$ , as expected.

We concentrate on the differences in the predictions of either peak or final size, between the generalized household-workplace model and the large population limit of Chapter 3. These differences are computed numerically, covering several epidemic parameter sets, structure size distributions and values of  $(p, q)$ . Our findings suggest that for a significant part of the explored parameter space, the large population limit gives a satisfying prediction (within an error margin of roughly 5%). Further, the within-workplace density  $p$  has a stronger impact than structure overlap  $q$ . Hence, taking into account a more detailed description of within-workplace contacts would be worth investigating.

These results are ongoing work. In particular, a wider sensitivity study is planned, which allows to assess the robustness of our present conclusions. These perspectives are detailed at the end of Chapter 4.

### 1.3.4 Chapter 5 - On spinal constructions for interacting populations

The previous chapters have allowed to propose reduced models which give access to macroscopic properties of the household-workplace model, such as the proportion of infected in the population, or the proportion of households containing a given number of susceptible and infected members. However, they are not designed for capturing microscopic aspects, such as chains of infections, which yield a detailed understanding of how the epidemic spreads in the network.

In order to track contamination chains, an analogy can be made between the tree of infections and a genealogical tree: infection events correspond to births, and recovery events to deaths. As a consequence, chains of infections then amount to ancestral lineages. While the study of such lineages is classical for the branching approximation which holds at the beginning of the epidemic, understanding them outside of the branching regime is an active field of research (Bansaye, 2024; Calvez et al., 2022; Duchamps et al., 2023; Medous, 2023).

Throughout this chapter, we thus consider the general setting of stochastic, density-dependent population processes, as our results are not restricted to epidemic models since such dynamics arise in other domains, including ecology. Here, we make some simplifying assumptions, notably that the individual's type space is finite. Indeed, the household-workplace model points out interesting difficulties, both technical and due to the representation of the process allowing to make emerge the infection tree, which are motivating perspectives for a future work. However, our results already open the door to pertinent applications in epidemiology, including models with contact heterogeneity, as emphasized in the chapter's discussion.

More precisely, the aim of this work is to study ancestral lineages of an individual which is sampled at some time  $t > 0$  in a population. For branching processes, such a typical lineage reduces to a Markov process, thanks to the many-to-one formula (Harris et al., 2016; Harris and Roberts, 2017), which is related to Feynman-Kac path equations (Del Moral,

2004, Sections 1.3 and 1.4.4). This so-called spinal construction greatly simplifies the study of typical lineages, both on the theoretical side as the reduced model is a Markov process, and from the numerical point of view as simulations are less demanding in computation time.

Recently, Bansaye, 2024 has developed a spinal process for multi-type populations with interaction, *i.e.* where reproduction rates depend on the population state. In this case, the author focuses on the lineage of an individual sampled at time  $t > 0$  with probability proportional to  $\psi(x, \mathbf{z})$ , where  $x$  is the individual's type,  $\mathbf{z}$  the current population type composition, and  $\psi$  a positive function on the appropriate space. A many-to-one formula relating such a lineage to a  $\psi$ -spinal process is established, both in finite populations and in the large population limit.

Here, we address some of the drawbacks and open questions of this approach. Namely, we first propose an alternative many-to-one formula, relating a  $\psi$ -sampled lineage to a time-inhomogeneous Markov process. This allows for a larger class of sampling strategies  $\psi$ , and can be easier to interpret and potentially numerically advantageous compared to the  $\psi$ -spine. Second, we focus on quantifying the convergence of the  $\psi$ -spine to its large population limit.

In order to introduce our time-inhomogeneous spinal construction, let us give a brief description of the population process. We consider a population whose size is bounded by  $K < \infty$ . Each individual is characterized by a type  $x \in \mathcal{X}$ , where  $\mathcal{X}$  is a finite set. The population is described by the vector  $\mathbf{z} = (\mathbf{z}_x, x \in \mathcal{X})$  belonging to the set  $\mathcal{Z}_K = \{\mathbf{z} \in (\mathbb{N} \cup \{0\})^{\text{Card}(\mathcal{X})} : \|\mathbf{z}\|_1 \leq K\}$ . Here,  $\mathbf{z}_x$  counts the number of individuals of type  $x$ . In a population of composition  $\mathbf{z}$ , each individual of type  $x$  dies and leaves offspring  $\mathbf{k} \in \mathcal{Z}_K$  at rate  $\tau_{\mathbf{k}}(x, \mathbf{z})$ . This means that the population composition jumps from  $\mathbf{z}$  to  $\mathbf{z} + \mathbf{k} - \mathbf{e}(x)$ , where  $(\mathbf{e}(x), x \in \mathcal{X})$  is the canonical basis of  $\mathcal{Z}_K$ . The rate  $\tau_{\mathbf{k}}(x, \mathbf{z})$  is supposed to be zero if the resulting population exceeds  $K$  individuals.

We let  $\mathbb{G}(t)$  be the set of individuals alive at time  $t$ , write  $u \geq v$  if individual  $u$  is a descendant of  $v$  and  $x_u(s)$  for the type of the unique ancestor of  $u$  alive at time  $s \leq t$ . The population process of interest is then given by  $X(t) = \sum_{u \in \mathbb{G}(t)} \delta_{(u, x_u(t))}$ , and  $Z(t)$  designates the associated population type composition, at time  $t \geq 0$ . Finally, for  $(x, \mathbf{z}) \in \mathcal{S}_K = \{(x, \mathbf{z}) \in \mathcal{X} \times \mathcal{Z}_K : \mathbf{z}_x \geq 1\}$ , we consider  $\mathfrak{X}(x, \mathbf{z})$  to yield an appropriate initial condition for  $X$ , such that the population is of composition  $\mathbf{z}$  with one distinguished individual  $u_x$  of type  $x$ .

We are now ready to give an intuitive description of our spinal process, which is inspired by Marguet, 2019 in continuation with which we refer to it as the  $\psi$ -auxiliary process. Let  $t \geq 0$  be the time at which sampling takes place. Similarly to Bansaye, 2024, our spinal process needs to take into account both the spine's type  $Y^{(t)}$ , and the composition of the population  $\zeta^{(t)}$ , due to the density-dependence of reproduction rates. Both reproduction rates of the spine and outside of it need to be biased, and this bias relies on the application  $m_\psi$  defined as follows. For any  $(x, \mathbf{z}) \in \mathcal{S}_K$ , for any  $s \in [0, t]$ ,

$$m_\psi(x, \mathbf{z}, t - s) = \mathbb{E} \left[ \sum_{u \in \mathbb{G}(t), u \geq u_x} \psi(x_u(t), Z(t)) \middle| X(s) = \mathfrak{X}(x, \mathbf{z}) \right].$$

This corresponds to the  $\psi$ -weighted average of the types of individuals alive at time  $t$ , whose ancestor at time  $s$  was a given individual of type  $x$  in a population of composition  $\mathbf{z}$ . Throughout the following, we require  $m_\psi$  to be a positive function.

Assume now that at time  $s \in [0, t]$ , the population is of composition  $\mathbf{z}$  and the spine of type  $x$ . Then on the one hand, the spine reproduces, leaves descendance  $\mathbf{k} \in \mathcal{Z}_K$  and switches to type  $y$  at rate

$$\tau_{\mathbf{k}}(x, \mathbf{z}) \mathbf{k}_y \frac{m_\psi(y, \mathbf{z} + \mathbf{k} - \mathbf{e}(x), t - s)}{m_\psi(x, \mathbf{z}, t - s)}.$$

In other words, such a transition is more likely if it increases the  $\psi$ -average of the spine's descendance at time  $t$ . Similarly, an individual of type  $y \in \mathcal{X}$  other than the spine leaves descendance  $\mathbf{k}$  at rate

$$\tau_{\mathbf{k}}(y, \mathbf{z}) \frac{m_\psi(x, \mathbf{z} + \mathbf{k} - \mathbf{e}(y), t - s)}{m_\psi(x, \mathbf{z}, t - s)}.$$

Again, this transition rate is positively biased if it yields a more favorable environment for the spine, in the sense that the  $\psi$ -average of the spine's descendance at time  $t$  is greater if at time  $s$ , the population is of composition  $\mathbf{z} + \mathbf{k} - \mathbf{e}(y)$  than if it is of composition  $\mathbf{z}$ . The spinal process  $(Y^{(t)}(s), \zeta^{(t)}(s))_{s \leq t}$  is thus a time-inhomogeneous Markov process, which can be rigorously defined as the unique solution to a system of SDEs (see Proposition 5.3.2).

The heart of the chapter is the following many-to-one formula. We let  $\mathbb{E}_{\mathfrak{X}(x, \mathbf{z})}$  and  $\mathbb{E}_{(x, \mathbf{z})}$  designate respectively the expectation conditionally on the events  $\{X(0) = \mathfrak{X}(x, \mathbf{z})\}$  and  $\{(Y^{(t)}(0), \zeta^{(t)}(0)) = (x, \mathbf{z})\}$ .

**Theorem 1.3.3.** *For any  $t \geq 0$  and any measurable function  $F : \mathbb{D}([0, t], \mathcal{S}_K) \rightarrow \mathbb{R}_+$ , for any  $(x, \mathbf{z}) \in \mathcal{S}_K$ ,*

$$\begin{aligned} \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{u \in \mathbb{G}(t), u \geq u_x} \psi(x_u(t), Z(t)) F((x_u(s), Z(s))_{s \leq t}) \right] \\ = m_\psi(x, \mathbf{z}, t) \mathbb{E}_{(x, \mathbf{z})} [F((Y^{(t)}(s), \zeta^{(t)}(s))_{s \leq t})]. \end{aligned} \quad (1.8)$$

The proof relies on identifying both sides of Equation (1.8) through the infinitesimal generators of the associated time-inhomogeneous, conservative semi-groups. This corresponds to exhibiting a time-inhomogeneous change in probability, inspired by Marguet, 2019, which relates the  $\psi$ -typical lineage of the population process to the  $\psi$ -auxiliary process.

Let us emphasize here that our assumption  $m_\psi > 0$  is less restrictive than  $\psi > 0$ . Notably, it allows to sample directly within subpopulations, which can find pertinent applications. For example, the  $\psi$ -auxiliary process may specifically capture contamination chains leading to infections of vulnerable individuals, such as the elderly, at time  $t$ . We refer to Chapter 5 for a more detailed discussion on strengths and weaknesses of the obtained spinal process.

Second, we consider the  $\psi$ -spine introduced in Bansaye, 2024, and quantify its convergence to its large population limit. We first set the context for this study. Let  $d = \text{Card}(\mathcal{X})$ . For a population of size at most  $K$ , we let  $Z^K(t) \in [0, 1]^d$  be its population state at time  $t$ , defined as its type composition normalized by  $K$ . Further, for any  $\mathbf{k} \in (\mathbb{N} \cup \{0\})^d$ , the reproduction rate  $\tau_{\mathbf{k}} : \mathcal{X} \times [0, 1]^d \rightarrow \mathbb{R}_+$  depends on the reproducing individual's type and the population state only, and the set  $\{\mathbf{k} : \tau_{\mathbf{k}} \neq 0\}$  is finite. Under mild conditions, the sequence of processes  $(Z^K)_{K \geq 1}$  then converges in probability, uniformly on finite time intervals, to a deterministic function  $z : \mathbb{R}_+ \rightarrow [0, 1]^d$  which is solution to a dynamical system.



Consider an application  $\psi : \mathcal{X} \times [0, 1]^d \rightarrow \mathbb{R}_+$  which is continuous on  $\mathcal{X} \times [0, 1]^d$  and continuously differentiable on  $\mathcal{X} \times (0, 1)^d$ . Let  $\psi_K$  be defined by  $\psi_K(x, \mathbf{z}) = \psi(x, \mathbf{z}/K)$  for  $(x, \mathbf{z}) \in \mathcal{S}_K$ . The  $\psi_K$ -spine can be described as follows. Assume that the population is in state  $\mathbf{z} \in [0, 1]^d$ , and that the spine is of type  $x \in \mathcal{X}$ . Then the spine leaves descendance  $\mathbf{k} \in (\mathbb{N} \cup \{0\})^d$  and becomes of type  $y$  at rate

$$\tau_{\mathbf{k}}(x, \mathbf{z}) \mathbf{k}_y \frac{\psi(y, \mathbf{z} + (\mathbf{k} - \mathbf{e}(x))/K)}{\psi(x, \mathbf{z})}.$$

Similarly, an individual of type  $y \in \mathcal{X}$  other than the spine leaves descendance  $\mathbf{k}$  at rate

$$\tau_{\mathbf{k}}(y, \mathbf{z}) \frac{\psi(x, \mathbf{z} + (\mathbf{k} - \mathbf{e}(y))/K)}{\psi(x, \mathbf{z})}.$$

The associated sequence of  $\psi_K$ -spinal processes  $(Y^K, \zeta^K)_{K \geq 1}$ , whose components describe the spine's type and population state, converges in law to  $(\Upsilon, z)$ , where  $\Upsilon$  is a time-inhomogeneous Markov process. More precisely, at time  $s \geq 0$ ,  $\Upsilon$  jumps from  $x \in \mathcal{X}$  to  $y \in \mathcal{X}$  at rate

$$\sum_{\mathbf{k} \in (\mathbb{N} \cup \{0\})^d} \mathbf{k}_y \tau_{\mathbf{k}}(x, z(s)) \frac{\psi(y, z(s))}{\psi(x, z(s))}.$$

Here, we do not go further into detail, and refer to Section 5.2.2 for assumptions and results of Bansaye, 2024.

Notably, the fluctuations of  $Z^K$  around its deterministic limit  $z$  are well understood, classical regimes being gaussian fluctuations, moderate and large deviations (Ethier and Kurtz, 1986, Chapter 11; Britton and Pardoux, 2019b; Pardoux, 2020). Technically, these results do not transfer immediately to  $\zeta^K$  as it distinguishes one individual, the spine, from the general dynamics. Intuitively, however, such control of fluctuations should hold, as the impact of the spine becomes negligible in the large population limit. Here, we thus work under Assumption 5.4.3 that there exist non-negative sequences  $(t_K)_{K \geq 1}$ ,  $(\varepsilon_K)_{K \geq 1}$  and  $(\alpha_K)_{K \geq 1}$  such that, for every  $K \geq 1$ ,

$$\mathbb{P} \left( \sup_{s \in [0, t_K]} \|\zeta^K(s) - z(s)\|_1 \geq \varepsilon_K \right) \leq \alpha_K.$$

The general idea is to transfer this control of the convergence of  $(\zeta^K)_{K \geq 1}$  to  $z$ , to the convergence of  $(Y^K)_{K \geq 1}$  to  $\Upsilon$ . This further requires some regularity of the reproduction rates, which are assumed to be Lipschitz-continuous (Assumption 5.4.2). Finally, a coupling argument leads to the following result:

**Theorem 1.3.4.** *Under Assumptions 5.4.2 and 5.4.3, there exists  $C > 0$  such that for every  $K \geq 1$ ,*

$$\mathbb{P}(\exists t \leq t_K : Y^K(t) \neq \Upsilon(t)) \leq \alpha_K + C(\varepsilon_K + K^{-1})t_K.$$

In particular, if the sequences  $(t_K/K)_{K \geq 1}$ ,  $(t_K \varepsilon_K)_{K \geq 1}$  and  $(\alpha_K)_{K \geq 1}$  tend to zero as  $K$  goes to infinity, with  $(t_K)_{K \geq 1}$  converging to  $t_\infty \in \mathbb{R}_+ \cup \{\infty\}$ , the coupling is asymptotically exact on  $[0, t_\infty]$ . In this regard, a particularly promising regime is that of moderate deviations in presence of an asymptotically stable equilibrium, a setting which emerges for instance in endemic models (Pardoux, 2020; Prodhomme, 2023). We refer to the discussion of Chapter 5 for detail.



## 1.4 Contributions principales de la thèse

Dans cette section, nous introduisons et résumons les résultats des Chapitres 2 à 5. La Figure 1.11 donne une vue d'ensemble des modèles considérés et des connexions entre-eux, indiquant le centre d'attention de chaque chapitre. Notamment, les Chapitres 2 à 4 visent à proposer et analyser des modèles réduits qui approchent le modèle foyer-travail, introduit en Section 1.1.2. Ici, nous rappelons la définition de ses paramètres, car ils seront utilisés dans les sections suivantes :

- $\pi^H$  et  $\pi^W$  désignent les distributions de tailles des foyers et lieux de travail. Nous supposons que la taille maximale  $n_{\max}$  des foyers et lieux de travail est finie.
- $\beta_G$ ,  $\lambda_H$  et  $\lambda_W$  sont les taux de contact au sein de chaque couche du graphe : en population générale, dans les foyers et dans les lieux de travail.
- Les durées des périodes infectieuses sont distribuées selon une mesure de probabilité  $\nu$  sur  $\mathbb{R}_+$ , supposée absolument continue par rapport à la mesure de Lebesgue. Dans le cas Markovien,  $\nu$  est exponentiellement distribuée de paramètre  $\gamma$ , qui correspond alors au taux de guérison.

Nous pouvons maintenant nous intéresser aux contributions de cette thèse.

### 1.4.1 Chapitre 2 - L'empreinte épidémique des structures de contact

*Ce chapitre est basé sur la prépublication "The epidemiological footprint of contact structures in models with two levels of mixing", réalisée en collaboration avec Vincent Bansaye, François Deslandes et Elisabeta Vergu (Bansaye et al., 2023a).*

Les structures de contact telles que les foyers, lieux de travail et écoles sont des composantes naturelles des réseaux de contact humain, qui sont donc agglomérés. Or, il est bien connu que le clustering a un impact significatif sur l'épidémie (House and Keeling, 2011), et que cet effet dépend de la façon dont le clustering est implémenté au sein du réseau (Volz et al., 2011). De plus, si l'on cherche à analyser certaines mesures de contrôle comme le télétravail ou la fermeture des écoles, Di Lauro et al., 2021 ont montré que la façon dont les contacts sont représentés par le modèle mathématique modifie les prédictions quant à l'effet de ces interventions. Puisque le télétravail agit en modifiant la distribution des tailles de lieux de travail, la question suivante se pose donc : quel est l'impact sur l'épidémie de la distribution des tailles de sous-groupes fortement connectés, tels que les lieux de travail ?

Étant donné l'impact du choix de la modélisation des contacts sur l'épidémie et l'effet des mesures de contrôle, cette question sera abordée en étudiant le modèle foyer-travail, puisqu'il distingue explicitement les contacts au sein des foyers et lieux de travail. Comme illustré par la suite, l'impact épidémique de telles petites structures de contact est significatif mais subtile, n'admettant pas d'expression analytique immédiate. Par conséquent, nous proposons deux indicateurs pour cet impact épidémique qui s'avéreront liés à deux autres enjeux majeurs : la conception de stratégies de télétravail efficaces, et la réduction du modèle.

Pour cela, nous avons étudié numériquement le modèle foyer-travail. Nous explorons par simulations stochastiques plusieurs jeux de paramètres épidémiques, inspirés de la COVID-19, de la grippe et de la varicelle. Ces scénarios sont combinés avec une variété de distributions

de tailles de lieux de travail, parmi lesquelles une distribution de référence basée sur les tailles de lieux de travail en Ile-de-France (INSEE, 2018). Au contraire, la distribution des tailles de foyers est fixée, car elle n'est généralement pas modifiée par les mesures de contrôle, et correspond à celle observée en France (INSEE, 2018).

Nous commençons par comparer deux stratégies de télétravail. La première consiste en une approche naïve, où le nombre d'employés autorisés en présentiel est proportionnel à la taille du lieu de travail (stratégie *linéaire*), tandis que ce nombre est proportionnel à la racine carrée de la taille du lieu de travail pour la seconde stratégie (stratégie *sous-linéaire*). Nous observons que la stratégie sous-linéaire est la plus performante des deux : à taux de télétravail fixé, la réduction de la taille du pique et de la taille finale est plus forte si on utilise la stratégie sous-linéaire au lieu de celle linéaire. Remarquons ici qu'à taux de télétravail fixé, le nombre moyen d'employés en présentiel par lieu de travail est fixé également, puisque le nombre de lieux de travail est constant par essence.

Ainsi, nous cherchons à proposer un indicateur de l'impact épidémique de la distribution des tailles de lieux de travail  $\pi^W$ , à moyenne fixée. Par simulation, nous montrons qu'en ce cas, le taux de croissance épidémique, la taille du pique et la taille finale sont linéairement corrélés à la variance de  $\pi^W$ , qui correspond donc à notre premier indicateur. En particulier, la stratégie de télétravail sous-linéaire induit par construction une réduction plus forte de la variance du nombre d'employés en présentiel que la stratégie linéaire. Cela illustre la pertinence de tels indicateurs, qui peuvent être mis en pratique pour concevoir des stratégies de contrôle efficaces.

Néanmoins, de telles explorations numériques présentent l'inconvénient que les simulations du modèle stochastique foyer-travail sont gourmandes en temps de calcul, en particulier pour de grandes populations. De surcroît, le modèle foyer-travail peut être délicat à calibrer en pratique, car il requière une bonne connaissance à la fois des paramètres du réseau et des paramètres épidémiques. Nous visons donc ensuite à développer un modèle réduit parcimonieux qui approche la dynamique du modèle foyer-travail. Plus précisément, nous proposons d'utiliser un modèle *SIR* déterministe, uniformément mélangeant, qui ne dépend que de deux paramètres : un taux de guérison  $\gamma$  et un taux de contact réduit  $\beta$ . Nous supposons que le taux de guérison  $\gamma$  est connu, ce qui est généralement le cas par expertise médicale, et il ne reste donc qu'à déterminer le taux de contact.

Deux façons naturelles de procéder seraient de choisir  $\beta$  de sorte que le modèle réduit et le modèle initial partagent soit le même nombre de reproduction, soit le même taux de croissance épidémique. D'une part, contrairement au cas uniformément mélangeant, il existe plusieurs nombres de reproduction pour le modèle foyer-travail (Ball et al., 2016). D'autre part, le taux de croissance épidémique est caractérisé pour notre modèle comme l'unique solution d'une équation implicite (Pellis et al., 2011), pour laquelle nous développons des formules partiellement plus explicites. Par conséquent, nous choisissons de calibrer le modèle réduit en utilisant le taux de croissance épidémique  $r$ , menant au système d'équations suivant, qui décrit les dynamiques des proportions de susceptibles  $s$  et infectés  $i$  :

$$s'(t) = -(\gamma + r)si, \quad i'(t) = (\gamma + r)si - \gamma i.$$

Afin d'évaluer la performance de ce modèle réduit, nous calculons par simulations l'erreur d'approximation pour la taille du pique et la taille finale, en considérant une variété de

scénarios épidémiques et de distributions de tailles de lieux de travail. Dans la plupart des cas, l'erreur commise est inférieure à 5%, ce qui est satisfaisant. De plus, nous remarquons que le modèle réduit approche mieux la taille du pique que la taille finale, ce qui est cohérent avec le fait qu'il ait été calibré sur la phase initiale de croissance exponentielle. Nous avons également évalué la robustesse de cette approche à diverses variations du modèle, à savoir en considérant un modèle *SEIR*, des périodes infectieuses distribuées selon une loi Gamma (*i.e.* sortant du cadre Markovien) ou encore des taux d'infection sous-linéaires au sein des structures. Dans tous les cas, l'approximation est raisonnable.

Cette réduction de modèle a plusieurs points forts. Elle est élégamment parcimonieuse, et relativement aisée à calibrer à partir de données, puisqu'elle ne repose que sur les taux de guérison et de croissance épidémique. En particulier, cela est avantageux en comparaison avec del Valle Rafo et al., 2021. Par ailleurs, ces résultats établissent que le taux de croissance épidémique est un bon indicateur de l'impact épidémique des distributions des tailles de structures  $\pi^H$  et  $\pi^W$ .

Néanmoins, il y a aussi quelques inconvénients. En effet, comme mentionné précédemment, la précision de ce modèle réduit décroît au cours du temps, et il n'y a ni garanties théoriques sur la qualité de l'approximation, ni quantifications analytiques de l'erreur commise. Cela motive le développement d'un autre modèle réduit, qui complète cette première approche. Ce sera l'enjeu du chapitre suivant.

### 1.4.2 Chapitre 3 - Limite grande population d'un modèle *SIR* multicouche incluant foyers et lieux de travail

*Ce chapitre est basé sur la prépublication "Large population limit for a multilayer SIR model including households and workplaces" (Kubasch, 2023).*

Nous étudions la limite grande population du modèle foyer-travail, ce qui nous permet de dériver un modèle réduit correspondant aux dynamiques épidémiques exactes en population infinie. Nous commençons par formaliser l'épidémie en population de taille finie  $K$  en tant que processus de Markov à valeur mesure. En effet, cela correspond à un cadre mathématique favorable pour l'étude de la limite grande population grâce au problème de martingale associé (*e.g.* Fournier and Méléard, 2004). Néanmoins, dans notre cas, cela n'est pas immédiat car les individus infectés corrént l'état épidémique de leur foyer et lieu de travail. Cela est illustré par les événements de guérison, puisqu'un infecté doit guérir simultanément dans les deux structures auxquelles il appartient. Par conséquent, la description intuitive des foyers et lieux de travail par le nombre de membres susceptibles, infectés ou guéris se révèle insuffisante pour représenter correctement la dynamique épidémique.

Pour y remédier, nous proposons d'incorporer à la description des structures la durée restante des périodes infectieuses des individus ayant été en contact avec la maladie, des approches similaires existant pour l'approximation branchement (Ball et al., 2014). En particulier, nous obtenons donc des dynamiques semblables aux modèles structurés en âge (Tran, 2006), ce qui nous permet de considérer une large gamme de distributions de la durée des périodes infectieuses tout en restant dans un cadre Markovien.

Remarquons ici que nous sommes en présence de deux sources d'aléa : la première correspond à la réalisation du réseau de contact, et la seconde à la propagation du processus épidémique. Notre modèle de graphe aléatoire, dont la construction est détaillée au Chapitre 3, assure que chaque individu choisit son foyer et son lieu de travail indépendamment l'un de

l'autre, et des autres individus. De plus, la suite obtenue de réseaux de contact en population de taille  $K \geq 1$  converge presque sûrement vers un réseau foyer-travail de taille infinie, où les distributions de tailles des foyers et lieux de travail sont données précisément par  $\pi^H$  et  $\pi^W$ , respectivement. Ici, nous considérons une réalisation donnée  $(\mathbf{G}^K)_{K \geq 1}$  de cette suite pour laquelle cette convergence a lieu, et nous ne soulignerons donc plus l'aléa lié au réseau de contact.

Étant donnée cette réalisation du réseau de contact pour une population de taille  $K$ , nous pouvons maintenant décrire le processus épidémique par un modèle basé sur agents, dont les agents sont les foyers et lieux de travail. Chaque structure est caractérisée par sa taille  $n$ , le nombre de membres susceptibles  $s$  et un vecteur  $\tau \in \mathbb{R}^{n_{\max}}$  qui représente les membres infectés et guéris. Plus précisément, pour  $k \leq n - s$ ,  $\tau_k$  est la durée restante de la période infectieuse du  $k$ -ème membre ayant contracté la maladie : si  $\tau_k > 0$ , l'individu reste infectieux pour  $\tau_k$  unités de temps, sinon il est guéri depuis  $|\tau_k|$  unités de temps. Pour  $k > n - s$ ,  $\tau_k = 0$  par défaut.

Détaillons maintenant l'évolution des types de structures au cours du temps. Suite à un évènement d'infection, le type du foyer du nouvel infecté saute de  $x = (n, s, \tau)$  à

$$j(x, \sigma) = (n, s - 1, \tau + \sigma \mathbf{e}(n - s + 1)),$$

pour  $\sigma > 0$ ,  $(\mathbf{e}(k), k \in \llbracket 1, n_{\max} \rrbracket)$  étant la base canonique de  $\mathbb{R}^{n_{\max}}$ . En effet, il y a un susceptible de moins, et le nouvel infecté est le  $(n - s + 1)$ -ème membre du foyer à être contaminé. La composante correspondante de  $\tau$  est donc initialisée à une valeur  $\sigma$  distribuée selon  $\nu$ . Simultanément, le lieu de travail du nouvel infecté change de type de façon analogue, en utilisant la même réalisation  $\sigma$  de la durée de période infectieuse.

Un tel évènement d'infection peut faire suite à une contamination au sein d'un foyer ( $X = H$ ) ou lieu de travail ( $X = W$ ) de type  $(n, s, \tau)$ , au taux  $\lambda_X s \sum_{k=1}^{n-s} \mathbf{1}_{\{\tau_k > 0\}}$  où nous rappelons que  $\lambda_H$  et  $\lambda_W$  désignent respectivement les taux de contact un-pour-un au sein des foyers et lieux de travail. La troisième source d'infection est la population générale, uniformément mélangeante de taux de contact un-pour-tous  $\beta_G$ . Au sein de celle-ci, une infection se produit donc au taux  $\beta_G SI/K$ , où  $S$  et  $I$  correspondent au nombre de susceptibles et infectés dans la population qui peuvent être déduits des types des foyers et lieux de travail.

Enfin, au sein d'une structure de type  $(n, s, \tau)$ , la durée restante des périodes infectieuses des membres infectés ou guéris décroît au cours du temps, suivant l'équation différentielle  $\tau'_k = -\mathbf{1}_{\{k \leq n-s\}}$ . Ainsi, nous obtenons un système fermé de dynamiques Markoviennes, décrit ci-dessous.

Soient  $K_H$  et  $K_W$  les nombres de foyers et lieux de travail, et  $x_k^X(t) = (n_k^X, s_k^X(t), \tau_k^X(t))$  pour  $k \in \llbracket 1, K_X \rrbracket$  le type du  $k$ -ème foyer ( $X = H$ ) ou lieu de travail ( $X = W$ ) au temps  $t$ . Notre processus d'intérêt  $\zeta^K = (\zeta^{H|K}, \zeta^{W|K})$  peut alors être décrit comme suit, pour  $X \in \{H, W\}$  et  $t \geq 0$  :

$$\zeta_t^{X|K} = \frac{1}{K_X} \sum_{k=1}^{K_X} \delta_{x_k^X(t)}.$$

Intuitivement,  $\zeta^{H|K}$  et  $\zeta^{W|K}$  correspondent aux distributions de types de foyers et lieux de travail, au cours du temps. Une définition rigoureuse du processus  $\zeta^K$  peut être obtenue en tant qu'unique solution d'un système d'équations différentielles stochastiques par rapport à des processus ponctuels de Poisson (voir Proposition 3.2.1).

Désignons par  $E$  l'espace de types des structures, *i.e.*

$$E = \{(n, s, \tau) \in \llbracket 1, n_{\max} \rrbracket \times \llbracket 0, n_{\max} \rrbracket \times \mathbb{R}^{n_{\max}} : s \leq n; \forall j > n - s, \tau_j = 0\}.$$

Soit  $\mathcal{M}_1(E)$  l'espace de mesures de probabilité sur  $E$ . Pour  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$ ,  $t \geq 0$  et  $x \in E$ , soit  $f_t(x) = f(t, x)$  et

$$f_t^{\mathcal{I}}(x) = \langle \nu, f_t(j(x, \cdot)) \rangle = \int_0^{+\infty} f(t, j(x, \sigma)) \nu(d\sigma).$$

Afin d'étudier la convergence de  $(\zeta^K)_{K \geq 1}$  en grande population, nous utilisons un argument de tension-identification-unicité. Pour cela, nous nous concentrons sur la moyenne de fonctions  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$  par rapport à  $\zeta_T^{X|K}$ , c'est-à-dire

$$\langle \zeta_T^{X|K}, f_T \rangle = \int_E f_T(x) \zeta_T^{X|K}(dx) = \sum_{k=1}^{K_X} f_T(x_k^X(T)).$$

Introduisons maintenant notre résultat de convergence. Celui-ci nécessite quelques hypothèses techniques sur la suite de conditions initiales  $(\zeta_0^K)_{K \geq 1}$ , qui ne sont toutefois pas très restrictives. Par conséquent, nous ne les détaillons pas ici et renvoyons à l'Hypothèse 3.3.1.

Considérons l'opérateur différentiel  $\mathcal{A}$  défini par

$$\forall x = (n, s, \tau) \in E, \quad \mathcal{A}f_t(x) = \partial_t f(t, x) - \sum_{k=1}^{n-s} \partial_{\tau_k} f(t, x).$$

Alors, pour tout  $x = (n, s, \tau) \in E$ , soient  $\mathbf{n}(x) = n$ ,  $\mathbf{s}(x) = s$  et  $\mathbf{i}(x) = i(\tau) = \sum_{k=1}^{n-s} \mathbf{1}_{\{\tau_k > 0\}}$  les fonctions qui à une structure de type  $x$ , associent sa taille, le nombre de membres susceptibles et le nombre de membres infectés. De plus, pour  $X \in \{H, W\}$ ,  $\bar{X}$  désigne l'autre sorte de structure, *i.e.*  $\bar{X} = H\mathbf{1}_{\{X=W\}} + W\mathbf{1}_{\{X=H\}}$ .

Nous pouvons désormais énoncer notre résultat.

**Theorem 1.4.1.** *Supposons que  $(\zeta_0^K)_{K \geq 1}$  satisfait l'Hypothèse 3.3.1 et converge en loi vers  $\eta_0 \in \mathcal{M}_1(E)^2$ . Alors  $(\zeta^K)_{K \geq 1}$  converge dans  $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_1(E))^2$  vers  $\eta = (\eta^H, \eta^W)$  défini comme l'unique solution du système suivant d'Équations (1.9). Pour tout  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$ , pour tout  $T \geq 0$  et  $X \in \{H, W\}$ ,*

$$\begin{aligned} \langle \eta_T^X, f_T \rangle &= \langle \eta_0^X, f_0 \rangle + \int_0^T \langle \eta_t^X, \mathcal{A}f_t \rangle dt + \lambda_X \int_0^T \langle \eta_t^X, \mathbf{si}(f_t^{\mathcal{I}} - f_t) \rangle dt \\ &+ \lambda_{\bar{X}} \int_0^T \frac{\langle \eta_t^{\bar{X}}, \mathbf{si} \rangle}{\langle \eta_t^{\bar{X}}, \mathbf{s} \rangle} \langle \eta_t^X, \mathbf{s}(f_t^{\mathcal{I}} - f_t) \rangle dt + \beta_G \int_0^T \frac{\langle \eta_t^H, \mathbf{i} \rangle}{\langle \eta_0^H, \mathbf{n} \rangle} \langle \eta_t^X, \mathbf{s}(f_t^{\mathcal{I}} - f_t) \rangle dt. \end{aligned} \quad (1.9)$$

De plus, il est possible de montrer que les marginales de  $\eta^X$ , conditionnellement à la taille  $n$  de la structure et au nombre  $s$  de susceptibles qu'elle contient, sont absolument continues. Soient  $\rho_{X,n,s}$  les densités associées, pour  $X \in \{H, W\}$ ,  $n \in \llbracket 1, n_{\max} \rrbracket$  et  $s \in \llbracket 1, n \rrbracket$ . Soit

$$\Lambda_X(t) = \lambda_{\bar{X}} \frac{\langle \eta_t^{\bar{X}}, \mathbf{si} \rangle}{\langle \eta_t^{\bar{X}}, \mathbf{s} \rangle} + \beta_G \frac{\langle \eta_t^H, \mathbf{i} \rangle}{\langle \eta_0^H, \mathbf{n} \rangle}.$$

Alors la famille de fonctions  $(\rho_{X,n,s})$  est une solution faible de l'équation différentielle suivante: pour tout  $\tau \in \mathbb{R}^{n-s}$  et  $t \in (0, T)$ ,

$$\begin{aligned} \partial_t \rho_{X,n,s}(t, \tau) - \sum_{k=1}^{n-s} \partial_{\tau_k} \rho_{X,n,s}(t, \tau) &= -s(\lambda_X i(\tau) + \Lambda_X(t)) \rho_{X,n,s}(t, \tau) \\ &+ \mathbf{1}_{\{s+1 \leq n\}} (s+1) (\lambda_X i(\tau_{1,n-s-1}) + \Lambda_X(t)) \rho_{X,n,s+1}(t, \tau_{1,n-s-1}) g_\nu(\tau_{n-s}), \end{aligned}$$

où  $\tau_{1,n-s-1} = (\tau_1, \dots, \tau_{n-s-1})$  et avec des conditions initiales appropriées. Cela relie l'Équation valeur mesure (1.6) à un système d'équations de transport non-linéaires et non locales.

Pour résumer, Théorème 1.4.1 donne accès aux dynamiques épidémiques exactes dans la limite grande population. L'objet limit  $\eta$  est riche, puisqu'il contient des informations détaillées sur l'état épidémique des foyers et lieux de travail, dont la durée restante de la période infectieuse de chaque individu ayant été en contact avec la maladie. Par conséquent,  $\eta$  reste infini-dimensionnel : nous aimerions donc obtenir une réduction plus forte, fini-dimensionnelle, en adoptant une description plus grossière de l'état épidémique de la population.

Une telle réduction plus forte peut être obtenue dans le cas Markovien où  $\nu$  est exponentiellement distribuée, de paramètre  $\gamma$ . Soient  $s(t)$  et  $i(t)$  les proportions d'individus respectivement susceptibles ou infectés, dans une population au temps  $t$  selon la distribution  $\eta_t$ . Par ailleurs, introduisons l'ensemble

$$\mathbb{S} = \{(n-i, i) : 2 \leq n \leq n_{\max}, 0 \leq i \leq n-1\}.$$

Pour  $(S, I) \in \mathbb{S}$  et  $X \in \{H, W\}$ , soit  $n_{S,I}^X(t)$  la proportion de foyers ( $X = H$ ) ou lieux de travail ( $X = W$ ) contenant  $S$  susceptibles et  $I$  infectés au temps  $t$ , selon  $\eta_t^X$ . Introduisons de plus les applications suivantes :

$$\tau_G(t) = \beta_G i(t), \text{ et } \tau_X(t) = \frac{\lambda_X}{\sum_{k \geq 1} k \pi_k^X} \sum_{(S,I) \in \mathbb{S}} SI n_{S,I}^X(t) \text{ pour } X \in \{H, W\}.$$

Supposons que, au temps 0, les durées restantes des périodes infectieuses des infectés sont *i.i.d.* de loi  $\nu$ , et qu'une fraction  $\varepsilon$  de la population est infectée uniformément au hasard, les autres étant susceptibles. Alors nous obtenons le résultat suivant.

**Theorem 1.4.2.** *Les fonctions  $(s, i, n_{S,I}^X : X \in \{H, W\}, (S, I) \in \mathbb{S})$  sont caractérisées comme étant l'unique solution du système dynamique suivant : pour tout  $t \geq 0$ ,  $X \in \{H, W\}$  et  $(S, I) \in \mathbb{S}$ ,*

$$\begin{aligned} \frac{d}{dt} s(t) &= -(\tau_H(t) + \tau_W(t) + \tau_G(t)s(t)), \\ \frac{d}{dt} i(t) &= -\frac{d}{dt} s(t) - \gamma i(t), \\ \frac{d}{dt} n_{S,I}^X(t) &= - \left( \lambda_X SI + \tau_{\bar{X}}(t) \frac{S}{s(t)} + \tau_G(t)S + \gamma I \right) n_{S,I}^X(t) \\ &+ \gamma(I+1) n_{S,I+1}^X(t) \mathbf{1}_{\{S+I < n_{\max}\}} \\ &+ \left( \lambda_X(S+1)(I-1) + \tau_{\bar{X}}(t) \frac{S+1}{s(t)} + \tau_G(t)(S+1) \right) n_{S+1,I-1}^X(t) \mathbf{1}_{\{I \geq 1\}}, \end{aligned} \quad (1.10)$$

avec des conditions initiales appropriées, détaillées dans l'énoncé du Théorème 3.3.3.

Remarquons que le système dynamique (1.10) décrit un modèle compartimental, dont les variables incluent les proportions de foyers et lieux de travail contenant un certain nombre de susceptibles et infectés. En particulier, cela implique que dans le cadre Markovien, la corrélation des états sanitaires des structures due aux infectés n'a pas besoin d'être prise en compte pour bien décrire la dynamique épidémique. C'est ce qui permet d'obtenir une réduction fini-dimensionnelle.

Néanmoins, le système dynamique est d'ordre  $O(n_{\max}^2)$ . Bien que ceci soit avantageux en comparaison d'un modèle EBCM foyer-travail à  $O(n_{\max}^3)$  équations (Section 3.3.2), cela soulève la question si la résolution numérique du système dynamique est pertinente vis-à-vis des simulations stochastiques. Des explorations numériques suggèrent que sur une large partie de l'espace de paramètres, résoudre le système dynamique (1.10) est significativement plus rapide que la simulation d'une trajectoire stochastique du processus épidémique initial. Cela indique que le modèle réduit obtenu peut être un choix pertinent pour des études par simulations.

### 1.4.3 Chapitre 4 - Étude de la sensibilité de la précision de réduction aux paramètres du réseau et de l'épidémie

L'objectif de ce chapitre est d'investiguer et quantifier l'impact des paramètres du réseau de contact et des paramètres épidémiques sur la précision des deux modèles réduits introduits dans les chapitres précédents. En effet, cela permet d'identifier les conditions sous lesquelles chacun de ces modèles réduits fournit une bonne approximation de l'épidémie. De plus, pour chaque réduction, cela met en lumière les paramètres influençant le plus la qualité d'approximation : la réduction bénéficierait donc d'une inférence précise de ces paramètres, indiquant l'importance de l'estimation de ces derniers.

Rappelons ici que les deux modèles réduits consistent en des approches complémentaires. D'une part, le modèle réduit uniformément mélangeant du Chapitre 2 est très parcimonieux, robuste aux variations du modèle, mais sans garanties théoriques de précision. D'autre part, la limite grande population du Chapitre 3 est exacte asymptotiquement, mais uniquement si le réseau de contact correspond exactement à celui du modèle foyer-travail. Par conséquent, notre étude de sensibilité sert à explorer deux questions différentes, qui sont soulevées par les forces et faiblesses de chaque modèle réduit.

Dans un premier temps, nous nous concentrons sur le modèle réduit uniformément mélangeant du Chapitre 2. Étant donnée l'absence de garanties théoriques de précision, nous avons précédemment exploré par simulations l'espace de paramètres pour identifier les régions de celui-ci où la réduction est pertinente. Ici, nous complétons cette approche par une analyse de sensibilité globale basée sur la décomposition de la variance de Sobol'. Cela permet de quantifier l'influence de chaque paramètre, et de vérifier si ces résultats sont cohérents avec les conclusions du Chapitre 2.

Pour cela, nous considérons le modèle foyer-travail, dont les paramètres sont les distributions des tailles de structures  $(\pi^H, \pi^W)$ , les taux de contact  $(\beta_G, \lambda_H, \lambda_W)$  et le taux de guérison  $\gamma$ , voir Section 1.1.2. Ici, nous explorons uniquement l'impact des taux de contact, les autres paramètres étant fixés. Nous concevons un design expérimental permettant de couvrir un domaine pertinent de l'espace de paramètres, basé sur les scénarios épidémiques



du Chapitre 2. Ensuite, nous calculons numériquement les indices de Sobol' principaux et totaux des quantités d'intérêt, à savoir de la différence entre la taille du pique ou la taille finale prédite par le modèle réduit uniformément mélangeant du Chapitre 2, et par la limite grande population du Chapitre 3.

Les résultats obtenus sont qualitativement très différents entre taille du pique et taille finale. Pour la taille du pique, les trois taux de contact ont un impact significatif comparable sur la précision de la prédiction du modèle réduit uniformément mélangeant, quasiment sans effets d'interactions. Pour la taille finale, en revanche, le taux de contact en population générale est responsable de la majorité de la variance, avec des effets d'interaction mineurs. Notamment, cela semble cohérent avec les résultats du Chapitre 2, qui indiquaient que la taille finale est mieux prédite si la proportion d'infections en population générale est élevée.

Dans un second temps, nous nous intéressons au modèle réduit du Chapitre 3, qui correspond à la limite grande population du modèle foyer-travail et donne donc accès aux dynamiques épidémiques asymptotiquement exactes. Néanmoins, ce résultat repose sur le choix du réseau de contact qui contient deux hypothèses fortes : (1) l'indépendance du choix des structures entre individus, et (2) l'uniforme mélange au sein de toutes les structures. En particulier, toutes les deux sont des simplifications de la vie réelle : par exemple, les deux partenaires d'un couple peuvent avoir le même lieu de travail, et les études empiriques indiquent que les lieux de travail ne sont pas uniformément mélangeants. Mais, si ces hypothèses sont relâchées, est-ce que la limite grande population du Chapitre 3 fournit malgré tout une bonne approximation de l'épidémie se propageant sur ce réseau de contact ?

Pour répondre à cette question, nous proposons un modèle foyer-travail généralisé qui ne satisfait pas nécessairement les hypothèses (1) et (2). Comme précédemment, les distributions (asymptotiques) de tailles de foyers et lieu de travail sont données par  $\pi^H$  et  $\pi^W$ , respectivement. Intuitivement, la première différence vis-à-vis du modèle foyer-travail est qu'au sein de chaque foyer, plusieurs individus choisissent leur lieu de travail ensemble, chaque membre du foyer participant à ce choix commun avec probabilité  $q$ . Ensuite, au sein d'un lieu de travail de taille  $n$ , le réseau de contact intra-travail est donné par un graphe Erdős-Rényi  $G(n, \rho(n))$ , où  $n \mapsto \rho(n)$  est une application décroissante, dont la valeur minimale  $p$  est atteinte si  $n = n_{\max} < \infty$ . Notamment, le modèle foyer-travail correspond exactement au choix de paramètres  $p = 1, q = 0$ . Ici, nous faisons varier la densité intra-travail  $p$  et le chevauchement des structures  $q$ .

Concernant la propagation de l'épidémie, nous considérons des dynamiques *SIR*, comme dans les chapitres précédents. La paramétrisation est la même que pour le modèle foyer-travail, avec les taux de contact  $\beta_G, \lambda_H, \lambda_W$  spécifiques à chaque couche du réseau, et le taux de guérison  $\gamma > 0$ . La seule différence consiste en l'interprétation du taux de contact  $\lambda_W$ , qui représente maintenant le taux auquel un contact est établi le long des arêtes de chaque réseau intra-travail. En particulier, si au sein d'un lieu de travail, le réseau de contact est complet et contient  $s$  susceptibles et  $i$  infectés, cela donne lieu à une infection au taux  $\lambda_W si$ , comme attendu.

Nous nous concentrons sur la différence de taille du pique ou taille finale, entre le modèle foyer-travail généralisé et la limite grande population du Chapitre 3. Ces différences sont calculées numériquement, en considérant plusieurs choix de paramètres épidémiques, distributions de tailles de structures et valeurs de  $(p, q)$ . Nos résultats suggèrent que sur une part significative de l'espace de paramètres exploré, la limite grande population donne une



prédiction satisfaisante (marge d'erreur d'environ 5%). De plus, la densité intra-travail  $p$  a un impact plus fort que le chevauchement des structures  $q$ . Ainsi, prendre en compte une description plus détaillée des réseaux de contact intra-travail serait pertinent.

Ces résultats font partie d'un travail en cours consistant en une analyse de sensibilité plus large, qui devrait nous permettre d'étudier la robustesse de nos conclusions. Ces perspectives sont détaillées à la fin du Chapitre 4.

#### 1.4.4 Chapitre 5 - Sur les constructions spinales pour populations en interaction

Les chapitres précédents nous ont permis de proposer des modèles réduits qui donnent accès à des propriétés macroscopiques du modèle foyer-travail, telles que la proportion d'infectés au sein de la population, ou la proportion de foyers contenant un nombre donné de membres susceptibles et infectés. Toutefois, ils ne sont pas conçus pour capter des aspects microscopiques, comme les chaînes de contamination, qui donnent accès à une compréhension précise de la propagation de l'épidémie sur le réseau.

Afin de suivre les chaînes d'infection, nous pouvons faire une analogie entre l'arbre d'infections et un arbre généalogique : les événements de contamination correspondent aux naissances, et les guérisons aux décès. Par conséquent, les chaînes de contamination peuvent être vues comme des lignées ancestrales. Tandis que l'étude de telles lignées est classique pour l'approximation branchement valable en début d'épidémie, leur compréhension en-dehors du régime de branchement est un domaine de recherche actif (Bansaye, 2024; Calvez et al., 2022; Duchamps et al., 2023; Medous, 2023).

Dans ce chapitre, nous considérons donc le cadre général de processus de population stochastiques, densité-dépendants, car nos résultats ne se limitent pas aux modèles épidémiques puisque ces dynamiques émergent dans d'autres domaines, tels que l'écologie. Ici, nous faisons quelques hypothèses simplificatrices, en supposant notamment que l'espace de types des individus est fini. En effet, le modèle foyer-travail présente des difficultés intéressantes, à la fois techniques et liées à la représentation du processus qui permet de faire émerger l'arbre d'infection, qui pourront faire l'objet d'un travail futur. Néanmoins, nos résultats ouvrent d'ores et déjà la porte à des applications pertinentes en épidémiologie, y compris pour les modèles tenant compte de l'hétérogénéité de contacts, comme souligné dans la discussion du chapitre.

L'objectif de ce travail est d'étudier la lignée ancestrale d'un individu échantillonné au temps  $t > 0$  dans une population. C'est un objet d'étude classique dans le cadre de processus de branchement, en quel cas il est bien connu qu'une telle lignée typique se réduit à un processus de Markov, grâce à la formule tous-pour-un (Harris et al., 2016; Harris and Roberts, 2017), qui est apparentée aux équations de Feynman-Kac (Del Moral, 2004, Sections 1.3 and 1.4.4). La *construction spinale* ainsi définie, simplifie grandement l'étude de lignées typiques tant sur le plan théorique, comme le modèle réduit est un processus de Markov, que sur le plan numérique, car les simulations deviennent moins coûteuses en temps de calcul.

Récemment, Bansaye, 2024 a développé un processus spinal pour des populations multi-type avec interactions, *i.e.* où les taux de reproduction dépendent de l'état de la population. Dans ce cas, l'auteur se concentre sur la lignée d'un individu échantillonné au temps  $t > 0$  avec probabilité proportionnelle à  $\psi(x, \mathbf{z})$ , où  $x$  est le type de l'individu,  $\mathbf{z}$  la composition en

types actuelle de la population, et  $\psi$  une fonction strictement positive sur l'espace approprié. Une formule tous-pour-un reliant une telle lignée à un processus  $\psi$ -spinal est établie, en population finie et dans la limite grande population.

Ici, nous abordons certains inconvénients et questions ouvertes de cette approche. Plus précisément, nous proposons d'abord une formule tous-pour-un alternative, reliant une lignée  $\psi$ -échantillonnée à un processus de Markov inhomogène en temps. Cela permet de considérer une classe plus large de stratégies d'échantillonnage  $\psi$ , donne lieu à une interprétation intuitive, et se révèle possiblement avantageux sur le plan numérique. Ensuite, nous quantifions la convergence de la  $\psi$ -épine de Bansaye, 2024 vers sa limite grande population.

Afin d'introduire notre construction spinale inhomogène en temps, commençons par donner une brève description du processus de population. Nous considérons une population dont la taille est bornée par  $K < \infty$ . Chaque individu est caractérisé par un type  $x \in \mathcal{X}$ , où  $\mathcal{X}$  est un ensemble fini. La population est décrite par le vecteur  $\mathbf{z} = (\mathbf{z}_x, x \in \mathcal{X})$  appartenant à l'ensemble  $\mathcal{Z}_K = \{\mathbf{z} \in (\mathbb{N} \cup \{0\})^{\text{Card}(\mathcal{X})} : \|\mathbf{z}\|_1 \leq K\}$ . Ici,  $\mathbf{z}_x$  compte le nombre d'individus de type  $x$ . Dans une population de composition  $\mathbf{z}$ , chaque individu de type  $x$  meurt et laisse la descendance  $\mathbf{k} \in \mathcal{Z}_K$  au taux  $\tau_{\mathbf{k}}(x, \mathbf{z})$ . Cela signifie que l'état de la population saute de  $\mathbf{z}$  à  $\mathbf{z} + \mathbf{k} - \mathbf{e}(x)$ , où  $(\mathbf{e}(x), x \in \mathcal{X})$  est la base canonique de  $\mathcal{Z}_K$ . Le taux  $\tau_{\mathbf{k}}(x, \mathbf{z})$  est supposé nul si la population qui en résulte contient plus de  $K$  individus.

Soit  $\mathbb{G}(t)$  l'ensemble des individus en vie au temps  $t$ , et écrivons  $u \geq v$  si l'individu  $u$  descend de  $v$  et  $x_u(s)$  pour le type de l'unique ancêtre de  $u$  vivant au temps  $s \leq t$ . Le processus de population d'intérêt est donné par  $X(t) = \sum_{u \in \mathbb{G}(t)} \delta_{(u, x_u(t))}$ , et  $Z(t)$  est la composition en types de la population associée, au temps  $t \geq 0$ . Enfin, pour  $(x, \mathbf{z}) \in \mathcal{S}_K = \{(x, \mathbf{z}) \in \mathcal{X} \times \mathcal{Z}_K : \mathbf{z}_x \geq 1\}$ , nous désignons par  $\mathfrak{X}(x, \mathbf{z})$  une condition initiale adéquate pour  $X$ , qui assure que la population est de composition  $\mathbf{z}$ , et qui distingue un individu  $u_x$  de type  $x$ .

Désormais, nous sommes prêts à donner une description intuitive du processus spinal, qui est inspiré par Marguet, 2019. En continuité avec cet article, nous l'appellerons également processus  $\psi$ -auxiliaire. Soit  $t \geq 0$  le temps auquel l'échantillonnage a lieu. De façon similaire à Bansaye, 2024, notre processus spinal prend en compte à la fois le type de l'épine  $Y^{(t)}$ , et la composition de la population  $\zeta^{(t)}$ , en raison de la densité-dépendance des taux de reproduction. À la fois les taux de reproduction de l'épine et hors de l'épine sont biaisés, et ce biais repose sur l'application  $m_\psi$  définie comme suit. Pour tout  $(x, \mathbf{z}) \in \mathcal{S}_K$ , pour tout  $s \in [0, t]$ ,

$$m_\psi(x, \mathbf{z}, t - s) = \mathbb{E} \left[ \sum_{u \in \mathbb{G}(t), u \geq u_x} \psi(x_u(t), Z(t)) \middle| X(s) = \mathfrak{X}(x, \mathbf{z}) \right].$$

Cela correspond à la moyenne  $\psi$ -pondérée des types d'individus en vie au temps  $t$ , dont l'ancêtre au temps  $s$  était un individu donné de type  $x$  dans une population de composition  $\mathbf{z}$ . Par la suite, nous faisons l'hypothèse que  $m_\psi$  est une fonction strictement positive.

Supposons maintenant qu'au temps  $s \in [0, t]$ , la population est de composition  $\mathbf{z}$  et l'épine de type  $x$ . Alors d'une part, l'épine se reproduit, laisse descendance  $\mathbf{k} \in \mathcal{Z}_K$  et devient de type  $y$  au taux

$$\tau_{\mathbf{k}}(x, \mathbf{z}) \mathbf{k}_y \frac{m_\psi(y, \mathbf{z} + \mathbf{k} - \mathbf{e}(x), t - s)}{m_\psi(x, \mathbf{z}, t - s)}.$$

En d'autres termes, une telle transition est plus probable si elle augmente la  $\psi$ -moyenne des descendants de l'épine au temps  $t$ . De façon similaire, un individu de type  $y \in \mathcal{X}$  autre que l'épine laisse descendance  $\mathbf{k}$  au taux

$$\tau_{\mathbf{k}}(y, \mathbf{z}) \frac{m_{\psi}(x, \mathbf{z} + \mathbf{k} - \mathbf{e}(y), t - s)}{m_{\psi}(x, \mathbf{z}, t - s)}.$$

De nouveau, ce taux de transition est biaisé positivement s'il crée un environnement plus favorable pour l'épine, au sens où la  $\psi$ -moyenne des descendants de l'épine au temps  $t$  est plus grande si au temps  $s$ , la population est de composition  $\mathbf{z} + \mathbf{k} - \mathbf{e}(y)$  que si elle est de composition  $\mathbf{z}$ . Le processus spinal  $(Y^{(t)}(s), \zeta^{(t)}(s))_{s \leq t}$  est donc un processus de Markov inhomogène en temps, qui peut être défini rigoureusement comme l'unique solution d'un système d'EDS (voir Proposition 5.3.2).

Le cœur de ce chapitre est la formule tous-pour-un suivante. Nous désignons par  $\mathbb{E}_{\mathfrak{X}(x, \mathbf{z})}$  et  $\mathbb{E}_{(x, \mathbf{z})}$  respectivement l'espérance conditionnellement à  $\{X(0) = \mathfrak{X}(x, \mathbf{z})\}$  et  $\{(Y^{(t)}(0), \zeta^{(t)}(0)) = (x, \mathbf{z})\}$ .

**Theorem 1.4.3.** *Pour tout  $t \geq 0$  et toute fonction mesurable  $F : \mathbb{D}([0, t], \mathcal{S}_K) \rightarrow \mathbb{R}_+$ , pour tout  $(x, \mathbf{z}) \in \mathcal{S}_K$ ,*

$$\begin{aligned} \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{u \in \mathbb{G}(t), u \geq u_x} \psi(x_u(t), Z(t)) F((x_u(s), Z(s))_{s \leq t}) \right] \\ = m_{\psi}(x, \mathbf{z}, t) \mathbb{E}_{(x, \mathbf{z})} [F((Y^{(t)}(s), \zeta^{(t)}(s))_{s \leq t})]. \end{aligned} \quad (1.11)$$

La preuve consiste à identifier les deux membres de l'Équation (1.11) au moyen du générateur infinitésimal des semi-groupes conservatifs, inhomogènes en temps, associés. Cela correspond à un changement de probabilité inhomogène en temps (Marguet, 2019) qui relie une lignée  $\psi$ -typique du processus de population au processus  $\psi$ -auxiliaire.

Soulignons ici que notre hypothèse  $m_{\psi} > 0$  est moins restrictive que  $\psi > 0$ . Notamment, elle permet d'échantillonner directement au sein de sous-populations, ce qui peut trouver des applications pertinentes. Par exemple, le processus  $\psi$ -auxiliaire est capable de capter les chaînes de contamination menant à l'infection d'individus vulnérables, comme les personnes âgées, au temps  $t$ . Le Chapitre 5 contient une discussion détaillée des forces et faiblesses du processus spinal obtenu.

Dans un second temps, nous considérons la  $\psi$ -épine introduite par Bansaye, 2024, et quantifions sa convergence vers sa limite grande population. Présentons le contexte de cette étude. Soit  $d = \text{Card}(\mathcal{X})$ . Pour une population de taille au plus  $K$ , soit  $Z^K(t) \in [0, 1]^d$  l'état de la population au temps  $t$ , défini comme sa composition en types normalisée par  $K$ . De plus, pour tout  $\mathbf{k} \in (\mathbb{N} \cup \{0\})^d$ , le taux de reproduction  $\tau_{\mathbf{k}} : \mathcal{X} \times [0, 1]^d \rightarrow \mathbb{R}_+$  dépend du type de l'individu et de l'état de la population uniquement, et l'ensemble  $\{\mathbf{k} : \tau_{\mathbf{k}} \neq 0\}$  est fini. Sous des hypothèses faibles, la suite de processus  $(Z^K)_{K \geq 1}$  converge alors en probabilité, uniformément sur des intervalles de temps finis, vers une fonction déterministe  $z : \mathbb{R}_+ \rightarrow [0, 1]^d$  solution d'un système dynamique.

Considérons une application  $\psi : \mathcal{X} \times [0, 1]^d \rightarrow \mathbb{R}_+$  continue sur  $\mathcal{X} \times [0, 1]^d$  et continûment différentiable sur  $\mathcal{X} \times (0, 1)^d$ . Soit  $\psi_K$  définie par  $\psi_K(x, \mathbf{z}) = \psi(x, \mathbf{z}/K)$  pour  $(x, \mathbf{z}) \in \mathcal{S}_K$ . La  $\psi_K$ -épine peut alors être décrite comme suit. Supposons que la population est dans l'état

$\mathbf{z} \in [0, 1]^d$ , et que l'épine est de type  $x \in \mathcal{X}$ . Alors l'épine laisse descendance  $\mathbf{k} \in (\mathbb{N} \cup \{0\})^d$  et devient de type  $y$  au taux

$$\tau_{\mathbf{k}}(x, \mathbf{z}) \mathbf{k}_y \frac{\psi(y, \mathbf{z} + (\mathbf{k} - \mathbf{e}(x))/K)}{\psi(x, \mathbf{z})}.$$

De même, un individu de type  $y \in \mathcal{X}$  autre que l'épine laisse descendance  $\mathbf{k}$  au taux

$$\tau_{\mathbf{k}}(y, \mathbf{z}) \frac{\psi(x, \mathbf{z} + (\mathbf{k} - \mathbf{e}(y))/K)}{\psi(x, \mathbf{z})}.$$

La suite associée de processus  $\psi_K$ -spinaux  $(Y^K, \zeta^K)_{K \geq 1}$ , dont les composantes décrivent le type de l'épine et l'état de la population, converge en loi vers  $(\Upsilon, \mathbf{z})$ , où  $\Upsilon$  est un processus de Markov inhomogène en temps. Plus précisément, au temps  $s \geq 0$ ,  $\Upsilon$  saute de l'état  $x \in \mathcal{X}$  à l'état  $y \in \mathcal{X}$  au taux

$$\sum_{\mathbf{k} \in (\mathbb{N} \cup \{0\})^d} \mathbf{k}_y \tau_{\mathbf{k}}(x, z(s)) \frac{\psi(y, z(s))}{\psi(x, z(s))}.$$

Nous renvoyons à la Section 5.2.2 pour un résumé détaillé des résultats et hypothèses de Bansaye, 2024.

Or, les fluctuations de  $Z^K$  autour de sa limite déterministe  $z$  sont bien comprises, les régimes classiques étant les fluctuations gaussiennes, ainsi que les déviations modérées et les grandes déviations (Ethier and Kurtz, 1986, Chapitre 11; Britton and Pardoux, 2019b; Pardoux, 2020). Techniquement, ces résultats ne se transfèrent pas immédiatement à  $\zeta^K$  puisque ce dernier distingue un individu, l'épine, de la dynamique générale. Intuitivement, toutefois, un tel contrôle des fluctuations devrait avoir lieu, puisque l'impact de l'épine devient négligeable en grande population. Ici, nous travaillons donc sous l'Hypothèse 5.4.3 qu'il existe des suites positives  $(t_K)_{K \geq 1}$ ,  $(\varepsilon_K)_{K \geq 1}$  et  $(\alpha_K)_{K \geq 1}$  telles que, pour tout  $K \geq 1$ ,

$$\mathbb{P} \left( \sup_{s \in [0, t_K]} \|\zeta^K(s) - z(s)\|_1 \geq \varepsilon_K \right) \leq \alpha_K.$$

L'idée générale est de transférer ce contrôle de la convergence de  $(\zeta^K)_{K \geq 1}$  vers  $z$ , à la convergence de  $(Y^K)_{K \geq 1}$  vers  $\Upsilon$ . Cela nécessite de plus la régularité des taux de reproduction, qu'on suppose ici Lipschitz-continus (Hypothèse 5.4.2). Enfin, un argument de couplage mène au résultat suivant :

**Theorem 1.4.4.** *Sous les Hypothèses 5.4.2 et 5.4.3, il existe  $C > 0$  tel que pour tout  $K \geq 1$ ,*

$$\mathbb{P}(\exists t \leq t_K : Y^K(t) \neq \Upsilon(t)) \leq \alpha_K + C(\varepsilon_K + K^{-1})t_K.$$

En particulier, si les suites  $(t_K/K)_{K \geq 1}$ ,  $(t_K \varepsilon_K)_{K \geq 1}$  et  $(\alpha_K)_{K \geq 1}$  tendent vers 0 lorsque  $K$  tend vers l'infini, avec  $(t_K)_{K \geq 1}$  convergeant vers  $t_\infty \in \mathbb{R}_+ \cup \{\infty\}$ , le couplage est asymptotiquement exact sur  $[0, t_\infty]$ . Ainsi, un régime particulièrement prometteur est celui des déviations modérées en présence d'un équilibre asymptotiquement stable, comme cela est le cas pour certains modèles endémiques (Pardoux, 2020; Prodhomme, 2023). Nous renvoyons à la discussion du Chapitre 5 pour plus de détails.



# The epidemic footprint of contact structures

---

*Models with several levels of mixing (households, workplaces), as well as various corresponding formulations for  $R_0$ , have been proposed in the literature. However, little attention has been paid to the impact of the distribution of the population size within social structures, effect that can help plan effective interventions. We focus on the influence on the model outcomes of teleworking strategies, consisting in reshaping the distribution of workplace sizes. We consider a stochastic SIR model with two levels of mixing, accounting for a uniformly mixing general population, each individual belonging also to a household and a workplace. The variance of the workplace size distribution appears to be a good proxy for the impact of this distribution on key outcomes of the epidemic, such as epidemic size and peak. In particular, our findings suggest that strategies where the proportion of individuals teleworking depends sublinearly on the size of the workplace outperform the strategy with linear dependence. Besides, one drawback of the model with multiple levels of mixing is its complexity, raising interest in a reduced model. We propose a homogeneously mixing SIR ODE-based model, whose infection rate is chosen as to observe the growth rate of the initial model. This reduced model yields a generally satisfying approximation of the epidemic. These results, robust to various changes in model structure, are very promising from the perspective of implementing effective strategies based on social distancing of specific contacts. Furthermore, they contribute to the effort of building relevant approximations of individual based models at intermediate scales.*

## **Code availability.**

<https://forgemia.inra.fr/francois.deslandes/communityepidemics>

*This chapter is based on the preprint "The epidemiological footprint of contact structures in models with two levels of mixing", a joint work with Vincent Bansaye, François Deslandes and Elisabeta Vergu (Bansaye et al., 2023a).*

## **Contents**

---

<b>2.1</b>	<b>Introduction</b>	<b>66</b>
<b>2.2</b>	<b>Model with two levels of mixing: description, simulation approach, key parameters, simulation scenarios</b>	<b>70</b>
2.2.1	General model description	70
2.2.2	Structure size distributions	70
2.2.3	Numerical simulation scenarios: structure size distributions and epidemiological parameters	71
2.2.4	Simulations of the population structure and epidemic process	73

2.2.5	Simulation of teleworking strategies . . . . .	74
<b>2.3</b>	<b>The impact of the size distribution of closed structures and assessment of teleworking strategies . . . . .</b>	<b>75</b>
2.3.1	Outbreak criterion, $R_I$ and type of infection . . . . .	75
2.3.2	The effect of structure size distribution on epidemic outcomes . . . . .	77
2.3.3	Teleworking strategies . . . . .	79
2.3.4	Robustness to the form of the infection term . . . . .	82
<b>2.4</b>	<b>Reduction to compartmental ODEs based on the initial growth rate . . . . .</b>	<b>83</b>
2.4.1	Laplace transform of the infection rate in a uniformly mixing population . . . . .	85
2.4.2	Characterization of the initial growth rate . . . . .	88
2.4.3	ODE reduction of the multilevel model based on the initial growth rate . . . . .	90
2.4.4	Assessment of the reduction robustness . . . . .	91
<b>2.5</b>	<b>Discussion . . . . .</b>	<b>92</b>
	<b>Appendix . . . . .</b>	<b>98</b>
<b>2.A</b>	<b>Generation of structure size distributions . . . . .</b>	<b>98</b>
<b>2.B</b>	<b>Parameter values for scenarios . . . . .</b>	<b>98</b>
<b>2.C</b>	<b>Numerical computations of epidemic parameters and outcomes . . . . .</b>	<b>99</b>
<b>2.D</b>	<b>Proof of Proposition 2.4.1 . . . . .</b>	<b>100</b>
<b>2.E</b>	<b>Computation of the exponential growth rate for the <i>SEIR</i> model with two levels of mixing . . . . .</b>	<b>103</b>
<b>2.F</b>	<b>Numerical aspects for the model reductions of Section 2.4.4 . . . . .</b>	<b>104</b>

---

## 2.1 Introduction

The dynamics of an epidemic relies on the contacts between susceptible and infected individuals in the population. The number and characteristics of contacts has a major quantitative effect on the epidemic. In addition to the main features playing a role in the description of contacts, such as the age and propensity to travel of individuals (Davies et al., 2020; Giles et al., 2020), the nature of the contact is also crucial: homogeneous mixing in closed structures (household, workplaces, schools,...) or related to other intermediate social structures (group of friends, neighbors...) e.g. House and Keeling, 2008. The heterogeneity of contacts can be captured in network based models (Keeling and Eames, 2005) or models with two levels of mixing (Ball and Neal, 2002). These models distinguish a global level of mixing corresponding to a uniformly mixing general population, as well as a local level consisting in an overlapping groups model, meaning that each individual belongs to one or several small contact groups such as households and workplaces and schools. These modeling frameworks or their simplified unstructured versions allow to tackle important questions related to the control of epidemic dynamics by acting specifically on these different population structures. Furthermore, the computation of the corresponding reproduction number, arguably one of the most important epidemic indicators, enables to assess control measures.

For the homogeneous mixing *SIR* model, several important characteristics can be summarized by the reproduction number  $R_0$ . This threshold parameter indicates whether there may be a large epidemic outbreak, allows to calculate the final epidemic size and the fraction of the population that needs to be vaccinated in order to stop an outbreak, see e.g. Heesterbeek and Dietz, 1996; Ball et al., 2016 and references therein. It is also directly linked to the exponential growth rate  $r$  at the beginning of the epidemic, and has a clear interpretation as the mean number of individuals contaminated by a single infected individual in a large susceptible population. For models with two levels of mixing, however, the definition of a unique reproduction number combining these criteria has not been achieved yet. Instead, various reproduction numbers have been proposed, of which Ball et al., 2016 have given an interesting overview. All of them respect the threshold of 1 for large epidemic outbreaks, and they generalize one or another aspect of the traditional  $R_0$ . Some of these reproduction numbers have the advantage of an intuitive interpretation. This is the case of the reproduction number  $R_I$  introduced in the supplementary material of Pellis et al., 2009 for the household-workplace model, and which was previously introduced for household models (Becker and Dietz, 1995; Ball et al., 1997). Its definition relies on a multi-type branching process which focuses on primary cases within households and workplaces, grouping all secondary cases as descendants of the primary cases. Then  $R_I$  is defined as the Perron root of the corresponding average offspring matrix. In this paper, we will show that this reproduction number has the advantage of being connected to further relevant information on the household-workplace epidemic, namely the proportions of infections occurring at each level of mixing.

Nevertheless, a drawback of most of the reproduction numbers for household-workplace models that are described by Ball et al., 2016 is that by construction, they lose track of time. Indeed, in an effort to construct meaningful generations of infected individuals, the timing of the infections is neglected. As a consequence, contrary to the case of homogeneous mixing, there is no simple link between these reproduction numbers and the initial exponential growth rate. The only exception is  $R_r$ , a reproduction number which has originally been introduced by Goldstein et al., 2009 for household models, and whose definition has been extended by Ball et al., 2016 to household-workplace models. The definition of this reproduction number depends explicitly on the exponential growth rate  $r$ . But as far as we see, it has no easy intuitive interpretation. It thus seems pertinent to complement the information yielded by a reproduction number such as  $R_I$  with the growth rate  $r$ . While simple closed analytic expressions seem out of reach, Pellis et al., 2011 have obtained an interesting characterization that we use and complement by more explicit expressions.

Given the relative difficulty for computing reproduction numbers for models with several levels of mixing, Goldstein et al., 2009 have suggested, in the case of the simplest model with two levels of mixing, namely structured only in general population and households, to first estimate the growth rate from data, and to then compute  $R_r$ . Trapman et al., 2016 have gone one step further, by proposing to first infer  $r$  from data, and then totally neglect the population structure and approximate a reproduction number from  $r$  using the formula linking the reproduction number to the exponential growth rate in the homogeneous mixing model. They find that this procedure is generally satisfactory, indicating that this procedure defines a homogeneous mixing model able to capture key aspects of the beginning of the epidemic. This makes one wonder to what extent it is possible in general to approach an epidemic spreading in a household-workplace model by a simple, unstructured, well parametrized compartmental model. Some work has been done in this direction by del Valle Rafo et al.,



2021. They have shown that it is possible to approach an *SIRS* household-workplace model by a homogeneously mixing *SIYRS* model, where *Y* stands for infected but no longer infectious individuals, once the parameters have been well chosen. Hence, they obtain the first approximation of multi-level epidemic dynamics using homogeneous mixing compartmental ODEs.

Naturally, models with two levels of mixing raise the question of the way their social organisation characterized by small contact structures has an impact on major features of an epidemic. From the point of view of control, they constitute minimal models allowing to account for closures of workplaces or schools. For the past years, governments worldwide have implemented such non-pharmaceutical interventions (NPIs) in reaction to the COVID-19 epidemic. Since then, several studies have assessed the impact of these measures on the epidemic spread. Both analysis of empirical studies (Mendez-Brito et al., 2021) and simulation studies (Backhausz et al., 2022; Simoy and Aparicio, 2021) come to the conclusion that especially (partial) school closure and/or home working have a substantial impact on the epidemic. Together, these findings motivate an interest in mathematical models enabling a closer study of school and workplace closures, and more generally the effect of control measures targeting small contact structures.

In this paper, we are interested in the impact of the distribution of individuals in closed structures on epidemic dynamics. In order to address this question, we consider a stochastic *SIR* model with two levels of mixing, namely a global and a local level. While the former corresponds to the general population, the latter is subdivided into two layers representing households and workplaces, respectively. Note that while our model does not explicitly distinguish schools, they can be considered as workplaces. In particular, we are motivated by and study the impact of control policies based on differentiated social distancing. For some structures, in particular for households, it is natural to assume that their size distribution is fixed and control policies cannot act on it. For others, such as workplaces and schools, control measures aiming at contact reduction can be considered, COVID-19 epidemic having raised this issue in new manners. Focusing on workplaces, we study here how control strategies which consist in modifying the structures' size distribution, can impact different epidemic outcomes. More generally, we demonstrate through simulations that the size distribution of closed structures has a significant effect on epidemic dynamics, as assessed by the total number of infections and by the initial growth rate of infection and by the maximal number of infected individuals along time. In particular, when both the number of individuals and structures are fixed, implying that the average structure size is constant as well, we show that these epidemic outcomes are sensitive to the variance of the structure size distribution. In short, balancing structure sizes reduces the impact of the epidemic.

One drawback of the model with two levels of mixing is that numerical simulations rely on good knowledge of several epidemic parameters, such as the rates of infection within each level, which may not be easy to assess. However, considering the significant impact of structure size distributions on epidemic outcomes and the fact that control measures may actively impact these distributions, it seems crucial not to neglect this particular population structure. This motivates the development of reduced epidemic models, which aim to be more parsimonious, while still being able to capture the impact of small structures on the epidemic thanks to a pertinent choice of parameters. Here, we propose such a reduced model, that we evaluate using simulations. It consists of a deterministic, homogeneously mixing *SIR* model, whose infectious contact rate is chosen as to ensure that the reduced

model and model with two levels of mixing share the same exponential growth rate. Hence, we will see that the initial growth rate is the key parameter for reducing the full epidemic process at the macroscopic level.

The questions we consider here involve quantities which capture some main features of the epidemics which are relevant for specific phases of an epidemic. Indeed, starting from a single infectious individual in a large population of size  $N$ , epidemic dynamics can be decomposed into three phases. This has been proven for simple models such as the homogeneously mixing *SIR* model, for which we will detail these phases below. However, this decomposition still holds in more complex models, including the model with two levels of mixing studied here.

*Phase 1: random behavior in small population.* When the number of infected individuals  $I(t)$  is of order 1,  $I(t)$  is approximated by a linear birth and death process. This approximation holds on finite time intervals, but also up to a time  $T_N$  which tends to infinity when  $N$  tends to infinity. More precisely, both processes coincide as long as the number of infected individuals is below  $\sqrt{N}$  (Ball and Donnelly, 1995). Let us also mention Barbour and Utev, 2004, for comparison results until the infected population reaches sizes of order of  $N^{2/3}$ , for a discrete time counterpart of the *SIR* model.

*Phase 2: deterministic evolution and linear behavior.* When  $1 \ll I(s), I(t) \ll N$ , the number of infected follows a deterministic and exponential dynamic:  $I(t) \approx I(s)e^{(\beta-\gamma)(t-s)}$ , where  $\beta$  is the transmission rate and  $\gamma$  the recovery rate. This approximation is valid as soon as  $s, t$  tend to infinity but remain far from the time  $\log(N)/(\beta - \gamma)$ , which corresponds to the entry in the macroscopic level. This deterministic phase allows to capture the initial growth rate of infection,  $\beta - \gamma$ , by considering the slope of the growth of  $I$  on a logarithmic scale. We refer to Bansaye et al., 2023b and references therein for more precise results.

*Phase 3: macroscopic deterministic behavior.* When the number of infected individuals is of order  $N$ , the proportion of susceptible, infected and recovered individuals can be approximated by a macroscopic deterministic system. More precisely, letting  $N$  go to infinity, the trajectories of  $(S/N, I/N, R/N)$  converge in law on finite time intervals to the solutions of the *SIR* dynamical system. The approximation is valid for any  $t$  greater than  $\log(N)/(\beta - \gamma)$ . For accurate results, we refer in particular to Barbour and Reinert, 2013 and Barbour, 1978. Let us also mention that fluctuations around the deterministic curve are of order  $1/\sqrt{N}$  by classical Gaussian approximation (Chapter 7, Sections 4 and 5, in Ethier and Kurtz, 1986). In our study, *phase 1* corresponds to the regime where stochasticity of the individual-based version of the *SIR* model is observed in simulations. *Phase 2* is the relevant regime for the definition of  $R_I$  and the initial growth rate  $r$ . *Phase 3* yields the deterministic macroscopic approximation, where stochasticity vanishes. It starts at a random time necessary to reach a macroscopic proportion of infected. This time represents the starting point of the comparison between the stochastic structured model and its reduced ODE-based counterpart we propose.

This paper is structured as follows. Section 2.2 presents the main modeling ingredients, such as a detailed description of the model with two levels of mixing and proper introduction of considered key parameters, as well as numerical settings for simulations. Section 2.3 is devoted to the study of the impact of the structure size distribution on some main epidemic outcomes, namely the reproduction number, the exponential growth rate, the peak size and the final epidemic size. For this purpose, two slightly different situations are considered. While the size of the population is always considered fixed, we first keep the total number of

workplaces constant as well but vary the way individuals are distributed among these given workplaces, see Section 2.3.2. Second, we consider teleworking strategies, which differ from the previous setting as for simulations, these strategies amount to creating a new workplace of size one for each teleworking employee, see Section 2.3.3. Finally, in Section 2.4, we propose an ODE reduction of the initial multi-level model based on the computation of the initial growth rate and assess its robustness. The paper concludes with a Discussion (Section 2.5) of the main results on the impact of structure size distributions on epidemic dynamics, their robustness to different modeling assumptions, and their implications for control measures.

## 2.2 Model with two levels of mixing: description, simulation approach, key parameters, simulation scenarios

### 2.2.1 General model description

We consider an *SIR*-type model with two levels of mixing by considering global and two types of local contacts following two local partitions of the population, see Ball and Neal, 2002. In addition to homogeneous mixing in the general population, contacts occur in households and workplaces of various sizes, in which the population is structured. Each individual belongs both to a household and a workplace, which are chosen independently from one another. Generally speaking, infection spreads through contacts between susceptible and infected individuals within each level of mixing, which are characterized by different contact rates among individuals as will be detailed below. Infected individuals recover at rate  $\gamma$ .

We distinguish two slightly different types of parametrization concerning contact description. For closed structures such as households and workplaces, we will use *one-to-one* infectious contact rates  $\lambda_H$  and  $\lambda_W$ , respectively. In other words, within a household, if there are  $s$  susceptibles and  $i$  infected individuals, each susceptible is infected at rate  $\lambda_H i$  (resp.  $\lambda_W i$  for workplaces). This has the disadvantage to make the average number of contacts established by each individual grow with the size of the structure. This is tractable for structures of finite size, and a good enough approximation of contacts within very small structures, but it is not realistic at the scale of the general population. Instead, within the general population, when there are  $s$  susceptibles and  $i$  infectious individuals, each susceptible individual becomes infected at rate  $\beta_G i / (N - 1)$  where  $N$  is the population size. Here, the parameter  $\beta_G$  represents the *one-to-all* infectious contact rate, which is the global rate at which an infected individual makes contact with all other individuals in the population. Hence the corresponding one-to-one infectious contact rate  $\lambda_G^{(N)} = \beta_G / (N - 1)$  is small. This allows to scale the contact rates when  $N$  tends towards infinity, so that the mean number of contacts made by an infected individual remains constant. The global rate of infection in the population is then  $\beta_G IS / (N - 1) = \lambda_G^{(N)} IS$  where  $S$  (resp.  $I$ ) is the number of susceptible (resp. infected) individuals, and  $S$  is indeed of order  $N$ .

### 2.2.2 Structure size distributions

Let us introduce the size distribution of households and workplaces, called  $\pi^H$  and  $\pi^W$ , respectively. When the number of structures is large,  $\pi_k^H$  (resp.  $\pi_k^W$ ) is the proportion of households (resp. workplaces) of size  $k \geq 1$ . The total number of individuals is  $N$ , which is fixed. Besides, all individuals belong to one (and only one) household and workplace, the

latter being of size one for teleworking employees. Notice that the following equivalences hold a.s.

$$N \sim N_H \sum_{k \geq 1} k \pi_k^H \sim N_W \sum_{k \geq 1} k \pi_k^W \quad (N \rightarrow \infty),$$

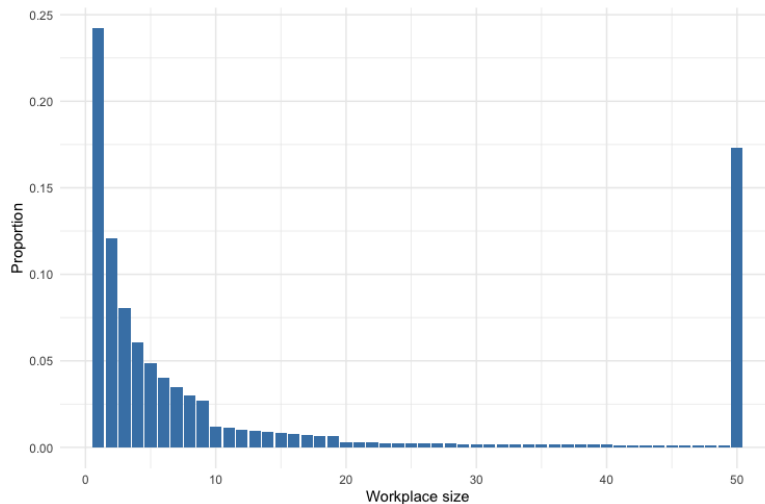
where  $N_H$  (resp.  $N_W$ ) is the total number of households (resp. workplaces). We define  $m_H = \sum_{k \geq 1} k \pi_k^H$  (resp.  $m_W = \sum_{k \geq 1} k \pi_k^W$ ) the average household (resp. workplace) size.

### 2.2.3 Numerical simulation scenarios: structure size distributions and epidemiological parameters

In the numerical explorations of the impact of the structure size distribution on the epidemic dynamics, we use the household size distribution observed in France in 2018 as reference distribution and also more generally, unless stated otherwise. We also provide a workplace distribution derived from the workplace size distribution of Ile-de-France in 2018, later called reference workplace size distribution, and we refer to 2.A for detail. In particular, we assume homogeneous mixing within structures, which is unrealistic for large workplace sizes, and we thus have limited workplaces to size 50 at most. The household reference distribution is stated in Table 2.1, while the workplace reference distribution is shown in Figure 2.1.

**Table 2.1:** Reference household size distribution corresponding to the size distribution of households in France in 2018.

Household size	1	2	3	4	5	6
Proportion	0.367	0.326	0.136	0.114	0.041	0.016



**Figure 2.1:** Reference workplace size distribution distribution derived from workplace size distributions in Ile-de-France.

To study the impact of the average workplace size and workplace variance we provide the following sets of exploratory workplace size distributions: (A) a set of 160 workplace

distributions with mean ranging from 3 to 30, different variances and maximal size 50; (B) a set of 100 workplace distributions with mean 20, different variances and maximal size 50; (C) a set of 100 workplace distributions with mean 7, different variances and maximal size 50. These workplace size distributions were generated using random mixtures of distributions, as explained in 2.A. The sets of structure size distributions are summarized in Table 2.2.

Name	Number of distributions	Average size (value/range)	Variance (value/range)	Maximum size	
Reference household size distribution	1	2.2	1.6	6	
Reference workplace size distribution	1	13.8	331.5	50	
Exploratory workplace size distributions	A	160	3-30	9.8-554.3	50
	B	100	20	42.4-544.0	50
	C	100	7	5.8-248.8	50

**Table 2.2:** Household and workplace size distribution sets.

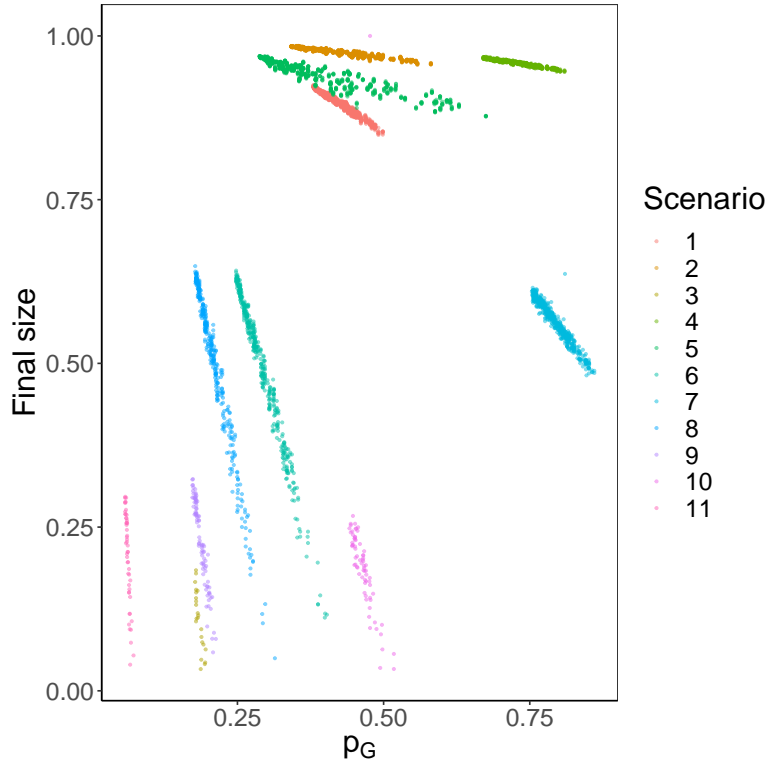
Numerical exploration of the model was performed using a combination of various epidemic parameters. We designed scenarios to cover a range of interesting situations that illustrate the mathematical properties of the model and its approximations as well as the expected epidemic behaviour of the model for several infectious diseases. Our study covers several values of the reproduction number, from threshold values to values observed in real world epidemic such as the flu (Ajelli et al., 2014) or COVID-19 (Locatelli et al., 2021; Galmiche et al., 2021). We also explore higher values of the reproduction number, closer to what is observed in highly contagious diseases such as chickenpox (Silhol and Boëlle, 2011). Similarly, we cover a range of distributions of infections between structures and global mixing. We study balanced scenarios where the proportion of infection in global mixing is between 30 and 40%, such as those observed in influenza or COVID-19 epidemics, as well as more contrasted situations where infections at the global or local level strongly dominate, as for chickenpox. An overview of all considered scenarios is given in Table 2.3.

**Table 2.3:** Main features of the epidemic scenarios considered.

Scenario	Reproduction number	Distribution of infections	Comment
1	COVID-19	balanced	COVID-19-like scenario
2	high	balanced	
3	threshold	balanced	
4	high	mostly global mixing	
5	high	mostly workplaces	
6	flu	balanced	
7	flu	mostly global mixing	
8	flu	balanced	flu-like scenario
9	threshold	mostly structures	
10	threshold	balanced	
11	threshold	fully structures	

Notice that the epidemic parameters  $\lambda_H$ ,  $\lambda_W$ ,  $\beta_G$  and  $\gamma$  need to be calibrated for each explored structure size distribution in order to approach the desired scenarios. More detail is given in 2.B, including values of  $R_I$ , epidemic growth rate and proportions of infection per layers for each scenario, and the corresponding epidemic parameters for the reference

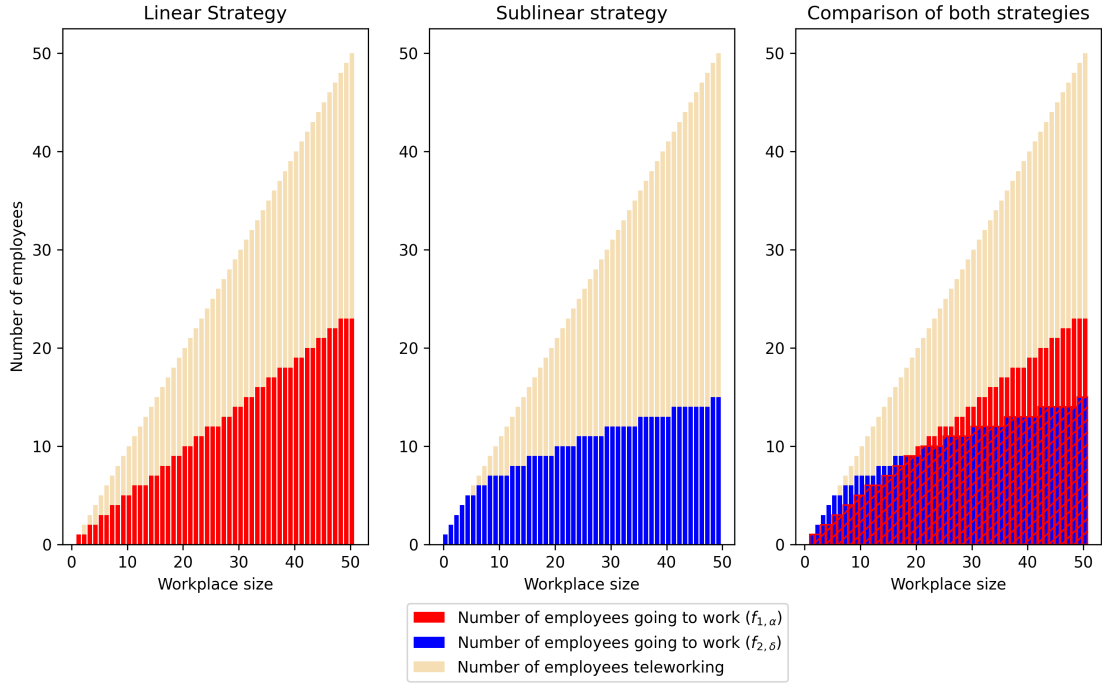
size distributions. Without loss of generality, the recovery rate is set to 1 in all simulations. Illustration of the simulated final size for each scenario and exploratory workplace size distribution in the simulation study are provided in Figure 2.2.



**Figure 2.2:** Simulated values of the final size as a function of the proportion of infection via global mixing ( $p_G$ ) for all scenarios of Table 2.3 and all exploratory workplace size distributions from set A, Table 2.2. For each epidemic scenario and each workplace size distribution, simulations were repeated 10 times. Each colored scatter plot on the figure thus contains  $160 \times 10$  points. The household size distribution is the reference household size distribution.

#### 2.2.4 Simulations of the population structure and epidemic process

The epidemic process was simulated using the Gillespie algorithm (Stochastic Simulation Algorithm). A population structure (contacts between individuals) is first generated from the size distributions of households and workplaces, generated as described in Section 2.2.3 and 2.A. Individuals are listed from 1 to  $N$  and placed in structures randomly by applying the following iterative process: (i) for each type of structure, randomly select a structure size  $k$  with probability given by the size distribution; (ii) if the number  $n$  of individuals that have not yet been assigned to a structure of the same type exceeds  $k$ , randomly select  $k$  individuals among those; otherwise, group all remaining individuals in a structure of size  $n < k$ . For a given population structure, the algorithm computes the rates of events, *i.e.* infection events in households, workplaces and the general population, and recovery, as described in Section 2.2.1. These rates are used to derive the next event, *i.e.* infection of a susceptible individual or recovery of an infected individual, and the corresponding event time. The epidemic is



**Figure 2.3:** For each workplace size, number of employees coming to work on site or teleworking according to either the linear or the sublinear strategy. The parameters  $\alpha$  and  $\delta$  are chosen in order to observe an average proportion of employees teleworking per workplace equal to 0.5 for the uniform workplace size distribution ( $\alpha = 0.46$ ,  $\delta = 2.01$ ).

initiated with a single infected individual, selected uniformly at random in the population. For each run of the epidemic process, we compute several classical summary statistics: (i) the final epidemic size, *i.e.* the number of individuals that are in the recovered state when the number of infected individuals becomes 0; (ii) the infectious peak size, *i.e.* the maximum number of infectious individuals occurring simultaneously over the course of the epidemic; (iii) the infectious peak time, *i.e.* the time at which the infectious peak occurs.

## 2.2.5 Simulation of teleworking strategies

We evaluated the evolution of different epidemic outcomes for varying proportions of teleworkers using two strategies, as illustrated in Figure 2.3:

(i) Linear strategy. The workplace size distribution is modified according to the function  $f_{1,\alpha}(k) = \lceil \alpha k \rceil$  for  $\alpha \in [0, 1]$ , where  $\lceil \cdot \rceil$  is the usual ceiling function. This means that a workplace of size  $k$  becomes a workplace of size  $\lceil \alpha k \rceil$  and the remaining  $k - \lceil \alpha k \rceil$  individuals now telework. (ii) Sublinear strategy. The workplace size distribution is modified according to the function  $f_{2,\delta}(k) = \lceil \delta k^{\frac{1}{2}} \rceil \wedge k$ , where  $\delta \geq 0$ . This means that a workplace of size  $k$  becomes a workplace of size  $\lceil \delta k^{\frac{1}{2}} \rceil$  and the remaining  $k - \lceil \delta k^{\frac{1}{2}} \rceil$  individuals now telework.

The rationale behind the sublinear strategy is that withdrawing an individual from a large structure has a stronger impact in terms of number of contacts. Notice that we consider here the exponent  $1/2$  for the sublinear strategy, but one could more generally consider any exponent  $q < 1$ .

## 2.3 The impact of the size distribution of closed structures and assessment of teleworking strategies

### 2.3.1 Outbreak criterion, $R_I$ and type of infection

Various notions of reproduction numbers have been proposed, as a compromise between complexity of the computations and epidemiological interpretation. The idea is to capture the mean number of infections caused more or less directly by one single "typical" individual. This concept is primarily defined in the first steps of the epidemic, which usually can be approximated by a branching process, whose mean satisfies a linear ODE. A typical individual corresponds then to a uniform sample in the corresponding population. The reproduction number is delicate to define for epidemiological processes with multi-level contacts, such as the one we consider here. We recall that each infected individual infects an individual outside his structure with rate  $\beta_G S / (N - 1)$ , which is in the *phase 1* approximated by  $\beta_G$ , since the number of susceptibles  $S$  is approximately  $N$ . As a consequence, the mean number of individuals directly infected in the general population by a single infected individual is  $\beta_G / \gamma$ .

Following the Supplementary Material of Pellis et al., 2009, we structure the infected population following the origin of the infection and consider successive generations of infected individuals:

$$(I_n^G, I_n^H, I_n^W)_{n \geq 0}, \quad I_n = I_n^G + I_n^H + I_n^W.$$

Processes  $I_n^G$ , resp.  $I_n^H$  and  $I_n^W$ , count the number of individuals in generation  $n$ , which have been infected through the mean field, respectively in the household and in the workplace. Hence  $I_n$  is the total number of infected individuals in generation  $n$ . At time 0, we assume  $I_0 = I_0^G = 1$ . The next generation  $n + 1$  of infected individuals is created by considering the number of direct infections  $I_{n+1}^G$  in the general population, plus the local epidemic triggered within structures. This process is illustrated in Figure 2.4.

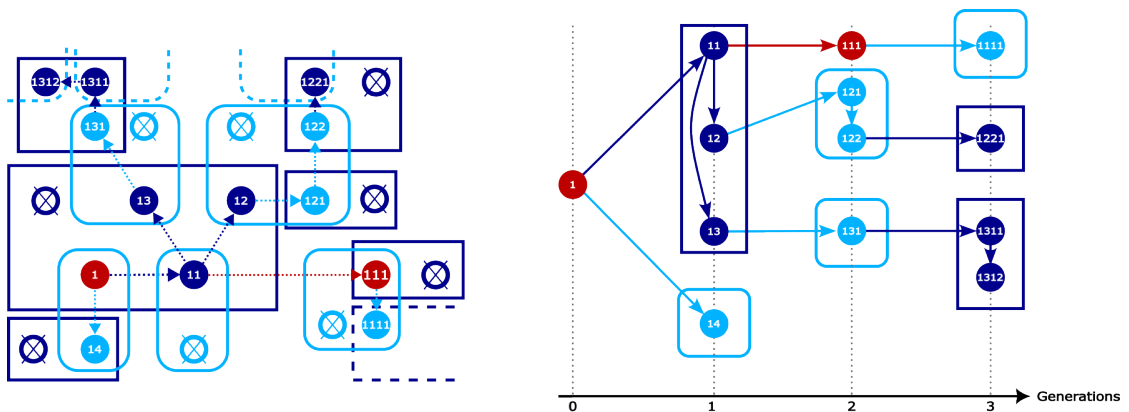
To compute the mean number of infections per generation, it is necessary to compute the mean number of individuals infected during the epidemic triggered by a single infected individual in a given structure. Thus, we introduce  $i_H(k)$  (resp.  $i_W(k)$ ), the average total number of infections starting from one infected individual in a closed population of size  $k$ , one-to-one contact rate  $\lambda_H$  (resp.  $\lambda_W$ ) and recovery rate  $\gamma$  as defined in Section 2.2.1. It corresponds to the number of infections caused by a single infected individual which introduces the epidemic into his household (resp. his workplace) of size  $k$ . Recalling that  $m_H = \sum_{k \geq 1} k \pi_k^H$  (resp.  $m_W = \sum_{k \geq 1} k \pi_k^W$ ), we define

$$\hat{\pi}_k^H = \frac{k \pi_k^H}{m_H}, \quad \text{resp.} \quad \hat{\pi}_k^W = \frac{k \pi_k^W}{m_W},$$

as the size biased distribution of structure sizes, which naturally defines the household (resp. workplace) size distribution of an individual chosen uniformly at random in the population. Then the numbers of infected individuals at each level triggered by an infected individual whose size structure is distributed according to the size biased law are defined by:

$$\mathcal{I}_G = \frac{\beta_G}{\gamma}, \quad \mathcal{I}_H = \sum_k \hat{\pi}_k^H i_H(k), \quad \mathcal{I}_W = \sum_k \hat{\pi}_k^W i_W(k). \quad (2.1)$$





**Figure 2.4:** Example of an epidemic with two levels of mixing (in general population and within structures). The spread of the epidemic is shown on the left. Households and workplaces are delimited respectively in dark and light blue. Only structures containing infected individuals are shown. Individuals who have been infected during the epidemic appear as plain dots, whose colors indicate the means by which the infection occurred: through the general population (red), within households (dark blue) or within workplaces (light blue). The arrows keep track of the spread of the disease, pointing from the infector to the newly infected. Their color refers again to the type of infection. Members of a structure who have not been infected during the within-structure epidemic are represented as crossed circles. In the case of the branching model, they are never going to be infected, as a secondary introduction of the disease in an already infected structure is negligible at the beginning of the epidemic. The different generations of infected as seen by the branching process are represented on the right. Colors still encode the way each infected is contaminated, and arrows represent the true order of infections as depicted on the left. The branching genealogy is depicted through lexicographic labeling of individuals. Labels have been reported on the left panel, as to simplify identification of infected individuals in both means of representation.

Following Pellis et al., 2009, the expectation  $\mathbb{E}((I_n^G, I_n^H, I_n^W)^T)$  can be approximated by a sequence  $X_n$  satisfying the following linear induction

$$X_{n+1} = AX_n$$

where  $A$  is the mean reproduction matrix:

$$A = \begin{pmatrix} \mathcal{I}_G & \mathcal{I}_G & \mathcal{I}_G \\ \mathcal{I}_H & 0 & \mathcal{I}_H \\ \mathcal{I}_W & \mathcal{I}_W & 0 \end{pmatrix}. \quad (2.2)$$

As  $A$  is a primitive matrix, Perron Frobenius theorem yields the asymptotic behavior of  $X_n = A^n X_0$  using its positive eigenlements, see Athreya and Ney, 1972. More precisely, the unique positive vector  $P = (p_G, p_H, p_W)^T$  solution of

$$AP = R_I P, \quad p_G + p_H + p_W = 1, \quad (2.3)$$

gives the proportion of infected individuals from each source: general population, households, workplaces. The associated positive eigenvalue is

$$R_I = \| AP \|_1$$

which corresponds to the mean reproduction number

$$R_I = \mathcal{I}_G + (1 - p_H)\mathcal{I}_H + (1 - p_W)\mathcal{I}_W. \quad (2.4)$$

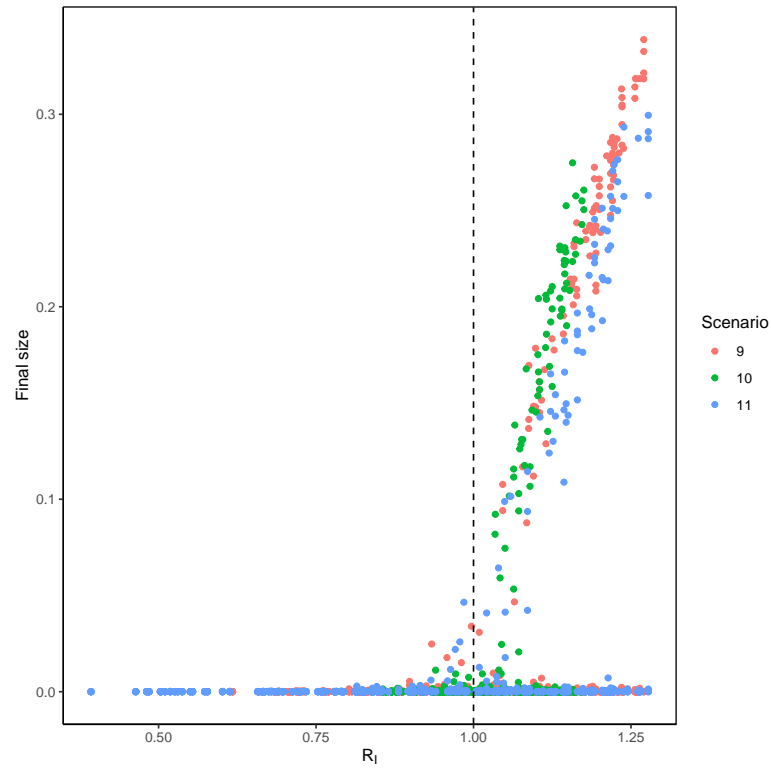
When  $R_I > 1$ , the process  $(I_n)_{n \geq 1}$  survives with positive probability and on this event a.s. grows geometrically fast with speed  $R_I$  yielding a supercritical regime, under an additional moment assumption on the number of infections, see also Athreya and Ney, 1972. We observe that the vector  $P$  gives the origin of infections for large times.

$R_I$  can play the role of an outbreak criterion, as illustrated in Figure 2.5. The final size of the epidemic is plotted against the value of  $R_I$  for three parameter sets (scenarios 9, 10 and 11 in Table 2.3). These epidemic parameter sets, combined with the set of workplace size distributions, allow to cover a large range of values for  $p_G$ ,  $p_H$  and  $p_W$  (more precisely,  $p_G$  between 0 and 0.5 and  $p_W$  between 0.05 and 0.65).

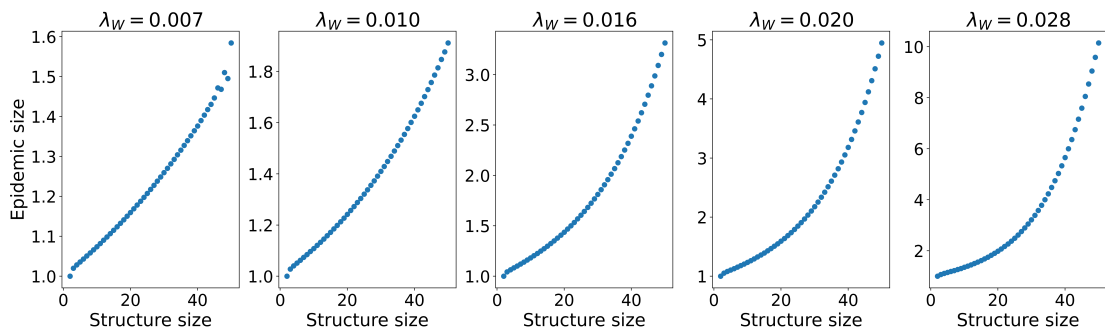
Let us end this section with a brief description of the numerical computation of  $R_I$ . For a given set of structure size, transmission parameter and recovery rate, we simulate the within structure epidemic using Gillespie's algorithm, and record the final epidemic size. We thus calculate the average size of epidemics in isolated structures from simulations of this epidemic process (default value of the number of runs: 10000, and 50000 for larger workplaces). Notice that the average final epidemic size could also be obtained through analytic results, see for instance Section 6.4 in Bailey, 1975. The value of  $R_I$  is then obtained from Equation (2.4), as the largest eigenvalue of the matrix defined in (2.2), by replacing unknown quantities by simulated quantities. The proportions of infections occurring within each layer of mixing are obtained from the associated eigenvector.

### 2.3.2 The effect of structure size distribution on epidemic outcomes

In this section, we will consider both the number of individuals and the number of workplaces as fixed. For the latter, one can imagine that this is due to logistic constraints, as there



**Figure 2.5:** Simulated values of the final size of epidemics of the stochastic structured model as a function of  $R_I$ . Simulations are performed for the reference household distribution, exploratory workplace size distributions from set A in Table 2.2 and epidemiological scenarios 9, 10 and 11 from Table 2.3. For each combination of workplace size distribution and epidemic scenario, 10 runs are performed.



**Figure 2.6:** Average final epidemic size within workplaces, i.e. for the uniformly mixing SIR model with one-to-one contact rate  $\lambda_W$  and removal rate  $\gamma = 1$ , as a function of workplace size. The average epidemic final size is computed using explicit formulas obtained from the Sellke construction, see for instance Section 2.4 in Andersson and Britton, 2000b. Different values of one-to-one contact rate  $\lambda_W$  have been considered, in order to cover a representative sample of scenarios 1 to 11 of Table B1, for the reference structure size distributions. More precisely, they correspond from left to right to scenarios 10, 4, 11, 2 and 5.

are only a certain number of offices (or classrooms) to dispose of. In other words, we are interested in understanding the epidemic impact of the way employees are assigned to those given workplaces.

From the previous section, one may notice that the workplace size distribution has a direct impact on  $R_I$  through  $\mathcal{I}_W$ , *i.e.* the average number of infections occurring within workplaces under the size-biased workplace size distribution  $\hat{\pi}^W$ . In particular, it follows from Equation (2.4) that diminishing  $\mathcal{I}_W$  is enough to ensure that  $R_I$  decreases. Further, Figure 2.6 illustrates that for most epidemic scenarios considered (Tables 2.3 and 2.4), the average number of infected  $i_W(k)$  caused by a within-workplace epidemic in a workplace of size  $k$  can be reasonably approximated in some scenarios by a linear function of the workplace size  $k$ . We deduce that, up to a constant  $c$ ,

$$\mathcal{I}_W = \sum_{k \geq 1} \hat{\pi}_k^W i_W(k) \approx c \sum_{k \geq 1} \hat{\pi}_k^W k = \frac{c}{m_W} \sum_{k \geq 1} k^2 \pi_k^W = c \frac{m_W^{(2)}}{m_W}, \quad (2.5)$$

where  $m_W^{(2)}$  designates the second moment of the workplace size distribution.

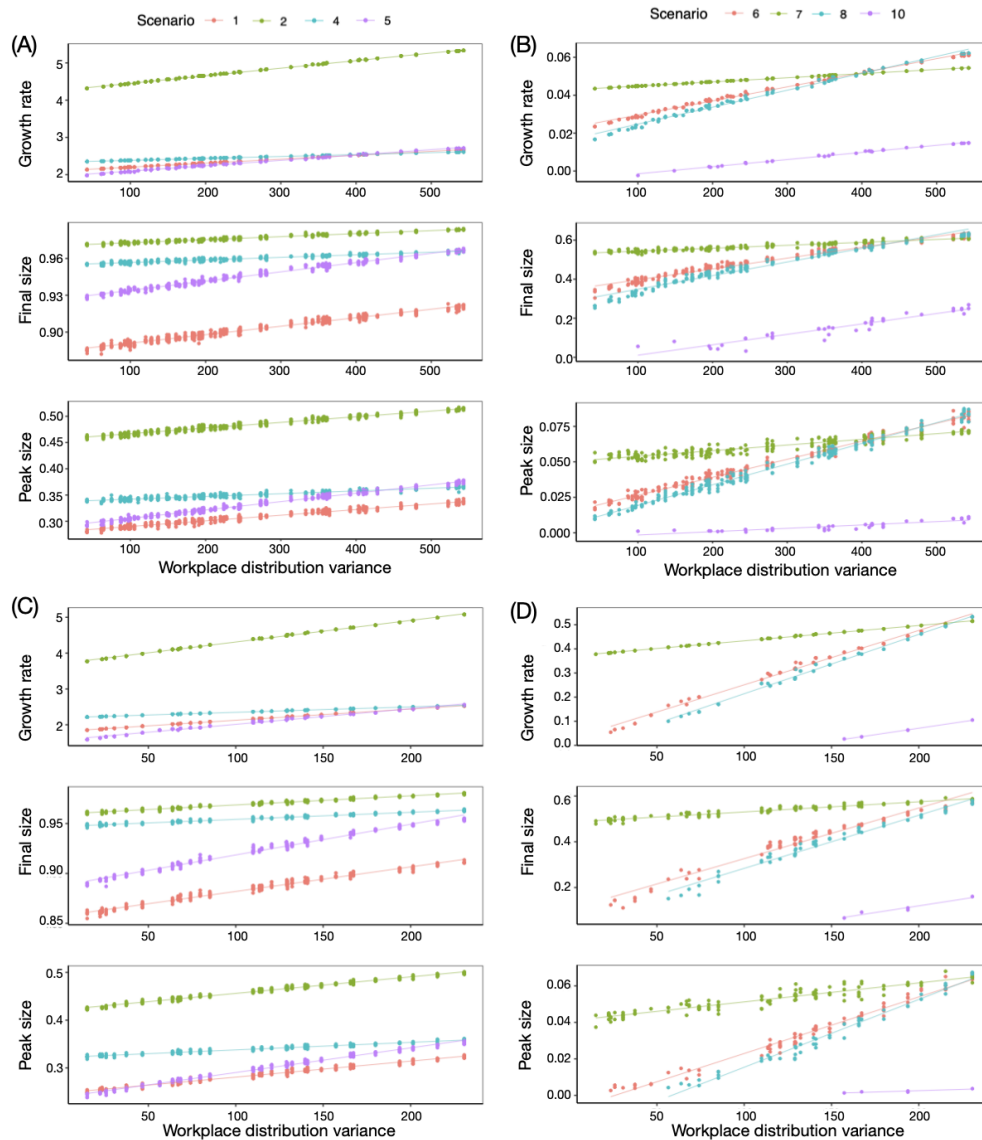
Since we suppose both the population size and the number of workplaces to be fixed, it follows that the average workplace size  $m_W$  is constant as well. As thus, in order to reduce  $\mathcal{I}_W$ , it is enough to reduce  $m_W^{(2)}$ . At fixed expected workplace size, modifying  $m_W^{(2)}$  is strictly equivalent to modifying the workplace size variance. Since the latter has a more direct and intuitive interpretation, we will focus on the variance of  $\pi^W$  as a natural candidate for the epidemic impact of the workplace size distribution.

In order to assess this impact, we will proceed by numerical exploration. A variety of workplace size distributions of average fixed at 20 and different variances have been considered, corresponding to exploratory workplace size distributions set B of Table 2.2. For each of these distributions and for epidemic scenarios 1, 2, 4 and 5 we have computed the epidemic growth rate as explained in 2.C, before evaluating through simulations the epidemic size and the peak size. Results have been reported in panels (A) of Figure 2.7, which is complemented by panel (B) for additional scenarios. These results thus illustrate the impact of the variance of the workplace size distributions on our selected epidemic outcomes (growth rate, final size, and peak size). This figure shows that the workplace size variance has a linear impact on these epidemic outcomes, observed for various values of average workplace size, see also panels (C) and (D) of Figure 2.7. Thus, the variance appears as a relevant indicator of the epidemic impact of the workplace size distribution. This also is of interest for the design of efficient control policies such as teleworking and (partial) closure of schools, as will be explored in the next section.

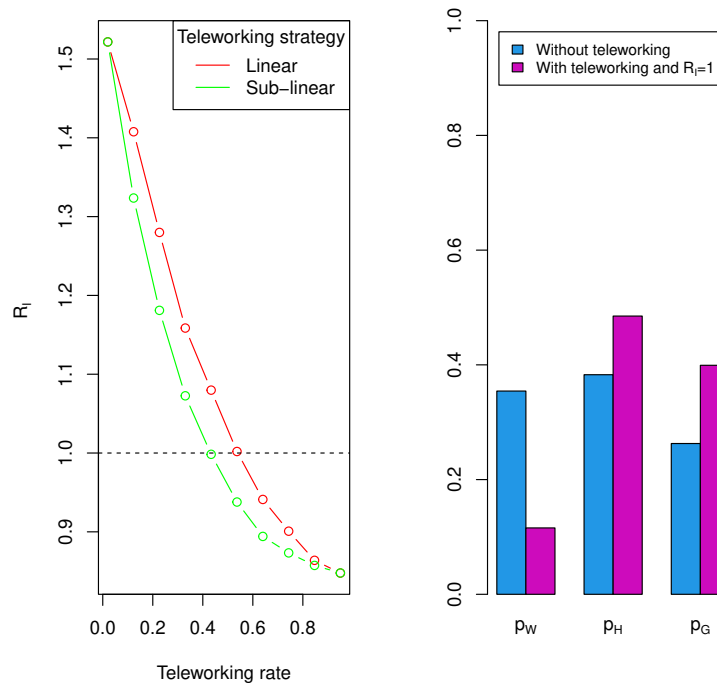
### 2.3.3 Teleworking strategies

Teleworking, a strategy to mitigate disease outbreaks, results in changes in the distribution of workplace sizes. These changes have an impact on the value of  $R_I$ , which has been shown to be a threshold criterion for epidemics, and more generally on the different epidemic outcomes.

Two teleworking strategies, formalized in Section 2.2.5, were assessed: (i) a linear strategy, where the same proportion of teleworking is applied equally to all workplaces, and (ii) a sublinear strategy where teleworking is more prevalent in larger workplaces. The motivation



**Figure 2.7:** Influence of the variance of the workplace size distribution on the growth rate, epidemic final size and epidemic peak size. For panels (A, B) (resp. (C, D)), simulations of the stochastic structured model were performed with the reference household size distribution, exploratory workplace size distribution set B (resp. C) with average workplace size of 20 (resp. 7) from Table 2.2. The epidemic scenario considered are 1, 2, 4, 5 (A, C) and 6, 7, 8, 10 (B, D) from Table 2.3. Simulations were repeated 10 times for each combination of scenario and workplace size distribution.



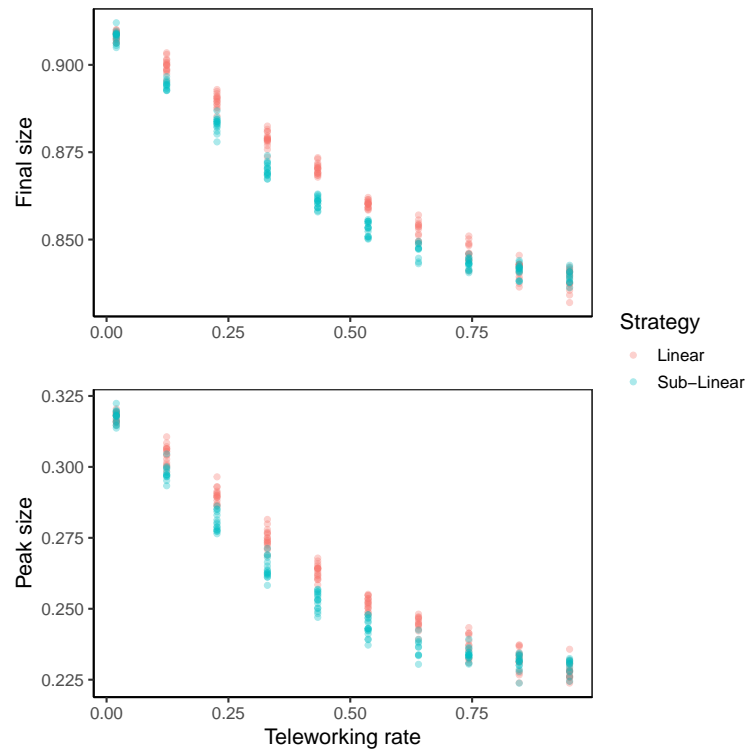
**Figure 2.8:** Effect of linear and sublinear teleworking strategies, using scenario 6 from Table 2.3 and reference household and workplace distributions from Table 2.2. Diminution of  $R_I$  as a function of teleworking rate (left) and proportions of infection in the different structures (right).

for such a strategy is twofold: larger workplaces allow for larger within-workplace epidemics, and they are expected to be better equipped to mitigate the economic impact of teleworking on the firm. We will indeed show that such a strategy has a beneficial health outcome.

Figure 2.8 illustrates the behaviour of the two teleworking strategies as a function of the teleworking rate, which is defined as the proportion of individuals in the population that do not have contacts in a workplace. Implementation of the teleworking strategies consists in adjusting parameters  $\alpha$  and  $\delta$  from Section 2.2.5 to obtain a prescribed value of the teleworking rate. The proportions of infections in the different structures are the same for both strategies at the threshold  $R_I = 1$ . The findings show a large reduction in the proportion of infections occurring in the workplaces. The sublinear strategy reaches the threshold for a lower teleworking rate, which indicates that this strategy has a lower impact on workplace organisation for similar epidemic outcomes.

Figure 2.9 illustrates, for the reference workplace size distribution and epidemic scenario 1 of Table 2.3, that even if the threshold cannot be reached by simply applying teleworking strategies, the sublinear strategy still outperforms the linear one. In particular, it shows that for the same global teleworking rate, the final size of the epidemic is lower for the sublinear strategy, except for the highest teleworking rates.

In other words, using sublinear teleworking policies (and more generally sublinear strategies for the closure of structures) allows either to reduce the need of teleworking in order to attain a given epidemiological outcome, or to reduce more strongly the epidemiological



**Figure 2.9:** Impact of linear and sublinear teleworking strategies on the final epidemic size (top) and epidemic peak size (bottom). Simulations of the stochastic structured model were performed using the reference household size distribution and workplace size distribution from Table 2.2, with epidemic parameters from scenario 1 of Table 2.3. 10 runs of were performed for each simulation scenario and distribution.

outcome for a given teleworking rate. Both effects may even be combined: more people go to work, but the epidemic outcome is reduced when compared to the linear strategy.

### 2.3.4 Robustness to the form of the infection term

Assuming linear growth of the number of infectious contacts per susceptible with the number of infected is often an overestimation. Thus, in this simulation study, we focus on having an infection rate within social structures growing sublinearly with the number of infected individuals in the structures. More precisely, we assume that within a household ( $X = H$ ) or workplace ( $X = W$ ) containing  $I$  infected and  $S$  susceptible individuals, the next infection occurs at rate  $\lambda_X S \sqrt{I}$ .

The observed linear effect of variance on the peak size and final size of the epidemic remains valid when the model is modified to use a sublinear infection rate in households and workplaces. Figure ?? shows the effect of the variance of the workplace size distribution with fixed mean on the epidemic outcomes for several scenarios (additional results for other scenarios can be found in Figure S5 of the Online Resource). This impact of the variance appears to hold for all scenarios, suggesting that the effect holds true regardless of both the epidemic speed and the proportions of infections that occur within structures. We also

confirmed the validity of this result for workplace distributions with smaller mean sizes, as illustrated in Figures S6 and S7 of the Online Resource.

## 2.4 Reduction to compartmental ODEs based on the initial growth rate

In this section, our aim is to propose a relevant reduction of the multi-level contact process, when the total population is large ( $N \gg 1$ ) and the number of infected individuals too, corresponding to *phase 2* and *phase 3* presented in the introduction. We propose a deterministic reduction which keeps track of the multi-level structuring of contacts, but has a low dimension and depends on few parameters only. It thus allows to see the effect of structure size distributions and control policies modifying them at a low computational cost. We show that the key parameter to achieve this reduction is the initial growth rate. As expected, it captures the initial growth of the size of the infected population. Actually, simulations show that it also allows for a relevant prediction of the rest of the epidemic, see Section 2.4.3 for details on the interest and limitations of this reduction.

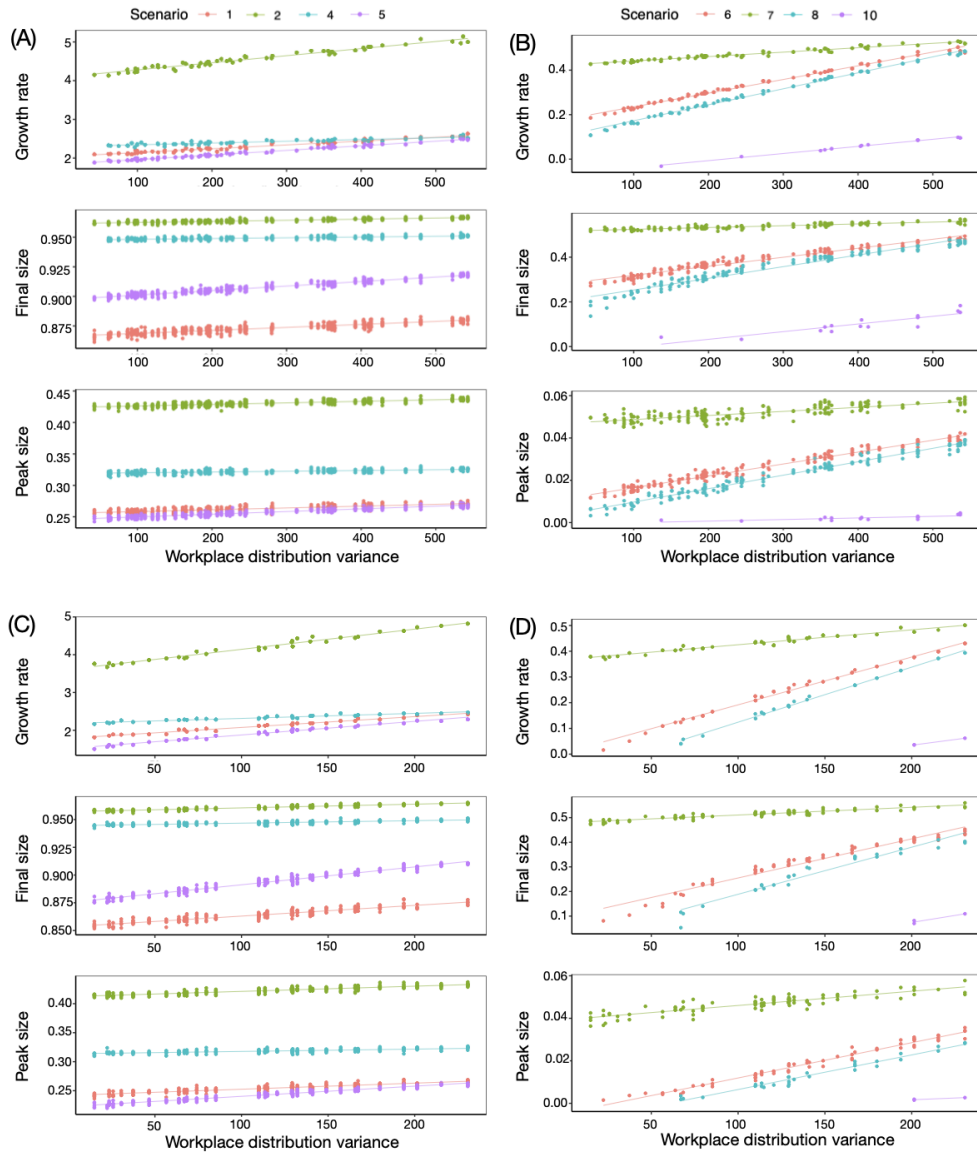
We assume that the total population size  $N$  is large and consider an approximation in an infinite population. As for the branching approximation considered in Section 2.3, we focus on the beginning of the epidemic (*phase 1* and *phase 2*). As households and workplaces are chosen independently from one another and for each individual, this implies that whenever an infection occurs in the general population, it will almost surely affect an individual whose household and workplace are entirely susceptible otherwise. Similarly, an infection taking place within a household will cause an infection within an otherwise susceptible workplace, and vice-versa. Some time is needed to reach a large (but still negligible compared to  $N$ ) number of infected individuals and forget the peculiar initial condition. Perron Frobenius theorem allows to get a deterministic growth rate, which is observable in *phase 2*. More precisely, it is observed at the beginning of phase 2 and more generally before the infected population is too close to  $N$ . For more quantitative results on this point for the SIR model, we refer to Proposition 5.1 of Bansaye et al., 2023b. It is called the initial growth rate  $r$ . It will play a crucial role in reducing and analyzing the process in the deterministic phase with a macroscopic number of infected individuals (*phase 3*).

For the stochastic *SIR* model in large homogeneously mixing populations, the initial growth rate can readily be obtained (Diekmann and Heesterbeek, 2000). Let us briefly explain the heuristics of the reasoning. Consider  $\mathbf{i}(t)$  the number of new infections in the population occurring at time  $t$  after the start of the epidemic. It is easy to see that  $\mathbf{i}(t)$  satisfies a renewal equation, which may be used to deduce an implicit equation for the exponential growth rate  $\bar{r}$ . Indeed, suppose that  $\mathbf{i}(t) = Ce^{\bar{r}t}$  for some constant  $C$ , and let  $\zeta(\tau)$  denote the average rate at which an individual who has been infected  $\tau$  units of time ago transmits the disease. Then  $\bar{r}$  is characterized as follows:

$$\mathfrak{L}(\zeta)(\bar{r}) := \int_0^\infty \zeta(\tau)e^{-\bar{r}\tau} d\tau = 1, \quad (2.6)$$

where  $\mathfrak{L}$  designates the Laplace transform operator. In order to conclude, it remains to make  $\zeta$  explicit, and at the beginning of the epidemic, one readily obtains the approximation  $\zeta(\tau) \approx \beta e^{-\gamma\tau}$ . Injecting this into the implicit equation  $\mathfrak{L}(\zeta)(\bar{r}) = 1$  leads to the well-





**Figure 2.10:** Influence of the variance of the workplace size distribution on the growth rate, epidemic final size and epidemic peak size. For (A, B) (resp. (C, D)), simulations of the stochastic structured model with sublinear infection rates were performed with the reference household size distribution, exploratory workplace size distribution set B (resp. C) with average workplace size of 20 (resp. 7) from Table 2.2. The epidemic scenario considered are 1, 2, 4, 5 (A, C) and 6, 7, 8, 10 (B, D) from Table 2.3. Simulations were repeated 10 times for each combination of scenario and workplace size distribution.

known growth rate  $\bar{r} = \beta - \gamma$ . This derivation can rigorously be obtained using branching approximations (Mode and Sleeman, 2000).

Exponential growth of infections is also observed when household-workplace structures are added to homogeneous mixing, and Pellis et al., 2011 characterize the associated growth rate  $r$ . Similarly to what has been done for the reproduction number, they aggregate within-structure epidemics to facilitate the mathematical analysis of the model. This leads to a point of view where an infected household contaminates a new workplace each time an infection occurs during the within-household epidemic, and vice-versa. Using Equation (2.6), this allows for the exact characterization of  $r$  as the unique solution of an implicit equation, which can be solved numerically. This motivates the study of the Laplace transform (2.6) of the average rate at which infections occur during the course of within-structure epidemics, which captures the dynamics of these infections.

Here, we follow and complement the approach in Pellis et al., 2011, mainly by providing more explicit expressions of the key quantities involved. The following Section 2.4.1 introduces our contribution, which lies in Proposition 2.4.1 and its corollaries. Subsequent Section 2.4.2 summarizes the work of Pellis et al., 2011, and allows to position our contribution in the context of their work.

### 2.4.1 Laplace transform of the infection rate in a uniformly mixing population

The main point lies in understanding the dynamics of the stochastic *SIR* model in a population of finite size  $k$ , with any one-to-one infectious contact rate  $\lambda$  and removal rate  $\gamma$ . The results on within-household or within-workplace epidemics will follow by choosing these parameters accordingly.

More precisely, consider the continuous-time Markov chain  $X_{k,\lambda,\gamma} = (S, I)$  taking values in  $\bar{\Omega}(k) = \{(s, i) \in (\mathbb{N} \cup \{0\})^2 : s + i \leq k\}$ , and whose transition rates are given by

Transition	Rate	
$(s, i) \rightarrow (s - 1, i + 1)$	$\lambda si$ ;	(2.7)
$(s, i) \rightarrow (s, i - 1)$	$\gamma i$ .	

Then  $S_t$  and  $I_t$  represent respectively the number of susceptible and infected individuals at time  $t$ . Furthermore, the initial condition of interest is  $x_0^k = (k - 1, 1)$ , and  $\mathbb{P}_{x_0^k}$  denotes the probability conditionally on  $X_{k,\lambda,\gamma}(0) = x_0^k$ .

Let us start by summarizing the results obtained by Pellis et al., 2011 on this matter. From (2.7), it is obvious that when the population is in state  $(s, i)$ , a new infection takes place at rate  $\lambda si$ . In particular, this rate is non-null if and only if  $si \geq 1$ , so we can restrict the study to the set of transient states  $\Omega(k) = \{(s, i) \in \bar{\Omega}(k) : i \geq 1\}$ . Let  $\zeta_{k,\lambda,\gamma}(t)$  be the average infection rate in a population of composition  $X_{k,\lambda,\gamma}(t)$ , conditionally on  $X_{k,\lambda,\gamma}(0) = x_0^k$ . It is clear that, by definition,

$$\zeta_{k,\lambda,\gamma}(t) = \sum_{(s,i) \in \Omega(k)} \lambda si \mathbb{P}_{x_0^k}(X_{k,\lambda,\gamma}(t) = (s, i)).$$

Consider  $\mathbf{Q}_{\lambda,\gamma}(k)$  the restriction of the generator of  $X_{k,\lambda,\gamma}$  to  $\Omega(k)$ , which is defined as the

following matrix indexed by states in  $\Omega(k)$ <sup>1</sup>

$$\forall (s, i) \in \Omega(k), \forall (s', i') \in \Omega(k),$$

$$(\mathbf{Q}_{\lambda, \gamma}(k))_{(s, i), (s', i')} = \begin{cases} -\lambda s i - \gamma i & \text{if } (s', i') = (s, i); \\ \lambda s i & \text{if } (s', i') = (s - 1, i + 1); \\ \gamma i & \text{if } (s', i') = (s, i - 1); \\ 0 & \text{otherwise.} \end{cases}$$

Then it is well known that for all  $(s, i)$  in  $\Omega(k)$ ,

$$\mathbb{P}_{x_0^k}(X_{k, \lambda, \gamma}(t) = (s, i)) = \left( e^{t \mathbf{Q}_{\lambda, \gamma}(k)} \right)_{x_0^k, (s, i)}.$$

Thus, a computation readily yields the following Laplace transform of  $\zeta_{k, \lambda, \gamma}$ , where  $I_{d(k)}$  is the identity matrix of appropriate dimension, namely  $d(k) = \#\Omega(k) = k(k + 1)/2$ : for any  $u \geq 0$ ,

$$\mathfrak{L}_{k, \lambda, \gamma}(u) := \mathfrak{L}(\zeta_{k, \lambda, \gamma})(u) = \sum_{(s, i) \in \Omega(k)} \lambda s i \left( (u I_{d(k)} - \mathbf{Q}_{\lambda, \gamma}(k))^{-1} \right)_{x_0^k, (s, i)}. \quad (2.8)$$

Let us now turn to our contributions. As we will see in the following proposition, we show that it is possible to go one step further and give an analytic expression for the relevant coefficients of

$$\widehat{\mathbf{Q}}_{k, \lambda, \gamma}(u) := (u I_{d(k)} - \mathbf{Q}_{\lambda, \gamma}(k))^{-1},$$

for any population size  $k$ . This will finally allow us to give a more explicit expression of  $\mathfrak{L}_{k, \lambda, \gamma}$ .

Studying the restricted generator  $\mathbf{Q}_{\lambda, \gamma}(k)$  leads to consider possible trajectories in  $\Omega(n)$  leading from some state  $(k - \ell, \ell)$  to another state  $(s, i)$ , for  $\ell \leq k$  and  $i \leq m = s + i \leq k$  (see 2.D). This incites us to introducing the following set, which allows to encode this set of trajectories:

$$\mathcal{I}_k(\ell, m, i) = \{(i_0, \dots, i_{m+1}) \in \{\ell\} \times \mathbb{N}^m \times \{i\} : \\ i_m \leq i, i_{j-1} - 1 \leq i_j \leq k - j \quad \forall 1 \leq j \leq m\}.$$

We are now ready to state our result.

**Proposition 2.4.1.** *Let  $\ell \in \{1, \dots, k\}$  and consider  $(s, i) \in \Omega(k)$  such that  $s \leq k - \ell$ . Then for any  $u \geq 0$ ,*

$$\left( \widehat{\mathbf{Q}}_{k, \lambda, \gamma}(u) \right)_{(k-\ell, \ell), (s, i)} = \frac{1}{u + \lambda s i + \gamma i} \sum_{\mathbf{i} \in \mathcal{I}_k(\ell, m, i)} \prod_{j=0}^m q_{k, \lambda, \gamma}(\mathbf{i}, j; u) g_{k, m, \lambda, \gamma}(\mathbf{i}, j; u) \quad (2.9)$$

where  $m = k - (s + i)$  and

$$q_{k, \lambda, \gamma}(\mathbf{i}, j; u) = \prod_{w=i_j}^{i_{j+1} - \mathbf{1}_{\{\ell=m\}}} \left[ 1 + \frac{u + \gamma w}{\lambda(k - j - w)w} \right]^{-1}$$

<sup>1</sup>We have made a slight change here compared to Pellis et al., 2011 since the mortality matrix  $\Delta$  needed to be deleted from the expression of  $Q$ .

and

$$g_{k,m,\lambda,\gamma}(\mathbf{i}, j; u) = \begin{cases} \left[ 1 + \frac{u + \lambda(k - j - i_{j+1} - 1)(i_{j+1} + 1)}{\gamma(i_{j+1} + 1)} \right]^{-1} & \text{for } j < m, \\ 1 & \text{for } j = m. \end{cases}$$

Furthermore, for every state  $(s, i) \in \Omega(k)$  such that  $s > k - \ell$ ,  $(\hat{\mathbf{Q}}_{k,\lambda,\gamma}(u))_{(k-\ell,\ell),(s,i)} = 0$ .

The proof of Proposition 2.4.1 uses arguments of linear algebra. Details can be found in 2.D. Using Equation (2.8), a more explicit expression of  $\mathfrak{L}_{k,\lambda,\gamma}$  follows from Equation (2.9). Let us define the ensemble

$$\mathcal{I}_k(m) = \{(i_0, i_1, \dots, i_m, i_{m+1}) \in \{1\} \times \mathbb{N}^{m+1} : \\ i_{m+1} \leq k - m, i_m \leq i_{m+1}, i_{j-1} - 1 \leq i_j \quad \forall 1 \leq j \leq m\}.$$

We can now state the result, using the same notations as in Proposition 2.4.1.

**Corollary 2.4.2.** For any integer  $k$  and any set of parameters  $\lambda, \gamma > 0$  and any  $u \geq 0$ ,

$$\mathfrak{L}_{k,\lambda,\gamma}(u) = \sum_{m=0}^{k-1} \sum_{\mathbf{i} \in \mathcal{I}_k(m)} c_{k,\lambda,\gamma}(\mathbf{i}; u) \prod_{j=0}^m q_{k,\lambda,\gamma}(\mathbf{i}, j; u) g_{k,m,\lambda,\gamma}(\mathbf{i}, j; u)$$

with

$$c_{k,\lambda,\gamma}(\mathbf{i}; u) = \left[ 1 + \frac{u + \gamma i_{m+1}}{\lambda(k - m - i_{m+1})i_{m+1}} \right]^{-1}.$$

In the following section, we will see how  $\mathfrak{L}_{k,\lambda,\gamma}$  intervenes in the computation of the growth rate  $r$  of the epidemic. Another quantity of similar nature will be needed, namely the Laplace transform  $\mathfrak{G}_{k,\lambda,\gamma,\beta_G}$  of the average rate  $\zeta_{k,\lambda,\gamma,\beta_G}^G$  at which all individuals of a structure of composition  $X_{k,\lambda,\gamma}(t)$ , conditionally on  $X_{k,\lambda,\gamma}(0) = x_0^k$ , contaminate individuals in the general population. Obviously, when the structure is in state  $(s, i) \in \bar{\Omega}(k)$ , this rate is given by  $\beta_G i$ , considering that the global proportion of susceptible individuals is close to one. In other words,

$$\mathfrak{G}_{k,\lambda,\gamma,\beta_G}(u) := \mathfrak{L}(\zeta_{k,\lambda,\gamma,\beta_G}^G)(u) = \sum_{(s,i) \in \bar{\Omega}(k)} \beta_G i (\hat{\mathbf{Q}}_{k,\lambda,\gamma}(u))_{x_0^k, (s,i)}.$$

This rate is positive if and only if  $(s, i) \in \Omega(k)$ . Proceeding like before, we obtain the following formula:

**Corollary 2.4.3.** For any integer  $k$ , for any set of parameters  $\lambda, \gamma, \beta_G > 0$ , for any  $u \geq 0$ ,

$$\mathfrak{G}_{k,\lambda,\gamma,\beta_G}(u) = \sum_{m=0}^{k-1} \sum_{\mathbf{i} \in \mathcal{I}_k(m)} c'_{k,\lambda,\gamma,\beta_G}(\mathbf{i}; u) \prod_{j=0}^m q_{k,\lambda,\gamma}(\mathbf{i}, j; u) g_{k,m,\lambda,\gamma}(\mathbf{i}, j; u)$$

with

$$c'_{k,\lambda,\gamma,\beta_G}(\mathbf{i}; u) = \frac{\beta_G i_{m+1}}{u + \lambda(k - m - i_{m+1})i_{m+1} + \gamma i_{m+1}}.$$

*Proof of Corollaries 2.4.2 and 2.4.3.* Start by noticing that

$$\Omega(k) = \bigsqcup_{m=0}^{k-1} \bigsqcup_{i=1}^{k-m} \{(k-m-i, i)\}.$$

Thus, Equation (2.8) becomes, using Proposition 2.4.1:

$$\mathfrak{L}_{k,\lambda,\gamma}(u) = \sum_{m=0}^{k-1} \sum_{i=1}^{k-m} \frac{\lambda(k-m-i)i}{u + \lambda(k-m-i)i + \gamma i} \sum_{\mathbf{i} \in \mathcal{I}_k(1,m,i)} \prod_{j=0}^m q_{k,\lambda,\gamma}(\mathbf{i}, j; u) g_{k,m,\lambda,\gamma}(\mathbf{i}, j; u)$$

and the conclusion follows by the definition of  $\mathcal{I}_k(m)$  and  $c_{k,\lambda,\gamma}(\mathbf{i}; u)$ . Similarly,

$$\mathfrak{G}_{k,\lambda,\gamma,\beta_G}(u) = \sum_{m=0}^{k-1} \sum_{i=1}^{k-m} \frac{\beta_G i}{u + \lambda(k-m-i)i + \gamma i} \sum_{\mathbf{i} \in \mathcal{I}_k(1,m,i)} \prod_{j=0}^m q_{k,\lambda,\gamma}(\mathbf{i}, j; u) g_{k,m,\lambda,\gamma}(\mathbf{i}, j; u)$$

and one concludes using the definition of  $\mathcal{I}_k(m)$  and  $c'_{k,\lambda,\gamma,\beta_G}$ .  $\square$

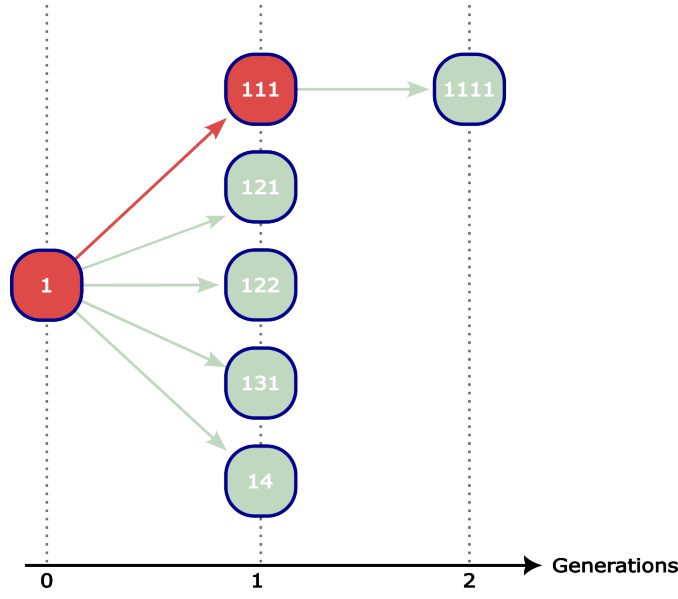
We now have introduced all necessary ingredients allowing for the computation of  $r$ , which is covered in detail in the next section.

## 2.4.2 Characterization of the initial growth rate

Let us now turn to the characterization of  $r$  for the multi-level model, which has been obtained by Pellis et al., 2011. We summarize their arguments for the sake of completeness and in order to illustrate how Corollaries 2.4.2 and 2.4.3 complement their approach.

Their main idea for computing the real time growth rate consists in considering the epidemic at the level of households instead of the individual level. Indeed, it is possible to reduce the epidemic dynamics to a two-type process, distinguishing households that have been contaminated *locally* (type  $L$ ), if the first infected of the household contracted the disease at his workplace, or *globally* (type  $G$ ) otherwise. Remember that, during the early phase of an epidemic, every newly contaminated household or workplace will be fully susceptible except for its member who has just been infected. Thus, a household infects another household *globally* whenever one of its contaminated members transmits the disease through the general population. Local transmission, on the other hand, occurs in the following way. Every time a member of a contaminated household  $H_1$  is infected during its within-household epidemic, a within-workplace epidemic is started at his workplace. Then again, each coworker who is infected during the within-workplace epidemic introduces the disease into his household, which is regarded as *locally* contaminated by household  $H_1$ . A slight subtlety is worth noticing: whether the first infected individual of a household participates in locally contaminating other households depends on the way he has been infected. If he has contracted the disease at his workplace, then he is not hold responsible for the the within-workplace epidemic there, and thus the other households that are infected through his workplace are not considered locally infected by his household. However, the opposite happens if he was infected through the general population, because he then launches a new within-workplace epidemic. This two-type process is depicted in Figure 2.11.

Suppose now that the epidemic is in its exponential growth phase, meaning that there exists a growth rate  $r$  such that the number of infected individuals at time  $t$  is proportional to



**Figure 2.11:** Illustration of the two-type household epidemic process. The example is the same as in Figure 2.4. The generations correspond here to generations of infected households, where the labels denote the first infected member of each household consistently with Figure 2.4. The colors of the arrows and of the households represent the type of infection, either globally (red) or locally (light green).

$e^{rt}$ . Pellis et al., 2011 argue that in this case, the household epidemic also grows exponentially at the same rate  $r$ . It thus is enough to study the previously introduced two-type process.

For  $x, y \in \{L, G\}$ , let  $\zeta_{xy}(t)$  denote the average rate at which a household of type  $y$  which was infected  $t$  units of time ago contaminates other households either locally if  $x = L$ , or globally if  $x = G$ . For  $u \geq 0$ , consider the matrix

$$K(u) = \begin{pmatrix} \mathfrak{L}\zeta_{GL}(u) & \mathfrak{L}\zeta_{GG}(u) \\ \mathfrak{L}\zeta_{LL}(u) & \mathfrak{L}\zeta_{LG}(u) \end{pmatrix}. \quad (2.10)$$

It is a classical result (Diekmann and Heesterbeek, 2000; Pellis et al., 2011) that for this two-type setting, the growth rate  $r$  is characterized as being the unique solution of the implicit equation

$$\rho(K(r)) = \frac{1}{2} \left( \text{Tr}(K(r)) + \sqrt{\text{Tr}(K(r))^2 - 4 \det(K(r))} \right) = 1, \quad (2.11)$$

where the operators  $\rho$ ,  $\text{Tr}$  and  $\det$  denote the spectral radius, trace and determinant, respectively.

It thus only remains to take a closer look at  $\zeta_{xy}(t)$  for all  $x, y \in \{L, G\}$ . As these are average rates of infection, one has to take into account the probability for a newly infected individual to belong to a household or workplace of a given size. Naturally, as households and workplaces are chosen independently uniformly at random, size-biased distributions appear both in the case of global and local infections. This leads us to introduce the following notation. For any application  $f : (n, z) \mapsto f_n(z)$  on  $\mathbb{N} \times \mathbb{R}$  and measure  $\nu$  on  $\mathbb{N}$ ,  $\nu(f_\bullet)$  defines the function on  $\mathbb{R}$  such that  $\nu(f_\bullet) : z \mapsto \sum_n \nu(n) f_n(z)$ .

Within a household of size  $k$ , by definition, the average rate at which global transmissions occur is  $\zeta_{k, \lambda_H, \gamma, \beta_G}^G$ . Since a newly locally or globally contaminated household is of size  $k$

with probability  $\hat{\pi}_k^H$ , it follows that

$$\zeta_{GG} = \zeta_{GL} = \hat{\pi}^H(\zeta_{\bullet, \lambda_H, \gamma, \beta_G}^G).$$

On the other hand, within a globally contaminated household of size  $n$ , new cases appear at rate  $\zeta_{k, \lambda_H, \gamma}$ . Once infected, each of these individual transmits the disease to his coworkers on average at rate  $\zeta_{w, \lambda_W, \gamma}$ , where  $w$  is the appropriate workplace size. This also applies to the globally infected individual who launched the within-household epidemic. If the household is contaminated locally, the reasoning is the same, except that the initially infected member is not hold responsible for his workplace epidemic, as he is himself a secondary case only. As a consequence,

$$\begin{aligned}\zeta_{LG} &= \hat{\pi}^W(\zeta_{\bullet, \lambda_W, \gamma}) + \hat{\pi}^H(\zeta_{\bullet, \lambda_H, \gamma}) * \hat{\pi}^W(\zeta_{\bullet, \lambda_W, \gamma}) \\ \zeta_{LL} &= \hat{\pi}^H(\zeta_{\bullet, \lambda_H, \gamma}) * \hat{\pi}^W(\zeta_{\bullet, \lambda_W, \gamma}).\end{aligned}$$

Using standard properties of Laplace transforms,  $K(u)$  thus admits the following expression as derived by Pellis et al., 2011, where the coefficients can now be computed using Corollaries 2.4.2 and 2.4.3:

$$K(u) = \left( \begin{array}{c} \hat{\pi}^H(\mathfrak{G}_{\bullet, \lambda_H, \gamma, \beta_G})(u) \\ \hat{\pi}^H(\mathfrak{L}_{\bullet, \lambda_H, \gamma})(u) \hat{\pi}^W(\mathfrak{L}_{\bullet, \lambda_W, \gamma})(u) \end{array} \right) \left( 1 + \hat{\pi}^H(\mathfrak{L}_{\bullet, \lambda_H, \gamma})(u) \hat{\pi}^W(\mathfrak{L}_{\bullet, \lambda_W, \gamma})(u) \right). \quad (2.12)$$

Numerical methods then allow to solve the implicit Equation (2.11) for the exponential growth rate  $r$ . We refer to 2.C for details on the computation procedure.

### 2.4.3 ODE reduction of the multilevel model based on the initial growth rate

The reduced model is a standard deterministic *SIR* model, with infection rate derived from the growth rate  $r$ , obtained from Equation (2.11). The reduction of the model is defined by the following set of ordinary differential equations:

$$\begin{cases} \frac{dS}{dt} = -(r + \gamma)SI \\ \frac{dI}{dt} = (r + \gamma)SI - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases} \quad (2.13)$$

The prediction accuracy of the structured model by the reduced model is illustrated in Figure 2.12, where 1% of individuals are initially infected. In this example, we compare simulated epidemics of the stochastic structured model with scenario 1 from Table 2.3 and reference household and workplace size distributions to the ODE reduction from Table 2.2. The reduced model, based on the initial growth of infected population, accurately predicts, as expected, the early stages of the epidemic. We note however, that the prediction accuracy decreases with time and that the epidemic peak and the final size are overestimated by the reduced model.

Further explorations, using all scenarios and exploratory workplace distributions A, indicate the same trends. Comparison of epidemic outcomes such as the epidemic peak and the

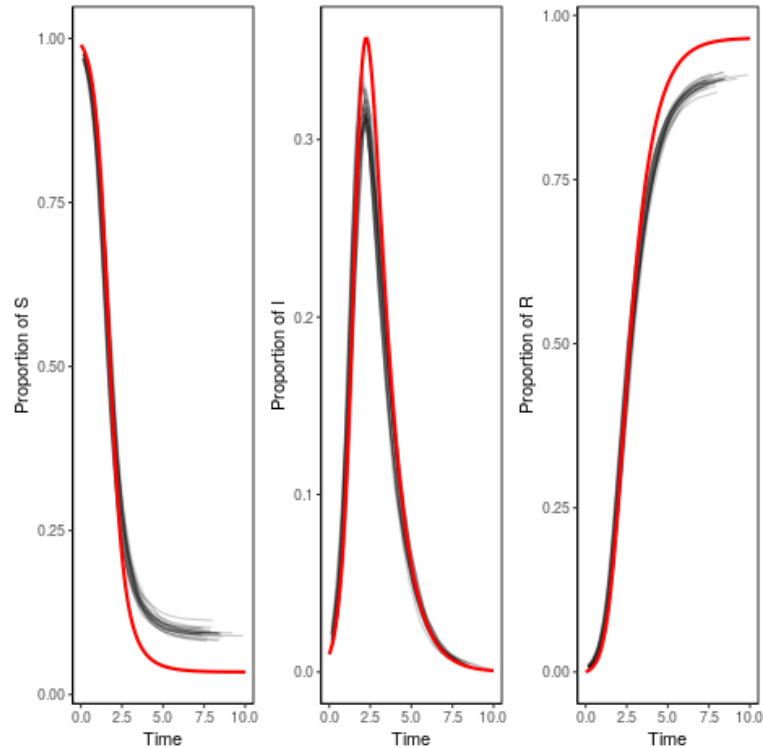
final size, between simulations of the stochastic structured model and their reduced counterpart are presented in Figure 2.13. We observe that, regardless of the scenario, the peak size and final epidemic size are largely correlated between the reduced model and numerical simulations of the stochastic structured model. A slight tendency to overestimate the epidemic outcomes with the reduced model can also be noticed. However, Figure 2.13 (top) illustrate that the epidemic outcomes of the reduced model remain close to the simulations of the stochastic structured model, with differences from the exact model simulations of less than 5% of the total number of individuals with the reduced model. The overestimation is more visible concerning the final epidemic size. This figure also shows that the epidemic parameters influence the prediction quality with the reduced model. This figure also illustrates the effect of the values of  $r$  and  $p_G$  on the prediction of the final epidemic size with the reduced model. Higher values of  $r$ , combined with low value for  $p_G$  (or high  $p_W$ ), such as in scenario 1 and 2, tend to decrease the prediction accuracy with the reduced model. For lower values of the growth rate, which also corresponds to lower epidemic final size, the prediction accuracy is high, see for example scenarios 6 to 11.

The quality of the prediction is expected to be good for high values of  $p_G$ , as for  $p_G = 1$  the reduced model is strictly equivalent to the deterministic approximation of the stochastic *SIR* model without structures. As the value of  $p_G$  decreases, propagation of the epidemic through households and workplaces becomes dominant and the structured model cannot be approximated by a uniformly mixing deterministic *SIR* model regardless of the growth rate, as illustrated in Figure 2.13 (bottom). This figure shows that the epidemic outcomes for the stochastic structured model cannot be obtained by any deterministic *SIR* model for scenarios with lower values of  $p_G$  such as scenario 1 and 2, which deviate from the black line representing the possible outcomes with a deterministic *SIR* model. In addition, Figure 2.13 (top) shows that, for scenarios with lower epidemic size, the prediction of epidemic outcomes by the reduced model is more accurate. This figure also illustrates that the peak size is accurately predicted by the reduced model in these cases, while the prediction of the final size is less accurate. This provides further evidence that the reduction captures the early phase of the epidemic.

#### 2.4.4 Assessment of the reduction robustness

So far, in the modeling and for the simulations and computations, we have restricted ourselves to Markovian models and to the *SIR* structure. The Markovian assumption simplifies the parametrization: there is one single rate of infection for each of three infection ways and one single rate of recovery. This assumption is not necessary and most of our results can be formalized in more general contexts. Similarly, the *SIR* model is the canonical one, but models with different structures can be envisaged. We have checked by simulation that the reduction procedure based on the initial growth rate, as defined in Section 2.4.3 and in 2.F, still provides good predictions in more general contexts. In Figure 2.14, we illustrate through two examples the prediction of epidemic outcomes between the structured model with Gamma distributed individual recovery times and the corresponding reduced deterministic *SIR* model. We can make similar observations as for the Markovian model that the reduced model generally provides a good prediction of the epidemic outcomes, even though the predictions are less accurate in some cases. Notice that the loss of accuracy was to be expected, as the reduced model differs both in terms of contact structure, and



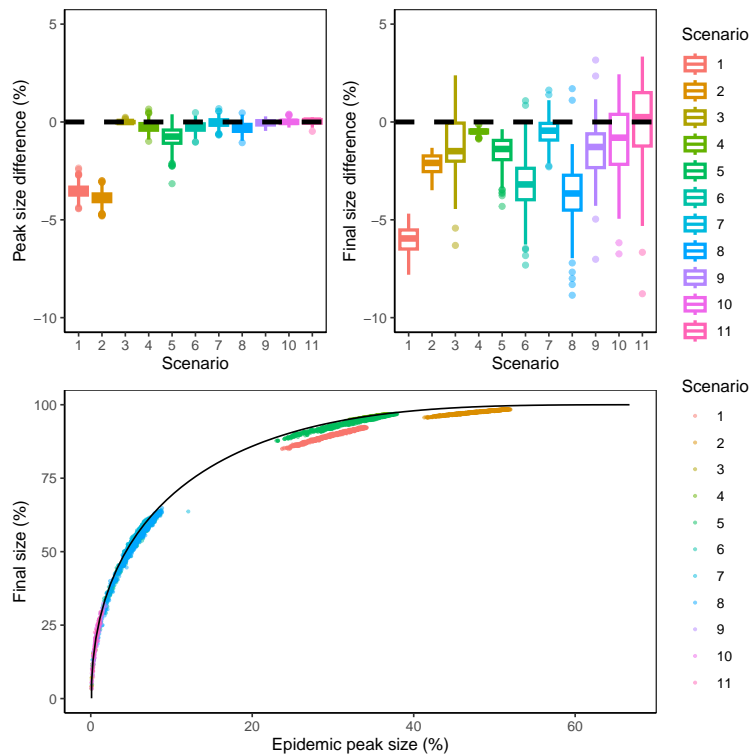


**Figure 2.12:** Evolution of the proportion of  $S$ ,  $I$ ,  $R$  individuals in 30 runs of the stochastic structured model (black), and the reduced model (red). Simulations are performed in a population with 100,000 individuals, with the reference household and workplace size distributions and epidemic parameters from scenario 1. The origin of time of the numerical simulation has been set to the time where 1% of individuals are infected. The simulation of the reduced model is performed with 1% of initially infected individuals.

by the choice of the distribution of the infectious period (exponential distribution instead of Gamma distribution). We also show with an example provided in Figure 2.15 that the reduction procedure provides good predictions of epidemic outcomes for an  $SEIR$  model, where an additional latent (and non infectious)  $E$  state is added to the  $SIR$  model. Details on the growth rate derivation for the  $SEIR$  model are provided in 2.E, while computational aspects are addressed in 2.F.

## 2.5 Discussion

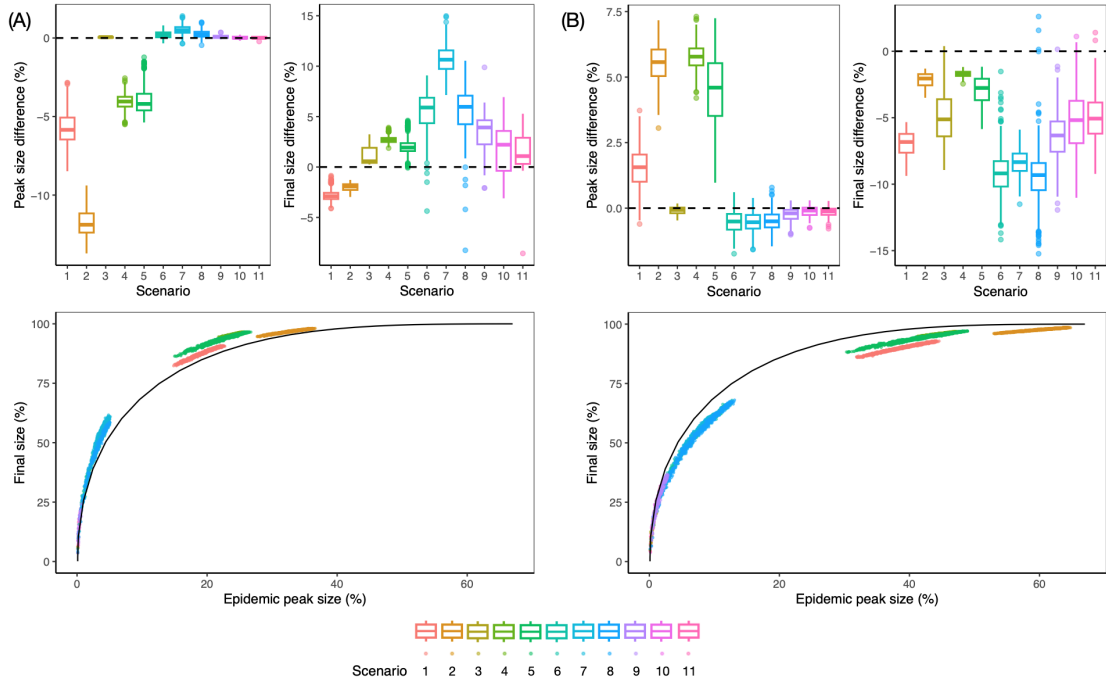
In this work, we study the effect of the size of closed structures (households, workplaces, schools...) on the propagation of epidemics, when individuals belong to different structures, chosen independently for a given individual. We assume that there is no dependence between the size of households and the size of workplaces. Our motivation comes in particular from the fact that control policies allow to change the distribution of structure sizes in various ways, for example by reducing the size of workplaces by teleworking or the size of schools by their (partial) closure. Optimizing control measures in terms of sanitary outcome has become a major challenge. Measuring the link between a social organisation, in terms of distribution



**Figure 2.13:** Boxplot of the differences in the peak sizes (top left) and final sizes (top right) between the complete and the reduced models, obtained from simulations of the stochastic structured model and the corresponding reduced model. Simulated final size as a function of the peak size is reported (bottom), the black line represents the value for the SIR model (2.13). Simulations of the stochastic structured model are performed with population size of 100,000, reference household distribution, workplace size distributions set A, for all scenarios from Table 2.3 (point color). Each combination of scenario and workplace size distribution from set A is repeated 10 times. Only simulations where an epidemic outbreak occurred (i.e. more than 3% of the population become infected) are reported in this figure.

of structure sizes, and the epidemic outcomes is a delicate issue. Indeed, explicit formula and low dimensional large population approximations are known only in uniformly mixing populations. These results concern explicit values for  $R_0$  and initial growth rate  $r$ , simple equation for the total size of the outbreak, reduction to an  $SIR$  three dimensional ODE with two parameters. Formulae become more complex when adding a local level consisting in one layer of structures (households), and even intractable with an additional structure like workplaces, where the chain of transmission does not need mean field infection any longer to propagate.

We have used both existing approaches (namely Pellis et al., 2011) and added new developments, by providing more explicit expressions (Section 2.4) and simulations to study the role of structure sizes on the key outcomes of epidemics. Notice that even though the results of Corollaries 2.4.2 and 2.4.3 have not been used in this paper for evaluating the growth rate of simulation scenarios (details on computations in 2.B), their numerical implementation would have been an alternative method for achieving these computations.

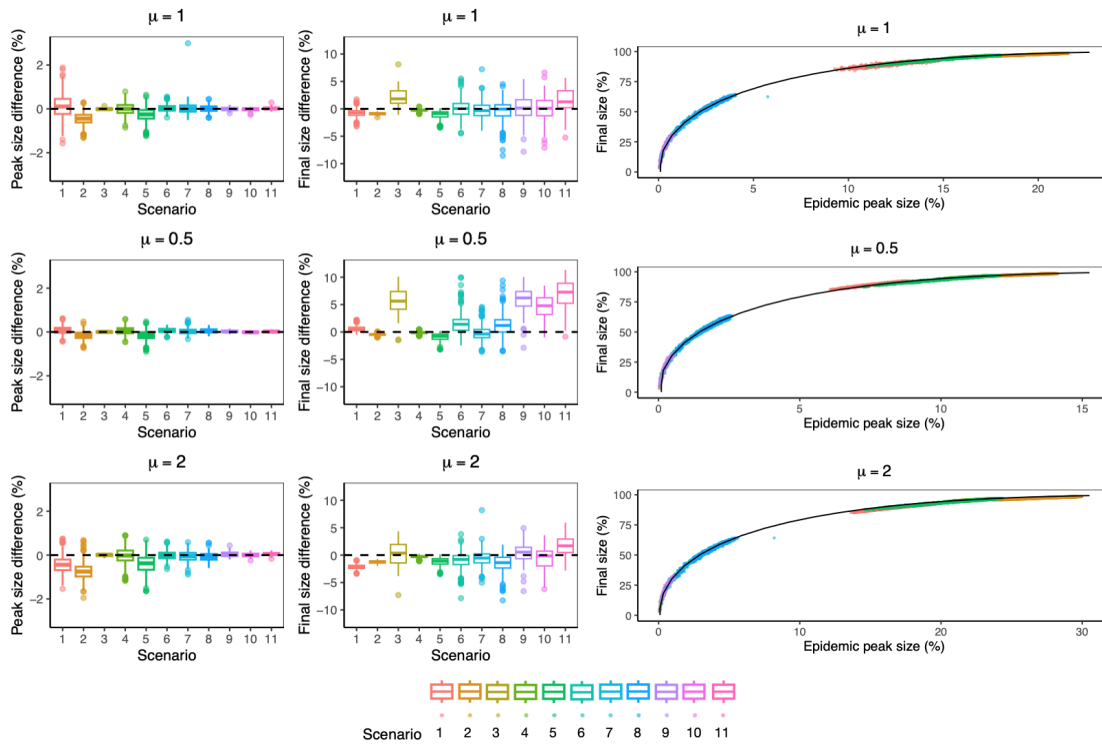


**Figure 2.14:** Boxplot (top) of the differences in the peak sizes and final sizes between the complete and the reduced models, obtained from simulations of the stochastic structured SIR model, with gamma distributed recovery rates. Simulated final size as a function of the peak size is reported (bottom), the black line represents the value for the SIR model (2.13). The shape parameter  $a$  and rate parameter  $r$  are set to values of 0.5 (panel A) or 2 (panel B), i.e. the density is given by  $(r^{-a}\Gamma(a))^{-1}x^{a-1}e^{-rx}$ . The average recovery rate is 1. Simulations of the stochastic structured model are performed with population size of 100,000, reference household distribution, workplace size distributions set A, for all scenarios from Table 2.3. Each combination of scenario and workplace size distribution from set A is repeated 10 times. Only simulations where an epidemic outbreak occurred (i.e. more than 3% of the population become infected) are reported in this figure.

In our setting, this would not be numerically pertinent as workplace sizes can be relatively large and the formulae obtained in Corollaries 2.4.2 and 2.4.3 need to iterate over all elements in a set whose cardinal grows exponentially with the structure size. However, making use of these results may be pertinent when computing the growth rate in models considering households only, as those typically are of smaller sizes.

We have focused on the Markovian case, i.e. time of infection exponentially distributed and constant infection rate, and a simple SIR structure, with local infection rates proportional to the number of susceptible and infected individuals. This basic framework was complemented by a robustness study on more complex settings (Sections 2.3.4 and 2.4.4). According to our findings, the structure size distribution plays a role on the key outcomes in most scenarios. More precisely, for a given number of structures and a given number of individuals and thus for fixed average structure size, the way individuals are distributed has a quantitative impact on the growth rate of infections, the total number of infected individuals and the size of the infected peak. In this setting, the variance of the structures size distribution provides a good proxy of this impact.

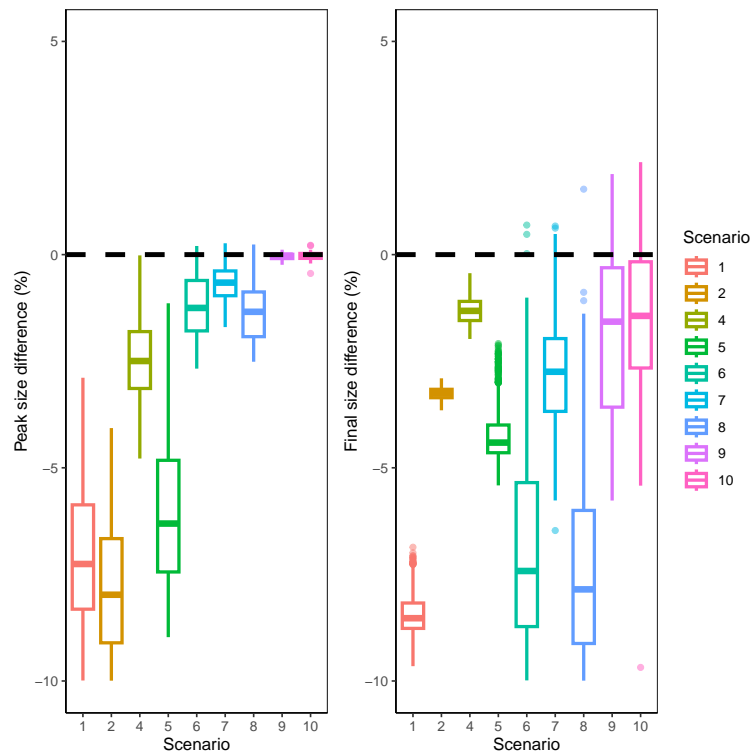
This finding may be related to previously known results on the importance of the variance



**Figure 2.15:** Boxplot of the differences in the peak sizes (left) and final sizes (middle) between the complete and the reduced models, obtained from simulations of the stochastic structured SEIR model and the corresponding reduced model defined in Appendix 2.F, for values of the transition rate from E to I,  $\mu = 1$ ,  $\mu = 0.5$  and  $\mu = 2$ . (Right) Plot of the epidemic final size as a function of the epidemic peak sizes obtained from simulations of the stochastic structured SEIR model, for values of the transition rate from E to I,  $\mu = 1$ ,  $\mu = 0.5$  and  $\mu = 2$ . The black line represents the epidemic peak size and epidemic final size for the standard SEIR model.

For all panels, simulations of the stochastic structured model are performed with population size of 100,000, reference household distribution, workplace size distributions set A, for all scenarios from Table 2.3. Each combination of scenario and workplace size distribution from set A is repeated 10 times. Only simulations where an epidemic outbreak occurred (i.e. more than 3% of the population become infected) are reported in these figures.

of the degree distribution in configuration model settings. Indeed, Britton et al., 2007 have pointed out the impact of the degree distribution variance on the reproduction number for *SIR* epidemics on configuration graphs. Similarly, Ma et al., 2013 have studied the case of a model with two levels of mixing, corresponding to a layer of households and a general population taking the form of a configuration graph. They have shown that in this case, the variance of the degree distribution in the general population has a strong influence on epidemic dynamics. In the case of the household-workplace model, we can consider the epidemic at the household level. In this situation, the size distribution of the workplace plays a crucial role in determining the number of households to which a given household is directly connected. Although the framework is clearly more complex than a simple configuration model, the distribution of workplace size may play a role similar in spirit to that of the degree distribution in the previously mentioned models based on configuration graphs.



**Figure 2.16:** Boxplot of the differences in the peak sizes and (left) and final sizes (right) between the complete and the reduced models, obtained from simulations of the stochastic structured model with sub-linear rate and the corresponding reduced model. Simulations of the stochastic structured model are performed with population size of 100,000, reference household distribution, workplace size distributions set A, for all scenarios from Table 2.3. Each combination of scenario and workplace size distribution from set A is repeated 10 times. Only simulations where an epidemic outbreak occurred (i.e. more than 3% of the population become infected) are reported in this figure.

As for the limitations of our study, the robustness analyses that we have carried out (Sections 2.3.4 and 2.4.4), even if they are rather summary, point to robustness of the main results when certain assumptions are modified. A more detailed analysis using sensitivity analysis is left for a future work.

On the one hand, the effect of the structure size distribution may be overestimated by the fact that we consider a linear infection rate. Indeed, the rate at which a susceptible individual is infected at a given time is assumed to be proportional to the number of infected individuals at the same time. This probably overestimates the real infection rate. We consider here structures which form a partition of the population, with uniform mixing within each structure. We think this assumption is rather relevant for households but it could be improved and extended for workplaces (and schools), where different levels of mixing could be considered (services/departments within companies, classes within schools, etc). The simulation study with sublinear infection rates in households and workplaces shows that the observation on the linear impact of variance on the final epidemic size is still valid (Figure 2.10). With sublinear infection rates, we also observed similar results regarding the robustness of the prediction of epidemic outcomes using the reduced model, namely that the reduction

is accurate for high values of  $p_G$  and a smaller epidemic size, see Figure 2.16. We note however, that the accuracy of the prediction using the reduced model is lower with sublinear infection rates, and the early epidemic is not predicted as accurately. When it comes to policies and controls such as teleworking, a notion of effective size and/or effective infection rate should probably be introduced, which is one of the interesting perspectives.

On the other hand, the effect of the structure size distribution is observed when looking at the characteristics of the epidemic. Variance arises using the size biased law and the fact that mean final size of epidemics is comparable to the size structure. Furthermore, following preliminary explorations, we observed that the dependence in the structure size distribution may imply a higher moment of it than the second moment linked to its variance. This could confer a greater impact of large structures than the variance would predict. This reinforces the importance of the structure size distribution and specifically the importance of the health benefit of moving toward structures (workplace offices, school classrooms...) of the same size. Variance remains the simplest proxy we have found in general, but further exploration may reveal another form of dependence.

Thus, based on our results, the variance of the structure size distribution, and more generally the ratio of the second to the first moment when the latter or the number of structures is not fixed, provides a good indicator to measure the impact of the distribution of individuals within small social structures. Nevertheless, much work remains to be done to identify the key parameters related to population structure that determine epidemic outcomes.

In this study, we are also interested in model reduction, in order to have a parsimonious model that is sufficiently accurate in terms of prediction and fast to run. This is an important issue, especially in a context where many scenarios and control policies need to be evaluated. Thus, our goal was to reduce the model to only a few parameters and variables. We obtained that by using the initial infection growth rate, which keeps track of the contact structure in a subtle (almost explicit but rather complex) way. We then make use of a classical *SIR* model (three variables and two parameters) as a reduction of the initial stochastic individual-centered model with three types of contacts.

In summary, our study highlights that knowledge and modeling of the size of contact structures appears to be important in characterizing the outcome of an epidemic. It also points to the major role of the initial growth rate of the infection, which is unique, contrary to the reproduction number which can have several interpretations in multilevel contact models, and also difficult to treat. We have tried to provide a more explicit expression for the initial growth rate, supplementing the literature. This allowed us first to conduct the study of the impact of the structure size distribution. As we already know, the initial growth rate provides the extinction/outbreak criterion by its sign and the rate of progression. It is related to the peak and timing of the peak, when the epidemic dynamics explode, and to the extinction rate when this dynamics decline. The initial growth rate has also been a key input to the reduction problem, allowing a complex contact structure to be reduced to a few parameters, while retaining the essence of the qualitative and quantitative behavior of the epidemic beyond the initial time.

## Appendix

### 2.A Generation of structure size distributions

Let us start by giving detail on the way the reference workplace size distribution is derived from the INSEE French workplace size distribution of 2018. Indeed, the reference workplace size distribution is obtained from the number of workers in workplace size classes provided by INSEE. On the one hand, workers belonging to workplaces in size classes lower than 50 are placed, for our workplace size distribution, in workplaces with size uniformly chosen in the size range covered by the size class. On the other hand, workers belonging to workplaces of over 50 employees are arbitrarily placed in workplaces of size 50 in our workplace distribution, as we assume uniform mixing within workplaces which becomes unrealistic for very large workplace sizes.

Finally, in order to generate distributions with given mean  $m$ , variance and maximum size  $n_{\max}$ , we proceed as follows. First, we create a set  $\mathcal{P}$  of distributions which each charge only two sizes in  $\{1, \dots, n_{\max}\}$  and mean  $m$ . In other words, for any  $1 \leq k < m < k' \leq n_{\max}$ , we define  $p = c\delta_k + (1 - c)\delta_{k'}$  where  $c = (k' - m)/(k' - k)$ . Second, we construct a new set of distributions  $\mathcal{D}$  from mixtures of two distributions of the previous set, such that the resulting distribution has a given variance. An element  $d \in \mathcal{D}$  hence is obtained by taking  $p_1, p_2 \in \mathcal{P}$  and letting  $d = cp_1 + (1 - c)p_2$ , where the weight  $c$  is chosen such that  $d$  has the prescribed variance. A target distribution  $D$  is then generated by a random mixture of those elementary distributions:

$$D = \sum_{d \in \mathcal{D}} w_d d,$$

where  $w_d$  are random weights such that  $\sum_{d \in \mathcal{D}} w_d = 1$ .

### 2.B Parameter values for scenarios

The size distribution of households and workplaces combined with structure-dependent infection rates determines a value for the initial growth rate of the epidemic  $r$ , as well as probabilities of infections corresponding to the three sources, global mixing  $p_G$ , households  $p_H$  and workplaces  $p_W$ . Due to the constraints imposed by the structure size distributions in the structured epidemic model, it is not always possible to find numerical values of infection rates that lead to given values of growth rate  $r$  and infection probabilities  $p_G$ ,  $p_H$  and  $p_W$ . Parameter selection for scenarios in Table 2.3 was performed, for the reference household and workplace distributions, using an optimisation procedure that yields infection rates leading to a solution for growth rate and infection probabilities values as close as possible to the target values. It relies on a cost function based on the mean square error between the target values and the trial values of  $r$ ,  $p_G$ ,  $p_H$  and  $p_W$ . A hyper-parameter controls the importance given to the error on the growth rate.

Table 2.4 summarizes the values of  $r$ ,  $R_I$ ,  $p_G$ ,  $p_H$  and  $p_W$  for each scenario of Table 2.3, as well as the obtained values of epidemic parameters for the reference structure size distributions.

Scenarios provided in Table 2.3 combined with the workplace distributions from Table 2.2 allow the exploration of the relevant behavior of the structured epidemic model, covering

**Table 2.4:** Values of growth rate, reproduction number and proportions of infection per layer for each simulation scenario, as well as the corresponding epidemic parameters, in the case of the reference structure size distributions (recall that  $\gamma = 1$  in all scenarios).

Scenario	Growth rate	$R_I$	$p_H$	$p_W$	$p_G$	$\lambda_H$	$\lambda_W$	$\beta_G$
1	2.4822	2.5028	0.4217	0.1788	0.3995	11.852	0.009	1.000
2	4.9937	4.6876	0.2796	0.3471	0.3732	11.443	0.020	1.750
3	0.0009	0.9923	0.4070	0.3927	0.2002	0.376	0.010	0.199
4	2.5203	3.6460	0.1521	0.1472	0.7008	0.356	0.010	2.555
5	2.5017	5.2021	0.1301	0.5445	0.3254	0.451	0.028	1.693
6	0.5054	1.5740	0.3868	0.3456	0.2676	0.653	0.012	0.421
7	0.5061	1.5187	0.1120	0.1122	0.7758	0.094	0.004	1.178
8	0.5020	1.5920	0.3975	0.4057	0.1968	0.730	0.013	0.313
9	0.1104	1.1330	0.3879	0.4182	0.1940	0.404	0.012	0.220
10	0.0900	1.0989	0.2631	0.2646	0.4723	0.196	0.007	0.519
11	0.0706	1.1068	0.3449	0.5883	0.0668	0.309	0.016	0.074

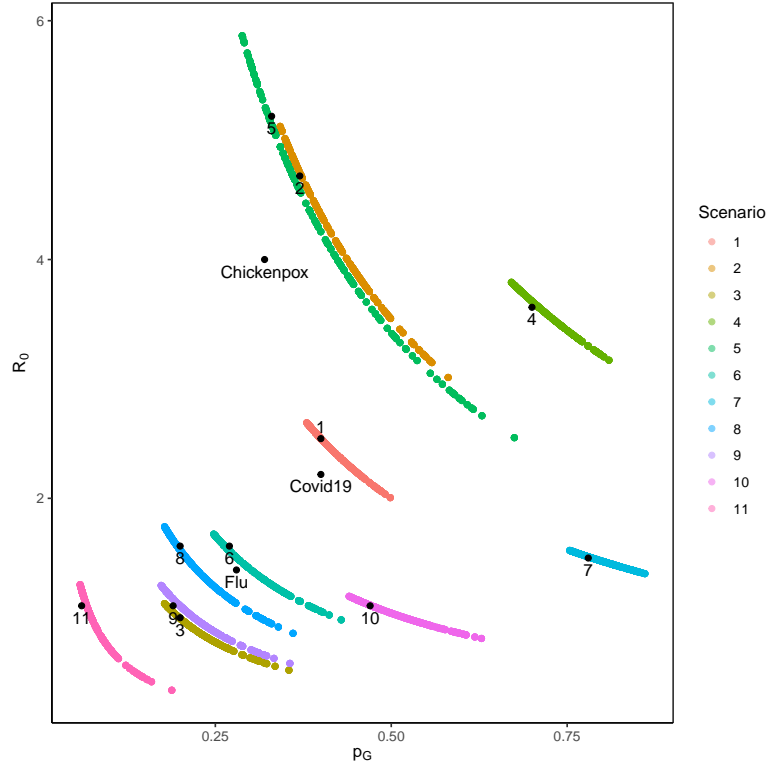
a wide range of epidemic settings. The epidemic final size for each scenario and workplace distribution is reported in Figure 2.2.

As mentioned in Section 2.2.3, our scenarios correspond to, or are close to, realistic epidemic settings for three diseases of interest: influenza (Ajelli et al., 2014), COVID-19 (Locatelli et al., 2021; Galmiche et al., 2021) and chickenpox (Silhol and Boëlle, 2011). Notice that while comparing proportions of infection at the local and global level is straightforward, the task is more delicate for reproduction numbers. As Ajelli et al., 2014 and Locatelli et al., 2021 infer the reproduction number based on the exponential growth rate, we have followed this approach and based the comparison on the reproduction number  $R_0$  defined as  $R_0 = 1 + r/\gamma$ , where  $r$  is the exponential growth rate (Trapman et al., 2016). Of course, Ajelli et al., 2014, Locatelli et al., 2021 and Silhol and Boëlle, 2011 do not use this precise definition of  $R_0$ , but it seemed the best compromise for comparison as it is closest in spirit to the the studies on influenza and COVID-19, while the study on chickenpox unfortunately does not detail their definition of  $R_0$ . The results are shown in Figure 2.17, which illustrates that our procedure covers a wide range of realistic epidemic settings.

## 2.C Numerical computations of epidemic parameters and outcomes

The values of  $R_I$ ,  $p_G$ ,  $p_H$  and  $p_W$  are obtained from Equations (2.2) and (2.3). They both require the values of  $\mathcal{I}_G$ ,  $\mathcal{I}_G$ ,  $\mathcal{I}_W$  which are given by Equation (2.1). Evaluations of  $\mathcal{I}_G$ ,  $\mathcal{I}_G$ ,  $\mathcal{I}_W$  require the values of  $i_H(k)$  and  $i_W(k)$  which we obtained from numerical simulations of the within structures epidemic. The growth rate  $r$  can be obtained in several ways, which all involve some form of solution for Equation (2.11), which we solved using a root finding algorithm. Elements of Matrix (2.10) can be obtained by numerical simulation of within structure epidemics, numerical integration and numerical matrix inversion. We also provide analytical formulations for the *SIR* and *SEIR* models, with linear infection rates, in Equations





**Figure 2.17:** Values of the reproduction number  $R_0$ , and proportion of infection via global mixing  $p_G$  for all scenarios of Table 2.3 and all exploratory workplace size distributions from set A, Table 2.2. The household size distribution is the reference household size distribution. For each scenario 1-11, labeled black points correspond to the values achieved for the reference workplace distribution. For comparison, values reported in the literature for COVID-19, influenza and chickenpox are also shown in black.

(2.8) and forthcoming Equation (2.17). Alternatively, we provide a fully analytical way to obtain the growth rate of the  $SIR$  model with linear infection rates, in Corollaries 2.4.2 and 2.4.3. The values of the epidemic peak and size are obtained by stochastic simulation of the structured model, and the values of the epidemic peak and size of the reduced model are obtained by simulation of Equation (2.13). For a summary of the numerical evaluation of epidemic quantities, see Table 2.5.

## 2.D Proof of Proposition 2.4.1

The notations are the same as previously introduced in Section 2.4.1. For two integers  $p, q$ , we further let  $I_p$  denote the identity matrix of dimension  $p$ , and  $M_{p,q}(\mathbb{R})$  (resp.  $M_p(\mathbb{R})$ ) the space of  $p \times q$  (resp.  $p \times p$ ) matrices with real coefficients. The aim is to compute

$$\hat{\mathbf{Q}}_{k,\lambda,\gamma}(u) = (uI_{d(k)} - \mathbf{Q}_{\lambda,\gamma}(k))^{-1}$$

for  $u \geq 0$ , which allows us to obtain Corollaries 2.4.2 and 2.4.3.

*Proof of Proposition 2.4.1.* We start by noticing that when  $k = 1$ , necessarily,  $\ell = 1$ ,  $(s, i) =$

**Table 2.5:** Numerical methods used to compute epidemic parameters and outcomes.

Epidemic parameter/outcome	Numerical method
$\mathcal{I}_G, \mathcal{I}_H, \mathcal{I}_W$	Equation (2.1), with $i_H(k)$ and $i_W(k)$ obtained from numerical simulations of the within structure epidemic.
$R_I$	Largest eigenvalue of Equation (2.2).
$p_G, p_H$ and $p_W$	Equation (2.3).
Growth rate $r$	Equation (2.11) with Laplace transforms for elements of matrix (2.10) which can be numerically evaluated. Alternatively, in the case of the <i>SIR</i> (or <i>SEIR</i> ) model with linear infection rates, these elements are defined in Equation (2.8) (resp. (2.17)), where matrices in the sum are inverted numerically.
Peak and final size	Numerical simulation of the structured epidemic model or numerical simulation of Equation (2.13) for the reduced epidemic model.

$(0, 1)$  and  $m = 0$ . The set of interest becomes  $\mathcal{I}_1(1, 0, 1) = \{(1, 1)\}$ , and the right side of Equation (2.9) equals  $(u + \gamma)^{-1}$ . On the other hand, it is obvious that  $(uI_1 - \mathbf{Q}_{\lambda, \gamma}(1)) = (u + \gamma)$ . Thus, (2.9) holds when  $k = 1$ .

Suppose now that Equation (2.9) is true for some integer  $k$ . We proceed by induction. Notice that it is possible to enumerate the states of  $\Omega(k+1)$  in such a way that  $\mathbf{Q}_{\lambda, \gamma}(k+1)$  is upper triangular. Indeed, it is enough to enumerate first all states in  $\widehat{\Omega}(k+1) = \{(s, i) \in \Omega(k+1) : s + i = k+1\}$  as  $\{(k+1-\ell, \ell) : 1 \leq \ell \leq k+1\}$ , so that progression from one state to another occurs by infections which are not reversible as individuals are immune after infection. This process is repeated for states  $\widehat{\Omega}(m) = \{(m-\ell, \ell) : 1 \leq \ell \leq m\}$  for  $m = k, \dots, 1$ , so that transition from one set of states to the next occurs by removal of an infected, which also is irreversible as individuals will remain immune afterwards.

This way of enumerating  $\Omega(k+1)$  is particularly interesting as  $\Omega(k+1) = \bigsqcup_{m=1}^{k+1} \widehat{\Omega}(m)$ , so that  $\mathbf{Q}_{\lambda, \gamma}(k+1)$  can be regarded as the following block matrix:

$$\mathbf{Q}_{\lambda, \gamma}(k+1) = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{Q}_{\lambda, \gamma}(k) \end{pmatrix}.$$

Here, blocks  $\mathbf{A} \in M_{k+1}(\mathbb{R})$  and  $\mathbf{B} \in M_{k+1, d(k)}(\mathbb{R})$  represent events of infections and removals in  $\widehat{\Omega}(k+1)$ , respectively, where we recall that  $d(k) = \#\Omega(k) = k(k+1)/2$ . As previously, the elements of  $\mathbf{Q}_{\lambda, \gamma}(k+1)$  can also be indexed by states in  $\Omega(k+1)$ . Naturally,  $uI_{d(k+1)} - \mathbf{Q}_{\lambda, \gamma}(k+1)$  also is a block matrix, simply replacing the blocks  $\mathbf{A}$  and  $\mathbf{Q}_{\lambda, \gamma}(k)$  by  $\mathbf{A}' = uI_{k+1} - \mathbf{A}$  and  $uI_{d(k)} - \mathbf{Q}_{\lambda, \gamma}(k)$  respectively.

More precisely,  $\mathbf{A}'$  is an upper bidiagonal matrix such that

$$\begin{aligned} \forall \ell \in \{1, \dots, k+1\}, \quad \mathbf{A}'_{\ell, \ell} &= \mathbf{A}'_{(k+1-\ell, \ell), (k+1-\ell, \ell)} = u + \lambda(k+1-\ell)\ell + \gamma\ell, \\ \forall \ell \in \{1, \dots, k\}, \quad \mathbf{A}'_{\ell, \ell+1} &= \mathbf{A}'_{(k+1-\ell, \ell), (k+1-(\ell+1), \ell+1)} = -\lambda(k+1-\ell)\ell. \end{aligned}$$

Since  $\mathbf{A}$  is upper diagonal, its inverse matrix is easily computable (Chatterjee, 1974), and

letting  $a_{\ell,i}^{(k+1)} = (\mathbf{A}')_{(k+1-\ell),(k+1-i,i)}^{-1}$ , we have:

$$a_{\ell,i}^{(k+1)} = \mathbf{1}_{i \geq \ell} \frac{1}{u + \lambda(k+1-i)i + \gamma i} \prod_{j=\ell}^{i-1} \left( 1 + \frac{u + \gamma j}{\lambda(k+1-j)j} \right)^{-1}.$$

Furthermore, as the only states of  $\Omega(k)$  that are directly accessible by removal from  $\widehat{\Omega}(k+1)$  belong to  $\widehat{\Omega}(k)$ , all coefficients of  $\mathbf{B}$  are null except for the following:

$$\mathbf{B}_{\ell,\ell-1} = \mathbf{B}_{(k+1-\ell,\ell),(k+1-\ell,\ell-1)} = -\gamma\ell, \quad \forall \ell \in \{2, \dots, k+1\}.$$

Using the fact that  $uI_{d(k+1)} - \mathbf{Q}_{\lambda,\gamma}(k+1)$  is a block matrix and that  $\mathbf{A}'$  and  $uI_{d(k)} - \mathbf{Q}_{\lambda,\gamma}(k)$  are upper triangular with positive diagonal coefficients and thus invertible, we have the following:

$$\widehat{\mathbf{Q}}_{k+1,\lambda,\gamma}(u) = (uI_{d(k+1)} - \mathbf{Q}_{\lambda,\gamma}(k+1))^{-1} = \begin{pmatrix} (\mathbf{A}')^{-1} & -(\mathbf{A}')^{-1}\mathbf{B}\widehat{\mathbf{Q}}_{k,\lambda,\gamma}(u) \\ 0 & \widehat{\mathbf{Q}}_{k,\lambda,\gamma}(u) \end{pmatrix}.$$

Thus, we obtain that, for all  $\ell, i \in \{1, \dots, k+1\}$  and  $(s, i) \in \Omega(n)$ ,

$$\begin{aligned} \left( \widehat{\mathbf{Q}}_{k+1,\lambda,\gamma}(u) \right)_{(k+1-\ell,\ell),(k+1-i,i)} &= a_{\ell,i}^{(k+1)}; \\ \left( \widehat{\mathbf{Q}}_{k+1,\lambda,\gamma}(u) \right)_{(k+1-\ell,\ell),(s,i)} &= \sum_{w=2}^{k+1} a_{\ell,w}^{(k+1)} \gamma w \left( \widehat{\mathbf{Q}}_{k,\lambda,\gamma}(u) \right)_{(k+1-w,w-1),(s,i)}. \end{aligned} \quad (2.14)$$

Let us turn to proving that Equation (2.9) holds true for  $k+1$ . Let  $\ell \in \{1, \dots, k+1\}$  and consider  $(s, i) \in \Omega(k+1)$  such that  $s \leq (k+1) - \ell$  and define  $m = (k+1) - (s+i)$ .

Notice that  $m = 0$  if and only if  $(s, i) \in \widehat{\Omega}(k+1)$  and  $s = k+1-i$ . As a consequence, if  $i < \ell$ , the set  $\mathcal{I}_{k+1}(\ell, 0, i)$  is empty; otherwise, if  $i \geq \ell$ ,  $\mathcal{I}_{k+1}(\ell, 0, i) = \{(\ell, i)\}$ . In both cases, the right hand side of Equation (2.9) equals  $a_{\ell,i}^{(k+1)}$ . Thus, by Equation (2.14), Assertion (2.9) follows.

Consider now the case  $m > 0$ , i.e.  $(s, i) \in \Omega(k)$ . It follows from Equation (2.14), using the change of variable  $\ell' = w - 1$ , that

$$\left( \widehat{\mathbf{Q}}_{k+1,\lambda,\gamma}(u) \right)_{(k+1-\ell,\ell),(s,i)} = \sum_{\ell'=1}^k a_{\ell,\ell'+1}^{(k+1)} \gamma (\ell' + 1) \left( \widehat{\mathbf{Q}}_{k,\lambda,\gamma}(u) \right)_{(k-\ell',\ell'),(s,i)}. \quad (2.15)$$

Using the inductive hypothesis and noticing that  $k - (s+i) = m - 1$ , we get that for any  $\ell' \in \{1, \dots, k\}$ ,

$$\begin{aligned} \left( \widehat{\mathbf{Q}}_{k,\lambda,\gamma}(u) \right)_{(k-\ell',\ell'),(s,i)} &= \\ &= \frac{1}{u + \lambda s i + \gamma i} \sum_{\mathbf{i} \in \mathcal{I}_k(\ell', m-1, i)} \prod_{j=1}^m q_{k,\lambda,\gamma}(\mathbf{i}, j-1; u) g_{k,m-1,\lambda,\gamma}(\mathbf{i}, j-1; u). \end{aligned}$$

Notice that if  $\mathbf{i} \in \mathbb{N}^{m+2}$  is such that  $i_{j-1} = i'_j$  for all  $j$ , then  $q_{k,\lambda,\gamma}(\mathbf{i}, j-1; u) = q_{k+1,\lambda,\gamma}(i', j; u)$  and  $g_{k,m-1,\lambda,\gamma}(\mathbf{i}, j-1; u) = g_{k+1,m,\lambda,\gamma}(i', j; u)$ . Letting  $\tau$  be the projection on  $\mathbb{N}^{m+2}$ : for

any  $\mathbf{i}' = (i_0, \dots, i_{m+1}) \in \mathbb{N}^{m+2}$ ,  $\tau(\mathbf{i}') = (i_1, \dots, i_{m+1})$ , we get that:

$$\sum_{\mathbf{i} \in \mathcal{I}_k(\ell', m-1, i)} \prod_{j=1}^m q_{k, \lambda, \gamma}(\mathbf{i}, j-1; u) g_{k, m-1, \lambda, \gamma}(\mathbf{i}, j-1; u) = \sum_{\substack{\mathbf{i}' \in \mathbb{N}^{m+2}: \\ \tau(\mathbf{i}') \in \mathcal{I}_k(\ell', m-1, i)}} \prod_{j=1}^m q_{k+1, \lambda, \gamma}(\mathbf{i}', j; u) g_{k+1, m, \lambda, \gamma}(\mathbf{i}', j; u).$$

Furthermore, for  $\mathbf{i}' \in \mathbb{N}^{m+2}$  such that  $i_0 = \ell$ ,  $i_1 = \ell'$ , it holds that

$$a_{\ell, \ell'+1}^{(k+1)} \gamma(\ell' + 1) = \mathbf{1}_{\{\ell'+1 \geq \ell\}} q_{k+1, \lambda, \gamma}(\mathbf{i}', 0; u) g_{k+1, m, \lambda, \gamma}(\mathbf{i}', 0; u). \quad (2.16)$$

Thus, noticing that the limits of the sum over  $\ell'$  in Equation (2.15) taken together with  $\mathbf{1}_{\{\ell'+1 \geq \ell\}}$  from Equation (2.16) induce that  $\ell-1 \leq \ell' \leq k$ , Equation (2.15) yields the desired result:

$$\left( \widehat{\mathbf{Q}}_{k+1, \lambda, \gamma}(u) \right)_{(k+1-\ell, \ell), (s, i)} = \frac{1}{u + \lambda s i + \gamma i} \sum_{\mathbf{i}' \in \mathcal{I}_{k+1}(\ell, m, i)} \prod_{j=0}^m q_{k+1, \lambda, \gamma}(\mathbf{i}', j; u) g_{k+1, m, \lambda, \gamma}(\mathbf{i}', j; u).$$

This completes the proof.  $\square$

## 2.E Computation of the exponential growth rate for the SEIR model with two levels of mixing

In the SEIR model, subsequently to an infectious contact between an infected and a susceptible individual, the susceptible first becomes *exposed* ( $E$  state, assimilated to an infected but not yet infectious state) for a duration distributed according to an exponential law of parameter  $\mu$ , before entering the infectious state. Thus, the computation of the exponential growth rate as proposed by Pellis et al., 2011 needs to be adapted.

For a population of size  $k$ , consider the Markov chain giving the numbers  $(S_t, E_t, I_t)_{t \geq 0}$  of susceptible, exposed and infected individuals in the population at time  $t \geq 0$  after the beginning of the epidemic, of transition rates

Transition	Rate
$(s, e, i) \rightarrow (s-1, e+1, i)$	$\lambda s i;$
$(s, e, i) \rightarrow (s, e-1, i+1)$	$\mu e;$
$(s, e, i) \rightarrow (s, e, i-1)$	$\gamma i.$

The set of transient states is then given by

$$\Omega'(k) = \{(s, e, i) \in (\mathbb{N} \cup 0)^3 : s + e + i \leq k, e + i \geq 1\},$$

of cardinal  $d'(k) = k(k+1)(k+5)/6$ . The restriction of the generator of this Markov chain to  $\Omega(k)$  is given by  $Q'_{\lambda, \gamma, \mu}(k)$  defined by:  $\forall (s, e, i), (s', e', i') \in \Omega'(k)$ ,

$$(Q'_{\lambda,\gamma,\mu}(k))_{(s,e,i),(s',e',i')} = \begin{cases} -\lambda si - \mu e - \gamma i & \text{if } (s', e', i') = (s, e, i); \\ \lambda si & \text{if } (s', e', i') = (s-1, e+1, i); \\ \mu e & \text{if } (s', e', i') = (s, e-1, i+1); \\ \gamma i & \text{if } (s', e', i') = (s, e, i-1); \\ 0 & \text{otherwise.} \end{cases}$$

Following the work of Pellis et al., 2011 and adopting the notations introduced in Section 2.4, one can then easily see that the exponential growth rate  $r'$  of the *SEIR* model with two levels of mixing is characterized by the implicit equation

$$\rho(K'(r')) = 1,$$

where for  $u \geq 0$ ,  $K'(u)$  is the following matrix:

$$\begin{pmatrix} \hat{\pi}^H(\mathfrak{G}'_{\bullet,\lambda_H,\mu,\gamma,\beta_G})(u) & \hat{\pi}^H(\mathfrak{G}'_{\bullet,\lambda_H,\mu,\gamma,\beta_G})(u) \\ \hat{\pi}^H(\mathfrak{L}'_{\bullet,\lambda_H,\mu,\gamma})(u)\hat{\pi}^W(\mathfrak{L}'_{\bullet,\lambda_W,\mu,\gamma})(u) & (1 + \hat{\pi}^H(\mathfrak{L}'_{\bullet,\lambda_H,\mu,\gamma})(u))\hat{\pi}^W(\mathfrak{L}'_{\bullet,\lambda_W,\mu,\gamma})(u) \end{pmatrix},$$

given the definition for  $\beta, \lambda, \mu, \gamma > 0$  of

$$\begin{aligned} \mathfrak{G}'_{k,\lambda,\mu,\gamma,\beta}(u) &= \sum_{(s,e,i) \in \Omega'(k)} \beta i ((uI_{d'(k)} - Q'_{\lambda,\gamma,\mu}(k))^{-1})_{(k-1,1,0),(s,e,i)}, \\ \mathfrak{L}'_{k,\lambda,\mu,\gamma}(u) &= \sum_{(s,e,i) \in \Omega'(k)} (\lambda si + \mu e) ((uI_{d'(k)} - Q'_{\lambda,\gamma,\mu}(k))^{-1})_{(k-1,1,0),(s,e,i)}. \end{aligned} \quad (2.17)$$

## 2.F Numerical aspects for the model reductions of Section 2.4.4

**SIR model with two levels of mixing and Gamma distributed individual recovery times.** The reduced model still takes the form of the dynamical system given in Equation (2.13), whose parameters are determined as follows.  $\gamma$  is set to the average recovery time and the growth rate  $r$  is obtained from Equation (2.11), where the coefficients of the relevant matrix defined in Equation (2.12) are estimated through simulations.

**SEIR model with two levels of mixing.** The reduced model is a standard uniformly mixing deterministic *SEIR* model, with infectious contact rate  $\hat{\lambda}$  and transition rate  $\mu$  for the transition from *E* to *I*. As before,  $\gamma$  designates the recovery rate of infected individuals.

Notice that, given the parameters  $\hat{\lambda}$ ,  $\mu$  and  $\gamma$ , the epidemic growth rate of the deterministic *SEIR* model can be computed by solving Equation (2.6) with  $\zeta(\tau) = (\hat{\lambda}/\gamma)\omega(\tau)$ , where  $\omega(\tau)$  is the distribution of infection times for an infected individual in a uniformly mixed population and  $\hat{\lambda}/\gamma$  is the average number of infections caused by an individual in early stages of the *SEIR* epidemic.

In order to choose the value of  $\hat{\lambda}$  for the model reduction, we proceed as follows. First, the epidemic growth rate  $r$  of the *SEIR* model with two layers of mixing is computed following section 2.E. It then remains to set  $\hat{\lambda}$  in such a way that the epidemic growth rate of the deterministic *SEIR* model, which can be computed as described above, is equal to  $r$ . This

can be achieved through simulations. Alternatively, it is possible to use the following formula derived from the Supplementary Material of Trapman et al., 2016 which states that

$$\hat{\lambda} = \gamma \left(1 + \frac{r}{\gamma}\right) \left(1 - \frac{r}{\mu}\right).$$

The reduced model is then defined by the following set of ordinary differential equations:

$$\begin{cases} \frac{dS}{dt} = -\hat{\lambda}SI \\ \frac{dE}{dt} = \hat{\lambda}SI - \mu E \\ \frac{dI}{dt} = \mu E - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases}$$



# Large population limit for a multilayer *SIR* model including households and workplaces

We study a multilayer *SIR* model with two levels of mixing, namely a global level which is uniformly mixing, and a local level with two layers distinguishing household and workplace contacts, respectively. We establish the large population convergence of the corresponding stochastic process. For this purpose, we use an individual-based model whose state space specifies the remaining infectious period length for each infected. This allows to deal with the natural correlation of the epidemic states of individuals whose household and workplace share a common infected. In a general setting where a non-exponential distribution of infectious periods may be considered, convergence to the unique deterministic solution of a measure-valued equation is obtained. In the particular case of exponentially distributed infectious periods, we show that it is possible to further reduce the obtained deterministic limit, leading to a closed, finite dimensional dynamical system capturing the epidemic dynamics. This model reduction subsequently is studied from a numerical point of view. We illustrate that the dynamical system derived from the large population approximation is a pertinent model reduction when compared to simulations of the stochastic process or to an alternative edge-based compartmental model, both in terms of accuracy and computational cost.

**Code availability.** <https://github.com/m-kubasch/household-workplace-model>

This chapter is based on the preprint "Large population limit for a multilayer *SIR* model including households and workplaces" (Kubasch, 2023).

## Contents

<b>3.1</b>	<b>Introduction</b>	<b>108</b>
<b>3.2</b>	<b>Presentation of the model</b>	<b>111</b>
3.2.1	General presentation of the model	111
3.2.2	The epidemic model at the level of households and workplaces	112
<b>3.3</b>	<b>Main results</b>	<b>117</b>
3.3.1	Large population approximation of $(\zeta^K)_{K \geq 1}$	117
3.3.2	Numerical assessment of the limiting dynamical system	120
<b>3.4</b>	<b>Proofs</b>	<b>123</b>
3.4.1	Proof of Theorem 3.3.2	123
3.4.2	Proof of Theorem 3.3.3	138
	<b>Discussion</b>	<b>147</b>



<b>Appendix</b> . . . . .	<b>148</b>
<b>3.A Absolute continuity of <math>\eta</math> and related PDE system</b> . . . . .	<b>148</b>
<b>3.B Implementation of the large population limit</b> . . . . .	<b>154</b>
3.B.1 Automatic implementation of the dynamical system . . . . .	154
3.B.2 Computational performance . . . . .	154
<b>3.C Edge-based compartmental model</b> . . . . .	<b>157</b>
3.C.1 Presentation of the EBCM . . . . .	157
3.C.2 Computational performance . . . . .	160
<b>3.D Proof of Proposition 3.4.10</b> . . . . .	<b>161</b>

---

### 3.1 Introduction

Epidemic spread depends by essence on the way individuals interact with one another. As a consequence, models have been developed which take into account main features of real-life contacts, for instance through contact networks (Kiss et al., 2017). In particular, some attention has been drawn to studying clustered networks, as clustering has a strong impact on epidemic spread (Hébert-Dufresne et al., 2010; Volz et al., 2011) which is intimately related to the way clustering is achieved within the network (House and Keeling, 2011). A particular form of clustering consists in the presence of entirely connected small social structures, such as households or workplaces, which exist in addition to random contacts in the general population. This kind of population structure is captured by models with two levels of mixing (Britton and Pardoux, 2019a, and references therein), which are related to efficient control measures. Indeed, both COVID-19 and influenza epidemics illustrate the pertinence of teleworking and school closures (Mendez-Brito et al., 2021; Simoy and Aparicio, 2021; Luca et al., 2018), and models explicitly distinguishing different contact types are well suited to simulate the impact of these measures (Di Lauro et al., 2021). However, precisely understanding the impact on disease propagation of the way individuals are organized in households and workplaces is not straightforward (Bansaye et al., 2023a).

This motivates the study of models with several levels of contact, which as we shall see lead to interesting mathematical issues due to their multiscale population structure. Pellis et al., 2009 have proposed a model with two levels of mixing structured in three layers of contacts: households, workplaces and the general population. This household-workplace model has already been studied to some extent, establishing for instance the epidemic growth rate (Pellis et al., 2011; Bansaye et al., 2023a) and several reproduction numbers (Ball et al., 2016). In particular, the  $R_I$  used throughout this paper was introduced in Pellis et al., 2009, and computations of  $R_I$  and proportions of infections per contact layer will make use of the working package associated to Bansaye et al., 2023a.

One drawback of this model is its complexity, both mathematically and numerically. Indeed, it is not simple to analyse due to correlations arising as soon as an individual may belong to several small contact structures at once. Also, simulations require a significant amount of computation time, especially when considering larger population sizes. As a consequence, it is of interest to develop reduced models, which may be more prone to theoretical studies and/or numerical exploration. In particular, large population approximations of stochastic

models have proven fruitful to achieve such model reductions in many contexts, among which epidemics on random graphs.

Historically, the standard *SIR* model developed by Kermack and McKendrick itself corresponds to the large population limit of the uniformly mixing stochastic *SIR* model. In the Markovian setting, the convergence of the stochastic model to its deterministic limit can be established using classical results on the convergence of finite type density-dependent Markov jump processes, e.g. Andersson and Britton, 2000a. When infectious periods are not restricted to being exponentially distributed, the large population convergence of the stochastic model to the unique deterministic solution of a system of integral equations can also be obtained (Kurtz, 1981; Forien et al., 2022). For more complex contact networks however, it often is challenging to propose closed systems of equations correctly describing the epidemic dynamics.

A well-understood case is the *SIR* model on the configuration graph, for which a reduced model referred to as edge-based compartmental model (EBCM) (Volz, 2008; Miller, 2011) has been proven to be the large population limit of the underlying stochastic model (Decreusefond et al., 2012; Janson et al., 2014). Since then, the equivalence with other reduced models has been established under appropriate assumptions (House and Keeling, 2011; Wilkinson et al., 2017; Jacobsen et al., 2018; Kiss et al., 2023), and the EBCM formalism has been extended to related models (Sherborne et al., 2018; Jacobsen et al., 2018). The configuration graph is a favorable setting for this analysis thanks to the absence of clustering in the large graph limit, which however also constitutes a major limitation, prohibiting for instance the existence of household-like structures.

Some attention has thus been drawn to models where each individual belongs to a random number of fully connected subgraphs (*cliques*) of the same type, hence being closely related to the household-workplace model. Several reduced models have been proposed, including an EBCM (Volz et al., 2011; St-Onge et al., 2023; Hébert-Dufresne et al., 2010). To our knowledge, the convergence of the underlying stochastic model to the proposed reduced model has not been established in any of these settings. Notice that these models share a major common point, which will also hold true in our setting: they focus on the epidemic at the level of structures, as they keep track of the proportions of cliques containing a certain number of susceptibles and infected, leading to high-dimensional dynamical systems for larger clique sizes.

When considering two levels of mixing, the first model for which a large population limit has been determined is the household model, which assumes a uniformly mixing general population and that each individual belongs to exactly one household (House and Keeling, 2008), thus being a special case of the household-workplace model. Here, the stochastic model can again be formalized as a finite type density-dependent Markov jump process, ensuring the large population convergence to the deterministic model. If either of these two assumptions is relaxed, e.g. considering a configuration graph at the global level (Di Lauro et al., 2021; Ma et al., 2013) or individuals belonging to several households (Barnard et al., 2018), reduced models have been proposed, but without rigorous derivation from stochastic models. In particular, the case of the household-workplace model is not covered, and the only reduced models proposed so far approach the epidemic dynamics using well calibrated uniformly mixing models (Bansaye et al., 2023a; del Valle Rafo et al., 2021). While these are capable of capturing some key characteristics of the epidemic, such as the epidemic peak size and final size, they do not allow for an accurate prediction of the epidemic dynamic over

time.

As a consequence, in this paper, we will study the large population limit of the multilayer *SIR* model with households and workplaces. In order to do so, we will formalize the model in a finite population as an individual-based stochastic process, and establish that this sequence of processes converges in law when the size of the population grows to infinity. This allows to identify a new model reduction, and establishes that it is asymptotically exact. Besides, it paves the way for more quantitative estimates on this approximation.

Notice here that each infected individual correlates the epidemic spread in his household and workplace, being infectious for exactly the same period of time in both structures. In order to deal with this dependence, the duration of infectious periods will explicitly be taken into account in the mathematical representation of the model. This difficulty actually arises as soon as one considers the probability of an individual belonging to several cliques at once, whether they are of different types or not, and we refer to Ball et al., 2014 where a similar approach has been developed for branching approximations. Let us emphasize that this model formulation allows to immediately consider a wide range of infectious period length distributions instead of being restricted to the Markovian case, which is a pertinent generalisation for many epidemic models (Sherborne et al., 2018; Forien et al., 2022; Feng et al., 2007; Lloyd, 2001).

The model will be represented by a measure-valued process mixing discrete and continuous components. More precisely, we establish the convergence of the individual-based process to the unique solution of an explicit measure-valued equation. In the particular case where this distribution is exponential, it is possible to go one step further and reduce the epidemic dynamics to a closed, finite dimensional dynamical system which is similar in spirit to reductions proposed in related settings (House and Keeling, 2008; St-Onge et al., 2023).

The present paper is structured as follows. Section 3.2 introduces the individual-based model, and Section 3.3.1 subsequently presents the convergence results in detail. Section 3.3.2 is devoted to numerical aspects. We first illustrate that the obtained dynamical system is in good accordance with stochastic simulations, discuss its implementation and examine its computational cost in terms of computation time compared to stochastic simulations. Next, we confront our reduced model to an alternative model reduction which we obtain using the EBCM formalism. Finally, Section 3.4 contains the proofs of our results.

Before proceeding, let us introduce some notations that will be used throughout the paper. For any integers  $n \leq m$ , we write  $\llbracket n, m \rrbracket = \{n, \dots, m\}$ . For a measurable space  $(E, \mathcal{E})$ , let  $\mathcal{M}_P(E)$  be the set of point measures,  $\mathcal{M}_F(E)$  the set of finite measures and  $\mathcal{M}_1(E)$  the set of probability measures on  $E$ . We define  $\mathcal{M}_{P,1}(E) = \mathcal{M}_P(E) \cap \mathcal{M}_1(E)$  the set of punctual probability measures on  $E$ . For a measure  $\mu$  on  $E$  and a suitable function  $f$  (either non-negative or belonging to  $L^1(\mu)$ ), let  $\langle \mu, f \rangle = \int_E f d\mu$ . Also, for  $x \in E$ ,  $\delta_x$  designates the Dirac measure at point  $x$ . Further, for any metric space  $E$  and any integer  $m$ , let  $\mathcal{C}(E, \mathbb{R}^m)$  be the set of continuous functions  $f : E \rightarrow \mathbb{R}^m$ . Similarly,  $\mathcal{C}_b(E, \mathbb{R}^m)$  is defined as the subset of bounded functions  $f \in \mathcal{C}(E, \mathbb{R}^m)$ . Finally, the space  $\mathcal{C}_b^1(E, \mathbb{R}^m)$  designates the set of bounded functions  $f : E \rightarrow \mathbb{R}^m$  such that  $f$  is differentiable and its differential is continuous and bounded.

## 3.2 Presentation of the model

Let us begin by introducing the epidemic model of interest, in two successive steps. At first, a general model description is yielded, which corresponds to a more intuitive presentation of the model, before stating the mathematical model in detail using a measure-valued stochastic differential equation.

### 3.2.1 General presentation of the model

Let us start by describing the population structure of interest. Consider a population of  $K$  individuals. Each individual is part of exactly one household and one workplace, which are chosen independently from one another, and independently for each individual.

More precisely, such a population structure can be obtained as described in Bansaye et al., 2023a. Suppose that households and workplaces are of size at least one and at most  $n_{\max}$ . Consider distributions  $(\pi_j^H)$  and  $(\pi_j^W)$  on  $\llbracket 1, n_{\max} \rrbracket$ . These distributions correspond to the large population limit of household and workplace size distributions, in the sense that in an infinite population, a proportion  $\pi_j^H$  of households would be of size  $j$ , while  $\pi^W$  would play a similar role for workplaces. In such an infinite population, the average household and workplace sizes, respectively  $m_H$  and  $m_W$ , would be given for  $X \in \{H, W\}$  by

$$m_X = \sum_{j=0}^{n_{\max}} j \pi_j^X.$$

On a probability space  $(\Omega_G, \mathbb{P}_G, \mathcal{F}_G)$ , we construct a sequence  $(\mathbf{G}^K)_{K \geq 1}$  of this random population structure as follows. For  $K \geq 1$ , let  $k \in \llbracket 0, K \rrbracket$  be the number of individuals who are not yet member of a household. While  $k > 0$ , choose a size  $\tilde{n}$  according to  $\pi^H$ , independently from the household sizes that were chosen during previous steps. The newly uncovered household is then of size  $n = \min(\tilde{n}, k)$ , and  $n$  individuals out of the  $k$  remaining ones are picked uniformly at random to assemble this new household. Consequently, it remains to update  $k$  to  $k - n$ . The process stops as soon as  $k = 0$ , as all individuals then belong to a household. Finally, this process is repeated independently for workplaces, using  $\pi^W$  instead of  $\pi^H$ .

It remains to describe the way the disease spreads in the population. The epidemic model considered here is an extension of the standard *SIR* model. At each time, each individual is either *susceptible* if he has never encountered the disease and may be contaminated; *infected* if the individual is currently infectious, in which case he may transmit the disease to other susceptibles; or *recovered*, once the infectious period is over, in which case the individual has become immune against the disease.

The disease is transmitted among individuals as follows. Within each household, each workplace and the general population, uniform mixing is assumed, but the parameterization differs slightly between the layers. Indeed, for households, we consider a *one-to-one* contact rate  $\lambda_H$ , meaning that whenever there are  $s$  susceptibles and  $i$  infected within a household, the next infectious contact occurs at rate  $\lambda_H s i$ . Similarly, another one-to-one contact rate  $\lambda_W$  is associated to workplace contacts. Within the general population, a *one-to-all* contact rate  $\beta_G$  is considered: when there are  $s$  susceptibles and  $i$  infected within a population of

size  $K$ , infectious contacts occur at rate  $\frac{\beta_G}{K} si$ . Indeed, each given encounter within the general population becomes less likely when the population size  $K$  increases.

Finally, infected individuals remain infectious for a period of time which is independent from  $(\mathbf{G}_K)_{K \geq 1}$  and distributed according to a probability distribution  $\nu$  on  $\mathbb{R}_+$ , which we assume to be absolutely continuous with regard to the Lebesgue measure. Once they recover, they are supposed to be immune against the disease from there on. In particular, if  $\nu$  is an exponential distribution, this corresponds to the Markovian *SIR* model.

### 3.2.2 The epidemic model at the level of households and workplaces

As we aim at investigating the large population limit of this model, we choose to enrich the population description as to obtain a closed Markov process. This corresponds to a favourable mathematical setting, as it allows us to use the associated martingale problem. In order to do so, we will represent the population in terms of particles which are described by a type. It seems natural to consider particles which correspond to entire structures, *i.e.* households and workplaces, which are characterized by their size and the number of susceptible and infected individuals they contain. Indeed, this point of view has already proven useful for deriving reproduction numbers for related models (Ball et al., 2016), as well as the epidemic growth rate of the household-workplace model (Pellis et al., 2011). However, this is not enough to obtain a closed system of Markovian dynamics. The problem is that each infected individual correlates the spread of the epidemic within his household and his workplace, leading to an intricate correlation network. In order to circumvent this difficulty, similarly to Ball et al., 2014, we will thus further characterize each structure by the infectious periods of its infected members. Adopting this point of view is key, as it allows to handle both the progressive discovery of the graph as the epidemic spreads, and the correlations arising from infected individuals, without explicitly keeping in memory the discovered graph.

Let  $\omega \in \Omega_G$ . For a population of size  $K \geq 1$ , let  $K_H$  be the number of households and  $K_W$  the number of workplaces in  $\mathbf{G}^K(\omega)$ . While  $K_H$  and  $K_W$  depend on  $\omega$ , this dependency is not specified explicitly for readability. This will also apply to the forthcoming notations. Label the  $K_H$  households in an arbitrary fashion  $1, \dots, K_H$ . Consider the set

$$E = \{(n, s, \tau) \in \llbracket 1, n_{\max} \rrbracket \times \llbracket 0, n_{\max} \rrbracket \times \mathbb{R}^{n_{\max}} : s \leq n; \forall j > n - s, \tau_j = 0\}.$$

Then for  $k \in \llbracket 1, K_H \rrbracket$ , the  $k$ -th household is characterized at time  $t \geq 0$  by its type

$$x_k^H(t) = (n_k^H, s_k^H(t), \tau_k^H(t)) \in E.$$

The first two components of  $x_k^H$  correspond respectively to the size of the household (which is constant over time), and the number of susceptible members of the household at time  $t$ . The third component  $\tau_k^H$  is a vector containing the *remaining infectious periods* of the members of the household. Indeed, at time  $t$ , there are  $n_k^H - s_k^H(t)$  infected or removed individuals within the household. For  $j \in \llbracket 1, n_k^H - s_k^H(t) \rrbracket$ , if  $\tau_{k,j}^H(t) > 0$ , the individual is still infectious and will remain so for  $\tau_{k,j}^H(t)$  units of time. Otherwise, if  $\tau_{k,j}^H(t) \leq 0$ , the individual has recovered, and the recovery has occurred at time  $t - |\tau_{k,j}^H(t)|$ . For  $j > n_k^H - s_k^H(t)$ ,  $\tau_{k,j}^H(t)$  has no interpretation, and is set to zero for convenience in computations. In other words, the infectiousness of a previously contaminated individual with remaining infectious period  $\tau$  is given by  $\mathbf{1}_{\{\tau > 0\}}$ .

Similarly, label the  $K_W$  workplaces in an arbitrary order  $1, \dots, K_W$ . For  $\ell \in \llbracket 1, K_W \rrbracket$ , the  $\ell$ -th workplace is characterized by its type  $x_\ell^W(t) = (n_\ell^W, s_\ell^W(t), \tau_\ell^W(t))$ , which is defined analogously to household types.

Notice that all of these quantities depend on the population size  $K$ , but this dependency is omitted to simplify notations.

By definition, these types evolve over time. On the one hand, for any  $X \in \{H, W\}$ , for any  $k \in \llbracket 1, K_X \rrbracket$  and  $j \in \llbracket 1, n_{\max} \rrbracket$ , the  $j$ -th component of  $\tau_k^X$  decreases linearly at unitary rate if it describes the remaining infectious period of an individual having contracted the disease at some previous time, and stays constant otherwise:

$$\forall j \in \llbracket 1, n_{\max} \rrbracket, \frac{d}{dt} \tau_{k,j}^X(t) = -\mathbf{1}_{\{j \leq n_k^X - s_k^X(t)\}}.$$

Let  $(e_j)_{1 \leq j \leq n_{\max}}$  denote the canonical basis of  $\mathbb{R}^{n_{\max}}$ . Then for any  $0 \leq t \leq T$ , and  $x = (n, s, \tau) \in E$ , we may define  $\Psi(x, T, t)$  as the type of a structure at time  $T$  given that it was in state  $x$  at time  $t$ , supposing that no infections occurred in the meantime:

$$\Psi(x, T, t) = \left( n, s, \tau - \sum_{j=1}^{n-s} (T-t)e_j \right).$$

On the other hand, infections within each level of mixing also cause the modification of the types of the household and the workplace of newly contaminated individuals. More precisely, consider a contamination occurring at time  $t$ . Suppose that the newly infected belongs to the  $k$ -th household and  $\ell$ -th workplace. Let  $\sigma$  be the realisation of a random variable of distribution  $\nu$ , which is drawn independently for each new infected. Then  $x_k^H$  and  $x_\ell^W$  jump from  $x_k^H(t-)$  and  $x_\ell^W(t-)$  to  $j(x_k^H(t-), \sigma)$  and  $j(x_\ell^W(t-), \sigma)$  respectively, where for any  $x = (n, s, \tau) \in E$ ,

$$j(x, \sigma) = (n, s-1, \tau + \sigma e_{n-s+1}).$$

It remains to describe how one identifies the household  $k$  and workplace  $\ell$  the newly infected belongs to. Let  $S(t-)$  be the number of susceptibles in the population previously to the infection event. If it takes place within the general population, any susceptible individual is chosen with uniform probability to be contaminated. The newly infected thus belongs to the  $k$ -th household with probability  $s_k^H(t-)/S(t-)$ , and independently to the  $\ell$ -th workplace with probability  $s_\ell^W(t-)/S(t-)$ . Similarly, if the infection occurs within a household, only the workplace of the newly infected needs to be uncovered, and corresponds to the  $\ell$ -th workplace with the same probability as previously. Within-workplace infections are treated analogously.

We are now ready to introduce the stochastic process  $(\zeta_t^K = (\zeta_t^{H|K}, \zeta_t^{W|K}))_{t \geq 0}$  taking values in  $\mathfrak{M}_{P,1} = \mathcal{M}_{P,1}(E) \times \mathcal{M}_{P,1}(E)$ .  $\zeta_t^{H|K}$  and  $\zeta_t^{W|K}$  correspond respectively to the normalized counting measures associated to the distributions of household and workplace types at time  $t$ , i.e. for any time  $t \geq 0$  and  $X \in \{H, W\}$ ,

$$\zeta_t^{X|K} = \frac{1}{K_X} \sum_{k=1}^{K_X} \delta_{x_k^X(t)}.$$

Start by noticing that, as both household and workplace sizes are bounded, for  $X \in \{H, W\}$ , the following inequality holds:

$$\frac{K}{n_{\max}} \leq K_X \leq K.$$

Thus, studying the asymptotic  $K \rightarrow \infty$  amounts to  $(K_H, K_W) \rightarrow (\infty, \infty)$ .

Observe that the number of infected individuals in a household of type  $(n, s, \tau)$  is given by  $i(\tau) = \sum_{k=1}^{n_{\max}} \mathbf{1}_{\{\tau_k > 0\}}$ . Then for any  $t \geq 0$ ,

$$I_H(t) = \frac{1}{K_H} \sum_{k=1}^{K_H} i(\tau_k^H(t))$$

corresponds to the average number of infected individuals per household at time  $t$ . Similarly, one may define  $S_H(t)$  as the average number of susceptibles per household at time  $t$ , as well as the workplace-related quantities  $I_W(t)$  and  $S_W(t)$ . Then

$$\forall X \in \{H, W\}, S(t) = K_X S_X(t) \text{ and } I(t) = K_X I_X(t). \quad (3.1)$$

Further, let  $N_H$  be the average household size, which is constant over time and always equal to  $K/K_H$ . This leads to  $I(t)/K = I_X(t)/N_X$ , which will be of use in computations. Notice that we will need to check that Equation (3.1) is well posed, as equalities of the type  $K_H S_H(t) = K_W S_W(t)$  technically need to be proven for the stochastic process formalizing the model. Notice that for  $X \in \{H, W\}$ ,  $S_X$  and  $I_X$  actually depend on the population size  $K$ , which is omitted in notations for readability.

Finally, let us briefly emphasize that the partition of the population in households and workplaces is entirely conveyed by  $\zeta_0^K \in \mathfrak{M}_{P,1}$ , as it does not vary over time. In particular, the proportions of households and workplaces of each size are supposed to correspond to those observed in  $\mathbf{G}^K(\omega)$ . Similarly, there are some natural constraints on  $\zeta_0^K$ , as  $\zeta^{H|K}$  and  $\zeta^{W|K}$  describe the same population, once dispatched into households, and once dispatched into workplaces. For instance, the total number of members in some epidemic state (susceptible, infected or recovered) within all households is equal to the total number of members in this state within all workplaces. Hence  $K_H S_H(0) = K_W S_W(0)$  and  $K_H I_H(0) = K_W I_W(0)$  almost surely. Further, at time 0, each infected or recovered needs to have the same remaining infectious period in both his household and workplace. In other words, almost surely,

$$\{\tau_{k,j}^H(0) : 1 \leq k \leq K_H, 1 \leq j \leq n_k^H - s_k^H(0)\} = \{\tau_{\ell,j}^W(0) : 1 \leq \ell \leq K_W, 1 \leq j \leq n_\ell^W - s_\ell^W(0)\}. \quad (3.2)$$

Forthcoming Lemma 3.2.2 shows that these conditions are enough to ensure that the previous characterization of  $S(t)$  and  $I(t)$  in terms of  $S_X(t)$  and  $I_X(t)$  is legitimate.

Before giving the proper definition of  $\zeta^K$ , let us introduce some necessary notations. Again,  $K$ -dependency is not specified explicitly in order to simplify notations. Let

$$U_G = (\mathbb{R}_+)^3 \times \llbracket 1, K_H \rrbracket \times \llbracket 1, K_W \rrbracket \times \mathbb{R}_+, \quad (3.3)$$

and consider the following measure on  $U_G$ :

$$\mu_G(du) = \mu_G(d\theta, dk, dl, d\sigma) = d\theta \otimes \mu_{\#}(dk) \otimes \mu_{\#}(dl) \otimes \nu(d\sigma),$$



where  $d\boldsymbol{\theta}$  and  $\mu_{\#}$  denote the Lebesgue measure on  $\mathbb{R}^3$  and the standard counting measure, respectively. For  $t \geq 0$  and  $u = (\boldsymbol{\theta}, k, \ell, \sigma) \in U_G$  where  $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ , let us define

$$\mathcal{I}_G(t, u) = \mathbf{1}_{\left\{ \theta_1 \leq \frac{\beta_G}{K} S(t) I(t), \theta_2 \leq \frac{s_k^H(t)}{S(t)}, \theta_3 \leq \frac{s_\ell^W(t)}{S(t)} \right\}}.$$

The idea is that  $\mathcal{I}_G$  will yield the correct rate for infection events in the general population. More precisely, the constraint on  $\theta_1$  corresponds to the rate of infectious contacts at that level of mixing, while the constraints on  $\theta_2$  and  $\theta_3$  are related to the probability that the newly infected belongs to the  $k$ -th household and  $\ell$ -th workplace.

Similarly, let

$$U = (\mathbb{R}_+)^2 \times \llbracket 1, K_H \rrbracket \times \llbracket 1, K_W \rrbracket \times \mathbb{R}_+, \quad (3.4)$$

endowed with the measure

$$\mu(du) = \mu(d\boldsymbol{\theta}, dk, d\ell, d\sigma) = d\boldsymbol{\theta} \otimes \mu_{\#}(dk) \otimes \mu_{\#}(d\ell) \otimes \nu(d\sigma),$$

where with slight abuse of notation,  $d\boldsymbol{\theta}$  designates the Lebesgue measure on  $\mathbb{R}^2$ . Then for  $t \geq 0$  and  $u = (\boldsymbol{\theta}, k, \ell, \sigma) \in U$  where  $\boldsymbol{\theta} = (\theta_1, \theta_2)$ , we further introduce

$$\mathcal{I}_H(t, u) = \mathbf{1}_{\left\{ \theta_1 \leq \lambda_H s_k^H(t) i(\tau_k^H(t)), \theta_2 \leq \frac{s_\ell^W(t)}{S(t)} \right\}}.$$

This time, the constraints on  $\theta_1$  and  $\theta_2$  correspond respectively to the rate of infection within the  $k$ -th household and the probability of the newly infected belonging to the  $\ell$ -th workplace. Also, for any  $T \geq t \geq 0$  and  $u = (\boldsymbol{\theta}, k, \ell, \sigma) \in U_G \cup U$ , consider the following quantity, which will allow to keep track of the change in the household population due to an infection within the  $k$ -th household:

$$\Delta_H(u, T, t) = \delta_{(\Psi(j(x_k^H(t-), \sigma), T, t))} - \delta_{(\Psi(x_k^H(t-), T, t))}.$$

Finally, define  $\mathcal{I}_W(t, u)$  and  $\Delta_W(u, T, t)$  analogously:

$$\begin{aligned} \mathcal{I}_W(t, u) &= \mathbf{1}_{\left\{ \theta_1 \leq \lambda_W s_\ell^W(t) i(\tau_\ell^W(t)), \theta_2 \leq \frac{s_k^H(t)}{S(t)} \right\}}, \\ \text{and } \Delta_W(u, T, t) &= \delta_{(\Psi(j(x_\ell^W(t-), \sigma), T, t))} - \delta_{(\Psi(x_\ell^W(t-), T, t))}. \end{aligned}$$

We are now ready to yield the main characterization of  $\zeta_t^K$ , as inspired by Tran, 2006.

**Proposition 3.2.1.** *Define on the same probability space as  $\zeta_0^K$ , and independently from  $\zeta_0^K$ , three independent Poisson point measures  $Q_Y$  on  $\mathbb{R}_+ \times U_Y$  with intensity  $dt \mu_Y(du)$ , for  $Y \in \{H, W, G\}$ . Then  $\zeta^K = (\zeta^{H|K}, \zeta^{W|K})$  is defined as the unique strong solution taking values in  $\mathbb{D}(\mathbb{R}_+, \mathfrak{M}_{P,1})$  of the following equation. For  $X \in \{H, W\}$  and  $T \geq 0$ ,*

$$\zeta_T^{X|K} = \frac{1}{K_X} \left( \sum_{j=1}^{K_X} \delta_{\Psi(x_j^X(0), T, 0)} + \sum_{Y \in \{H, W, G\}} \int_0^T \int_{U_Y} \mathcal{I}_Y(t-, u) \Delta_X(u, T, t) Q_Y(dt, du) \right), \quad (3.5)$$

where  $U_G$  and  $U_H = U_W = U$  are defined by Equations (3.3) and (3.4), respectively.



The idea behind Equation (3.5) goes as follows. Let us focus for example on the distribution  $\zeta_T^{H|K}$  of household types at time  $T$ . Each household's type contributes to the distribution at uniform weight  $1/K_H$ . If no infection event occurs between times 0 and  $T$ , then the state of the  $k$ -th household at time  $T$  is given by  $\Psi(x_k^H(0), T, 0)$ . However, suppose now that before time  $T$ , at least one initially susceptible member of the  $k$ -th household is infected, and let  $t$  be the first time at which such an event occurs. Then  $x_k^H(t) = j(x_k^H(t-), \sigma)$  where  $\sigma$  is distributed according to  $\nu$ . If no other infections affect this household up to time  $T$ , it will be in state  $\Psi(j(x_k^H(t-), \sigma), T, t)$  instead of  $\Psi(x_k^H(0), T, 0) = \Psi(x_k^H(t-), T, t)$ . This reasoning is reflected in  $\Delta_H$ , and can be iterated over the whole of  $[0, T]$ . Finally, the terms  $\mathcal{I}_Y$  for  $Y \in \{G, H, W\}$  assure that all infection events occur at the corresponding rates.

*Proof.* The proof uses classical arguments, which will only be outlined here. Start by establishing existence of  $\zeta^K$ . Consider the sequence  $(T_n)_{n \geq 0}$  of successive jump times of  $\zeta^K$ , where we define  $T_0 = 0$ . Then using a method similar to rejection sampling,  $(T_n)_{n \geq 0}$  can be obtained as a subsequence of the jump times of a Poisson process with intensity  $n_{\max}(\lambda_H n_{\max} + \lambda_W n_{\max} + \beta_G)K$ , whose only limiting value is  $+\infty$ . Thus  $\lim_{n \rightarrow +\infty} T_n = +\infty$  almost surely, ensuring that  $\zeta^K$  takes values in  $\mathbb{D}(\mathbb{R}_+, \mathfrak{M}_{P,1}(E))$ .

Finally, uniqueness is obtained by an induction argument which proves that for any  $n \geq 0$ ,  $(T_n, \zeta_{T_n}^K)$  is uniquely determined by  $(\zeta_0^K, (Q_Y)_{Y \in \mathcal{S}})$  where  $\mathcal{S} = \{G, H, W\}$ . This obviously is true for  $n = 0$ . The induction step relies on the observation that  $T_{n+1}$  is uniquely determined by  $(T_n, \zeta_{T_n}^K, (Q_Y)_{Y \in \mathcal{S}})$  and  $\zeta_{T_{n+1}}^K$  by  $(T_{n+1}, T_n, \zeta_{T_n}^K, (Q_Y)_{Y \in \mathcal{S}})$ . The induction hypothesis allows to conclude.  $\square$

Let us briefly show that it follows from Proposition 3.2.1 that Equation (3.1) is well posed.

**Lemma 3.2.2.** *Suppose that almost surely,  $K_H S_H(0) = K_W S_W(0)$  and Equation (3.2) holds. Then for any  $t \geq 0$ ,  $K_H S_H(t) = K_W S_W(t)$  and  $K_H I_H(t) = K_W I_W(t)$ , almost surely.*

*Proof.* Let  $T \geq 0$ , and for any  $x = (n, s, \tau) \in E$ , let  $\mathbf{n}(x) = n$ ,  $\mathbf{s}(x) = s$  and  $\mathbf{i}(x) = i(\tau)$ . Start by focusing on  $S_X(T) = \langle \zeta_T^{X|K}, \mathbf{s} \rangle$  for  $X \in \{H, W\}$ . It follows from Equation (3.5) that

$$K_X S_X(T) = \sum_{j=1}^{K_X} \mathbf{s}(\Psi(x_j^X(0), T, 0)) + \sum_{Y \in \{H, W, G\}} \int_0^T \int_{U_Y} \mathcal{I}_Y(t-, u) \langle \Delta_X(u, T, t), \mathbf{s} \rangle Q_Y(dt, du).$$

Notice that on the one hand, for any  $x \in E$  and  $0 \leq t \leq T$ ,  $\mathbf{s}(\Psi(x, T, t)) = \mathbf{s}(x)$ . Hence the first term of the right-hand side equals  $K_X S_X(0)$ , and for any  $u = (\theta, k, \ell, \sigma)$ ,

$$\langle \Delta_H(u, T, t), \mathbf{s} \rangle = \mathbf{s}(\Psi(j(x_k^H(t-), \sigma), T, t)) - \mathbf{s}(\Psi(x_k^H(t-), T, t)) = -1.$$

The analogous computation yields  $\langle \Delta_W(u, T, t), \mathbf{s} \rangle = -1$ . Thus  $K_H S_H(T) = K_W S_W(T)$  almost surely.

Let us now turn to  $I_X(T) = \langle \zeta_T^{X|K}, \mathbf{i} \rangle$ . This time, for any  $u = (\theta, k, \ell, \sigma)$  and  $0 \leq t \leq T$ , it holds that  $\langle \Delta_X(u, T, t), \mathbf{i} \rangle = \mathbf{1}_{\{\sigma > (T-t)\}}$ . Finally, Equation (3.2) ensures that

$$\sum_{k=1}^{K_H} \mathbf{i}(\Psi(x_k^H(0), T, 0)) = \sum_{k=1}^{K_H} \sum_{j=1}^{n_k^H - s_k^K(0)} \mathbf{1}_{\{\tau_{k,j}^H(0) > T\}} = \sum_{\ell=1}^{K_W} \mathbf{i}(\Psi(x_\ell^W(0), T, 0)).$$

The conclusion follows as previously from Equation (3.5).  $\square$

### 3.3 Main results

In this section, we are going to present our main results on the convergence of  $(\zeta^K)_{K \geq 1}$  in the Skorokhod space  $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_1(E))^2$ . For  $X \in \{H, W\}$ , let  $\bar{X}$  be the complementary structure type, i.e.  $\bar{X} = W$  if  $X = H$  and vice-versa.

As we are interested in studying the limit  $K \rightarrow \infty$ , an important ingredient will be the asymptotic behavior of the sequence of random graphs on which the epidemic spreads. Let  $\pi^{H|K}$  and  $\pi^{W|K}$  be the household and workplace size distributions observed in  $\mathbf{G}^K$ . The law of large numbers ensures that  $(\pi^{H|K}, \pi^{W|K})_{K \geq 1}$  converges  $\mathbb{P}_{\mathcal{G}}$ -almost everywhere to  $(\pi^H, \pi^W)$ . We hence define

$$\Omega_{\mathcal{G}}^* = \{\omega \in \Omega_{\mathcal{G}} : (\pi^{H|K}(\omega), \pi^{W|K}(\omega)) \xrightarrow{K \rightarrow \infty} (\pi^H, \pi^W)\},$$

and our main results will hold for  $\omega \in \Omega_{\mathcal{G}}^*$ .

Further, the following assumption on the sequence of initial conditions  $(\zeta_0^K)_{K \geq 1}$  will be required from now on.

**Assumption 3.3.1.** For any  $X \in \{H, W\}$  and  $T \geq 0$ , suppose that:

(i)

$$\lim_{N \rightarrow \infty} \sup_{K \geq 1} \mathbb{E} \left[ \sup_{0 \leq t \leq T} \frac{1}{K_X} \sum_{k=1}^{K_X} \sum_{i=1}^{n_{\max}} \mathbf{1}_{\{n_k^X - s_k^X(0) \geq i, |\tau_{k,i}^X(0) - t| \geq N\}} \right] = 0.$$

(ii) For any  $c \in \mathbb{R}$ , for any  $i \in \llbracket 1, n_{\max} \rrbracket$ ,

$$\lim_{\epsilon \rightarrow 0} \sup_{K \geq 1} \mathbb{E} \left[ \frac{1}{K_X} \sum_{k=1}^{K_X} \mathbf{1}_{\{n_k^X - s_k^X(0) \geq i, |(\tau_{k,i}^X(0) - T) - c| \leq \epsilon\}} \right] = 0.$$

Briefly, the first assumption allows to control the impact of the initial condition on the queues of the distribution of remaining infectious periods at each time, while the second condition is related to aspects of absolute continuity. These conditions are for instance satisfied if for any  $K \geq 1$ , at time 0, the remaining infectious periods of infected individuals are i.i.d. of law  $\nu$ , while those of recovered individuals are set to be equal to zero. Notice that this choice for recovered individuals does not represent a loss of generality, as it does not affect the epidemic spread and initially recovered individuals will remain recognizable at any time  $T$  as the only ones whose remaining infectious periods equal  $-T$ .

#### 3.3.1 Large population approximation of $(\zeta^K)_{K \geq 1}$

For any  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$ , let  $f_t(x) = f(t, x)$  and  $f_t^{\mathcal{I}}(x) = \langle \nu, f_t(j(x, \cdot)) \rangle$  for every  $(t, x) \in \mathbb{R}_+ \times E$ . Consider the differential operator  $\mathcal{A}$  defined as follows. For any  $x = (n, s, \tau) \in E$ ,

$$\mathcal{A}f_t(x) = \partial_t f(t, x) - \sum_{k=1}^{n-s} \partial_{\tau_k} f(t, x).$$

Also, for any  $x = (n, s, \tau) \in E$ , let  $\mathbf{n}(x) = n$ ,  $\mathbf{s}(x) = s$  and  $\mathbf{i}(x) = i(\tau)$  be the functions which to a structure in state  $x$  associate the corresponding structure size, number of susceptible and number of infected members, respectively. For instance, for any  $X \in \{H, W\}$

the average rate of within-structure infections at time  $t$  is given by

$$\lambda_X \langle \zeta_t^{X|K}, \mathbf{si} \rangle = \frac{\lambda_X}{K_X} \sum_{k=1}^{K_X} \mathbf{s}(x_k^X(t)) \mathbf{i}(x_k^X(t)).$$

Notice also that as mentioned previously, the average size of structures is constant over time, hence  $\langle \zeta_t^{X|K}, \mathbf{n} \rangle = \langle \zeta_0^{X|K}, \mathbf{n} \rangle$  for all  $t \geq 0$ . Finally, let  $\mathfrak{M}_1 = \mathcal{M}_1(E)^2$ . We are now ready to state our first result, whose proof is postponed to Section 3.4.1.

**Theorem 3.3.2.** *Let  $\omega \in \Omega_G^*$ . Suppose that  $(\zeta_0^K)_{K \geq 1}$  satisfies Assumption 3.3.1 and converges in law to  $\eta_0 \in \mathfrak{M}_1$ . Then  $(\zeta^K)_{K \geq 1}$  converges in  $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_1(E))^2$  to  $\eta = (\eta^H, \eta^W)$  defined as the unique solution of the following system of Equations (3.6). For any  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$ , for any  $T \geq 0$  and  $X \in \{H, W\}$ ,*

$$\begin{aligned} \langle \eta_T^X, f_T \rangle &= \langle \eta_0^X, f_0 \rangle + \int_0^T \langle \eta_t^X, \mathcal{A}f_t \rangle dt + \lambda_X \int_0^T \langle \eta_t^X, \mathbf{si}(f_t^I - f_t) \rangle dt \\ &+ \lambda_{\bar{X}} \int_0^T \frac{\langle \eta_t^{\bar{X}}, \mathbf{si} \rangle}{\langle \eta_t^{\bar{X}}, \mathbf{s} \rangle} \langle \eta_t^X, \mathbf{s}(f_t^I - f_t) \rangle dt + \beta_G \int_0^T \frac{\langle \eta_t^H, \mathbf{i} \rangle}{\langle \eta_0^H, \mathbf{n} \rangle} \langle \eta_t^X, \mathbf{s}(f_t^I - f_t) \rangle dt. \end{aligned} \quad (3.6)$$

This measure-valued equation can further be related to a system of PDEs. Indeed, it is possible to establish an absolute continuity result for the marginals of  $\eta^X$  conditioned on the structure's size and number of susceptible members. The associated densities can be shown to satisfy, in the sense of distributions, a system of differential equations related to non-linear nonlocal transport equations. We refer to Appendix 3.A for detail.

From now on, let us assume that  $\nu$  is the exponential distribution of parameter  $\gamma$ . As we shall see, it then is possible to deduce from Theorem 3.3.2 that the proportion of susceptible and infected individuals in the population converges to the solution of a dynamical system, when the size of the population grows large.

Let  $s(t)$  and  $i(t)$  be the proportions of susceptible and infectious individuals, respectively, in the population at time  $t$  according to distribution  $\eta_t$ . Further introduce the set

$$\mathbb{S} = \{(n - i, i) : 2 \leq n \leq n_{\max}, 0 \leq i \leq n - 1\}.$$

For  $(S, I) \in \mathbb{S}$ , let  $n_{S,I}^H(t)$  be the proportion of households containing  $S$  susceptible and  $I$  infected individuals at time  $t$ , according to distribution  $\eta_t^H$ . Define  $n_{S,I}^W(t)$  analogously for workplaces. Finally, consider

$$\tau_G(t) = \beta_G i(t), \text{ and } \tau_X(t) = \frac{\lambda_X}{m_X} \sum_{(S,I) \in \mathbb{S}} SI n_{S,I}^X(t) \text{ for } X \in \{H, W\}.$$

We assume that at time 0, a fraction  $\varepsilon$  of uniformly chosen individuals are infected amidst an otherwise susceptible population. Furthermore, at time 0, the remaining infectious period of each infected individual is supposed to be distributed according to  $\nu$ , independently from one another. Let us emphasize here that actually, only this second assumption is crucial for the results to hold, while the original distribution of infected individuals does not need to be uniform (in which case forthcoming Equations (3.7) and (3.9) need to be adapted). We have

chosen this particular initial condition as it has been previously considered in the literature, and refer to the Discussion for further comments.

In practice, this setting corresponds to the following probability distribution  $\eta_{0,\varepsilon} = (\eta_{0,\varepsilon}^H, \eta_{0,\varepsilon}^W) \in \mathfrak{M}_1$  characterized for  $X \in \{H, W\}$  as follows. For any  $n \in \llbracket 1, n_{\max} \rrbracket$  and  $s \in \llbracket 0, n \rrbracket$ :

$$\eta_{0,\varepsilon}^X(n, s, d\tau) = \pi_n^X \binom{n}{s} (1 - \varepsilon)^s \varepsilon^{n-s} \left( \nu^{\otimes(n-s)} \otimes \delta_0^{\otimes(n_{\max}-n+s)} \right) (d\tau). \quad (3.7)$$

It then is possible to describe the epidemic dynamics by a finite, closed set of ordinary differential equations, as shown in the following result whose proof is postponed to Section 3.4.2.

**Theorem 3.3.3.** *Let  $\varepsilon > 0$  and suppose that  $\eta$  satisfies Equation (3.6) with  $\eta_0 = \eta_{0,\varepsilon}$ . Then the functions  $(s, i, n_{S,I}^X : X \in \{H, W\}, (S, I) \in \mathbb{S})$  are characterized as being the unique solution of the following dynamical system: for any  $t \geq 0$ ,  $X \in \{H, W\}$  and  $(S, I) \in \mathbb{S}$ ,*

$$\frac{d}{dt}s(t) = -(\tau_H(t) + \tau_W(t) + \tau_G(t)s(t)), \quad (3.8a)$$

$$\frac{d}{dt}i(t) = -\frac{d}{dt}s(t) - \gamma i(t), \quad (3.8b)$$

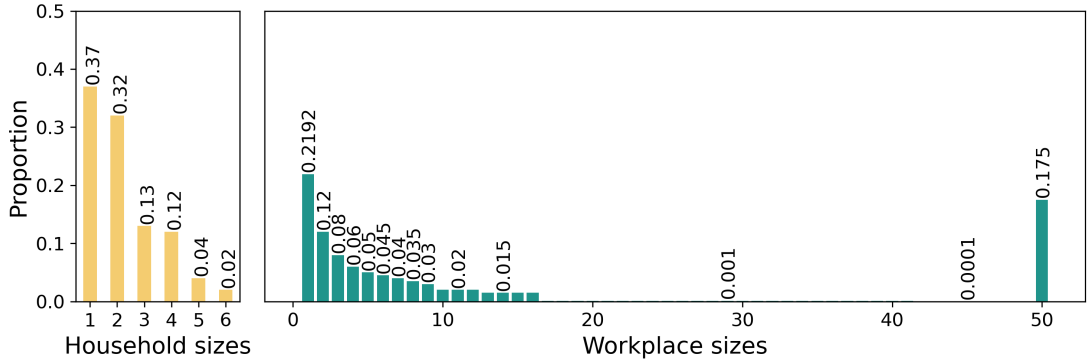
$$\begin{aligned} \frac{d}{dt}n_{S,I}^X(t) = & - \left( \lambda_X S I + \tau_{\bar{X}}(t) \frac{S}{s(t)} + \tau_G(t) S + \gamma I \right) n_{S,I}^X(t) \\ & + \gamma (I + 1) n_{S,I+1}^X(t) \mathbf{1}_{\{S+I < n_{\max}\}} \\ & + \left( \lambda_X (S + 1) (I - 1) + \tau_{\bar{X}}(t) \frac{S + 1}{s(t)} + \tau_G(t) (S + 1) \right) n_{S+1,I-1}^X(t) \mathbf{1}_{\{I \geq 1\}}, \end{aligned} \quad (3.8c)$$

with initial conditions given by

$$s(0) = 1 - \varepsilon; \quad i(0) = \varepsilon; \quad n_{S,I}^X(0) = \binom{S+I}{I} \pi_{S+I}^X (1 - \varepsilon)^S \varepsilon^I. \quad (3.9)$$

This dynamical system may be understood as follows. Equation (3.8a) corresponds to the fact that the proportion of susceptibles decreases whenever a new infection occurs within the general population, or within a household or workplace. Similarly, Equation (3.8b) is due to newly contaminated individuals moving from the susceptible to the infected state, which they in turn leave at rate  $\gamma$ . It remains to take an interest in Equation (3.8c). The first line indicates that a structure of type  $(S, I) \in \mathbb{S}$  changes its composition upon either the infection of one of its susceptible members which may occur in any layer of the graph, or upon the removal of one of its infected members. Simultaneously, a structure of type  $(S, I + 1)$  transforms into a structure of type  $(S, I)$  whenever one of its infected members recovers, while a structure of type  $(S + 1, I - 1)$  becomes of type  $(S, I)$  upon infection of a susceptible member. In particular, this result shows that under the assumptions of Theorem 3.3.3, in the large population limit, we may neglect the natural correlation between structures caused by the fact that infected individuals belong to two structures at once. This allows to obtain a stronger model reduction than in Theorem 3.3.2, in the sense that the model reduces to a finite-dimensional ODE-system instead of a measure-valued equation.

Before detailing the proofs of Theorems 3.3.2 and 3.3.3, let us examine the latter from a numerical point of view.



**Figure 3.1:** Household and workplace size distributions  $\pi^H$  (left) and  $\pi^W$  (right) used in simulations.

### 3.3.2 Numerical assessment of the limiting dynamical system

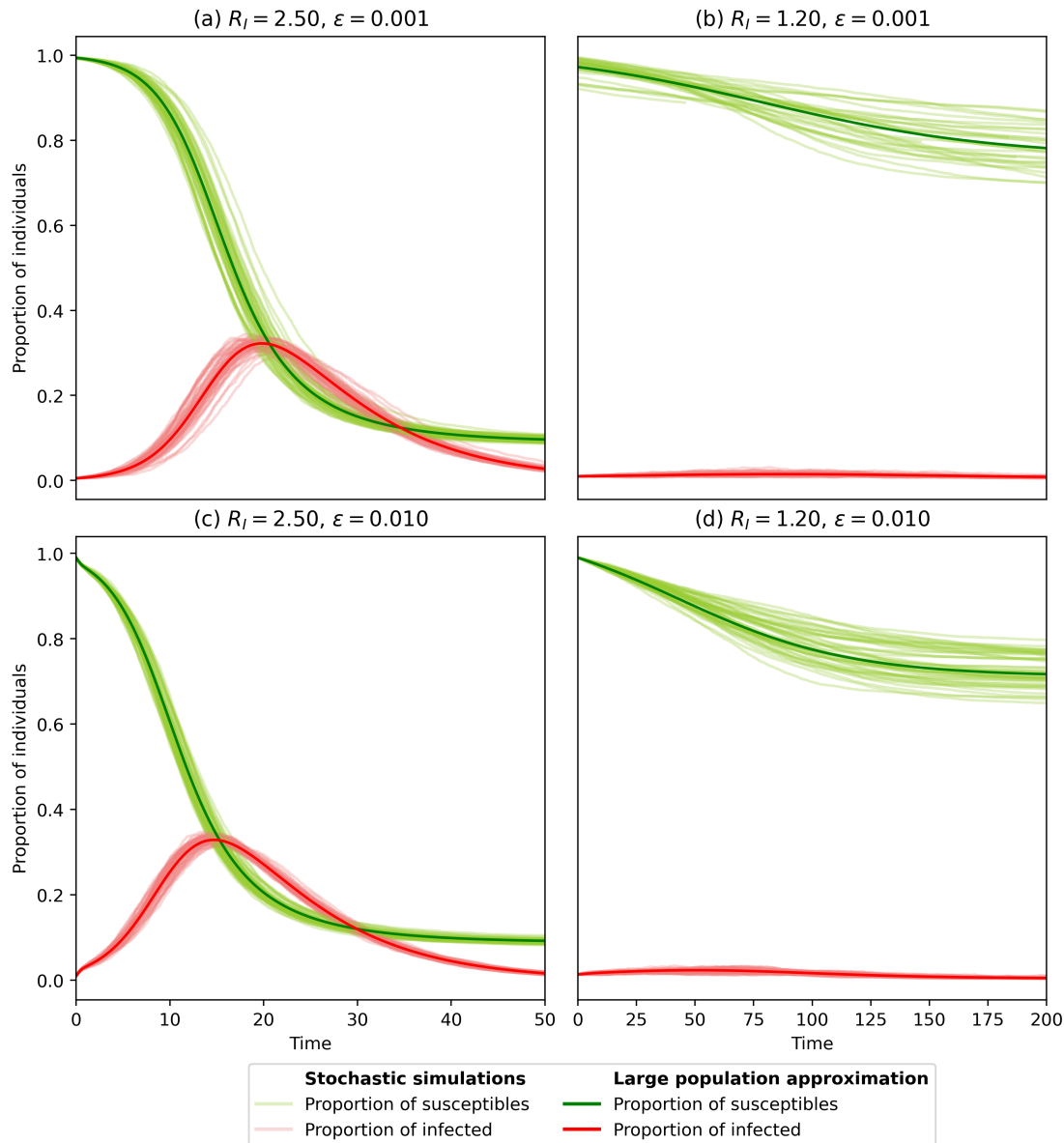
The aim of this section is first to portray that the proposed large population limit, under the form of dynamical system (3.8a–c), is in good accordance with the original stochastic model for large population sizes. This secondarily leads to some practical comments on the implementation of the dynamical system. Finally, a comparison with another reduced model for epidemics with two layers of mixing will be established, namely with an edge-based compartmental model (EBCM) in the line of work of Volz et al., 2011.

#### Implementation of the dynamical system and illustration of Theorem 3.3.3

Let us start by illustrating the result of Theorem 3.3.3 through numerical simulations. Using Gillespie’s algorithm, we have performed fifty simulations of the epidemic within a population of  $K = 10000$  individuals, where  $\pi^H$  and  $\pi^W$  are roughly inspired by the French household and workplace distributions as observed in 2018 by Insee (Bansaye et al., 2023a). These distributions are represented in Figure 3.1. Two sets of epidemic parameters have been considered, leading to either  $R_I = 2.5$  or  $R_I = 1.2$ . In both cases, the majority of contaminations take place at the local level. Indeed, for the first scenario with  $R_I = 2.5$ , 42% and 18% of infections occur within households and workplaces, respectively, and these proportions both equal 40% in the second scenario. Further, the epidemic is started by infecting either 10 or 100 individuals chosen uniformly at random at time 0. For each simulation, we have followed the evolution of the proportion of susceptible and infected individuals within the population over time.

The simulation outcomes are presented in Figure 3.2. For each choice of parameters, the solutions  $s$  and  $i$  of dynamical system (3.8a–c) with initial condition given by (3.9) are plotted on the same graph. As expected, one observes good accordance of the stochastic simulations and the deterministic functions  $(s, i)$ .

Before proceeding further, let us briefly emphasize a few aspects of the implementation of the proposed deterministic model. A potential drawback of dynamical system (3.8a–c) consists in its large dimension. Indeed, it holds that  $\#\mathbb{S} = n_{\max}(n_{\max} + 1)/2 - 1$ . The number of equations of dynamical system (3.8a–c) is hence of order  $O(n_{\max}^2)$ . However, this fast-growing number of equations actually is manageable, as it is possible to implement the



**Figure 3.2:** Comparison of the stochastic model with its large population approximation given by dynamical system (3.8a–c). Household and workplace distributions are those of Figure 3.1. Two sets of epidemic parameters are considered, namely  $(\beta_G, \lambda_H, \lambda_W, \gamma) = (0.125, 1.5, 0.00115, 0.125)$  and  $(\beta_G, \lambda_H, \lambda_W, \gamma) = (0.03, 0.05, 0.0015, 0.125)$  for the left and right column respectively ( $R_I = 2.5$  and  $R_I = 1.2$ ). The initial conditions are either  $\varepsilon = 0.001$  in Panels (a) and (b), or  $\varepsilon = 0.01$  in (c) and (d). For each of these scenarios, Gillespie’s algorithm is used to simulate 50 trajectories of the stochastic model defined in Proposition 3.2.1 in a population of  $K = 10000$  individuals (faint lines). For Panels (a) and (b), only trajectories reaching a threshold proportion of 0.005 infected are kept, and time is shifted so that time 0 corresponds to the moment when this threshold is reached. Finally, the deterministic solution  $(s, i)$  of (3.8a–c) is represented for each scenario (thick lines). For Panels (a) and (b), the same time shifting procedure as for simulations is applied.

dynamical system in an automated way, in the sense that each equation does not need to be written one-by-one by the programmer. We refer to Appendix 3.B.1 for details.

Nevertheless, the large dimension of the dynamical system of interest raises the question whether it is numerically speaking interesting to actually use it for numerical explorations. We have compared the average time needed to either solve once dynamical system (3.8a–c), or to simulate one trajectory of the stochastic model using Gillespie’s algorithm, for different choices of epidemic parameters. In practice, for stochastic simulations, it is often necessary to compute several individual trajectories in order to obtain the general behavior of the epidemic. However, as it is possible to execute these simulations in parallel, comparison to one individual simulation seemed the most pertinent. The procedure and results are detailed in Appendix 3.B.2. In summary, solving the reduced model is up to one order of magnitude faster than performing one stochastic simulation for values of  $R_I > 1$  that are not too close to the critical case  $R_I = 1$ . This shows that the reduced model is pertinent for numerical exploration.

### Comparison to edge-based compartmental models

One may notice that the population structure, as described in Section 3.2.1, can be regarded as a modification of the well-studied configuration model. Indeed, our network of household- and workplace-contacts may be seen as a two-layer graph, where each layer corresponds to a random graph generated as described in Miller, 2009 and Newman, 2009, that we shall call clique configuration model (CCM) hereafter. This random graph model generalizes configuration models to include small, totally connected sub-graphs referred to as *cliques*. It then is possible to derive an EBCM for our household-workplace model by reasoning as in Volz et al., 2011. Details are provided in Appendix 3.C.

Edge-based compartmental models on CCM variants have been known to be in good accordance with simulations of the corresponding stochastic epidemic models, under the assumption of a very small initial proportion of infected. In our case, we have confronted the EBCM with dynamical system (3.8a–c), as well as simulated trajectories of our stochastic model. As expected, for very small values of  $\varepsilon$ , the EBCM and dynamical system (3.8a–c) both yield the correct epidemic dynamics, whereas for larger values of  $\varepsilon$ , the EBCM does not fit the simulated epidemic trajectories. We refer to Appendix 3.C for details.

Finally, proceeding like before, we obtain that the number of equations of the EBCM is of order  $O(n_{\max}^3)$ . This has a strong negative impact on computation time, as briefly illustrated in Appendix 3.C, arguing against the applicability of this EBCM for numerical explorations.

To conclude, in the particular case of the household-workplace model studied in this article, the EBCM seems to be equivalent to the large population approximation described by dynamical system (3.8a–c), under the condition that the initial proportion of infected is very small. However, considering both the computational cost of its higher dimension and the loss of accuracy for more general initial conditions of the EBCM, the large population approximation given by dynamical system (3.8a–c) seems more pertinent in the case of the epidemic model under consideration.

## 3.4 Proofs

This section is devoted to establishing Theorems 3.3.2 and 3.3.3. As we will see, the proof of Theorem 3.3.2 has the intrinsic difficulty of all convergence results for measure-valued processes, with some technical difficulty arising from the infectiousness being a discontinuous function of the remaining infectious period. This will become apparent in the proof of forthcoming Proposition 3.4.8. Nevertheless, it allows us to obtain a deterministic prediction of the dynamics of the structure type distributions during the course of an epidemic. At this level, the limiting object is rich, allowing it to convey detailed information on the distribution of remaining infectious periods within structures. This however comes at the cost of an infinite-dimensional limiting object, which motivates the interest in trying to further reduce its dimension by adopting a coarser population description. In the case where  $\nu$  is the exponential distribution, Theorem 3.3.3 shows that this actually is possible, the final reduced model taking the form of dynamical system (3.8a–c). As mentioned previously, the existence of an asymptotically exact, closed, finite-dimensional ODE-system capturing the epidemic dynamics was not obvious from the beginning. Indeed, in order to obtain this result, we need to show that structuring the population of previously contaminated individuals by remaining infectious period is not necessary to handle the correlation of epidemic states of structures sharing a common infected. While this result fundamentally relies on the memory-less property of the exponential distribution, it will demand some effort, as illustrated in forthcoming Propositions 3.4.12 and 3.4.14.

### 3.4.1 Proof of Theorem 3.3.2

Let us start with the proof of Theorem 3.3.2. It follows a classical scheme, establishing tightness of  $(\zeta^K)_{K \geq 1}$ , whose limiting values are shown to satisfy Equation (3.6). Uniqueness of the solutions of this equation given the initial condition then ensures the desired convergence result. In particular, the proof is inspired by Fournier and Méléard, 2004 and Tran, 2006.

#### Uniqueness and continuity of the solution of Equation (3.6).

We are first going to establish a uniqueness result for the solutions of Equation (3.6). Notice that we do not need to prove existence of solutions in this section, as forthcoming Proposition 3.4.8 constructs such solutions as limiting values of  $(\zeta^K)_{K \geq 1}$ .

Let us start with a technical lemma, whose proof we present for sake of completeness.

**Lemma 3.4.1.** *Let  $f \in \mathcal{C}_b(E, \mathbb{R})$ . There exists a sequence  $(f_k)_{k \geq 1}$  taking values in  $\mathcal{C}_b^1(E, \mathbb{R})$  such that  $f_k$  converges simply to  $f$  and  $\sup_{k \geq 1} \|f_k\|_\infty \leq \|f\|_\infty$ .*

*Proof.* Consider a mollifier  $\psi$ , i.e.  $\psi \in \mathcal{C}^\infty(\mathbb{R}^{n_{\max}})$  is of compact support in  $\mathbb{R}^{n_{\max}}$ , its mass  $\int_{\mathbb{R}^{n_{\max}}} \psi(x) dx$  equals 1, and for  $k \geq 1$ , the function  $\psi_k : x \mapsto k^{n_{\max}} \psi(kx)$  satisfies  $\lim_{k \rightarrow \infty} \psi_k = \delta_0$  in the sense of distributions. Let  $f \in \mathcal{C}_b(E, \mathbb{R})$ . Define the sequence  $(f_k)_{k \geq 1}$  as follows:

$$f_k : (n, s, \tau) \in E \mapsto f(n, s, \cdot) * \psi_k(\tau).$$

Then, for any  $x \in E$ , by definition of  $(\psi_k)_{k \geq 1}$  it holds that  $\lim_{k \rightarrow \infty} f_k(x) = f(x)$ . Further, as  $\psi_k$  is of integral 1 for any  $k$ , it is obvious that for any  $k$ ,  $\|f_k\|_\infty \leq \|f\|_\infty$ . Finally, it also follows from the usual properties of convolution that for any  $k \geq 1$ ,  $f_k$  is smooth



with respect to its last variable and the corresponding partial derivatives are bounded, hence  $f_k \in \mathcal{C}_b^1(E, \mathbb{R})$ .  $\square$

We may now turn to the main result of this paragraph. With slight abuse of notation, for an element  $\eta = (\eta_1, \eta_2) \in \mathfrak{M}_1$ , we define its total variation norm by  $\|\eta\|_{TV} = \|\eta_1\|_{TV} \vee \|\eta_2\|_{TV}$ .

**Proposition 3.4.2.** *Let  $\eta_\star \in \mathfrak{M}_1$ . Then Equation (3.6) admits at most one measure-valued solution  $\eta$  which belongs to  $\mathcal{C}(\mathbb{R}_+, (\mathfrak{M}_1, \|\cdot\|_{TV}))$ , such that  $\eta_0 = \eta_\star$ .*

From now on, for  $X \in \{H, W\}$ , define  $n_X = \langle \eta_0^X, \mathbf{n} \rangle$ ,  $s_X(t) = \langle \eta_t^X, \mathbf{s} \rangle$  and  $i_X(t) = \langle \eta_t^H, \mathbf{i} \rangle$ . Let us establish the proposition.

*Proof.* First, notice that it follows immediately from Equation (3.6) that  $\langle \eta_T^X, 1 \rangle = \langle \eta_0^X, 1 \rangle$  for  $X \in \{H, W\}$ , thus  $\eta_0 \in \mathfrak{M}_1$  implies that for any  $T \geq 0$ ,  $\eta_T \in \mathfrak{M}_1$ .

Let us show that any solution  $\eta$  of Equation (3.6) belongs to  $\mathcal{C}(\mathbb{R}_+, (\mathfrak{M}_1, \|\cdot\|_{TV}))$ . In order to do so, it is enough to show that  $\eta^X \in \mathcal{C}(\mathbb{R}_+, (\mathcal{M}_1(E), \|\cdot\|_{TV}))$  for any  $X \in \{H, W\}$ . We are going to detail the proof for  $\eta^H$  only, as  $\eta^W$  can be handled in the same way.

Let  $T \geq 0$  and  $g \in \mathcal{C}_b^1(E, \mathbb{R})$  such that  $\|g\|_\infty \leq 1$ . Consider the function defined by

$$\forall (t, x) \in \mathbb{R} \times E, \quad f_t(x) = g(\Psi(x, T, t))$$

and recall that  $f_t^{\mathcal{I}}(x) = \langle \nu, f_t(\mathbf{j}(x, \cdot)) \rangle$ . Then by definition,  $f_T(x) = g(x)$ . It follows from the assumption  $g \in \mathcal{C}_b^1(E, \mathbb{R})$  that  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$ . The advantage of this construction is that  $t \mapsto \Psi(x, T, t)$  corresponds to a reversal of time, which cancels out the deterministic dynamics described by the differential operator  $\mathcal{A}$ . Indeed, letting  $x = (n, s, \tau)$ , a brief computation shows that

$$\partial_t f_t(x) = \sum_{k=1}^{n-s} \partial_{\tau_k} g(\Psi(x, T, t)) \quad \text{and} \quad \partial_{\tau_k} f_t(x) = \partial_{\tau_k} g(\Psi(x, T, t)),$$

which yields that  $\mathcal{A}f_t(x) = 0$  for all  $(t, x) \in \mathbb{R} \times E$ . Using the fact that  $\langle \eta_T^H, g \rangle = \langle \eta_T^H, f_T \rangle$ , it follows from Equation (3.6) that

$$\begin{aligned} \langle \eta_T^H, g \rangle &= \langle \eta_0^H, f_0 \rangle + \lambda_H \int_0^T \langle \eta_t^H, \mathbf{si}(f_t^{\mathcal{I}} - f_t) \rangle dt \\ &\quad + \lambda_W \int_0^T \frac{1}{s_W(t)} \langle \eta_t^W, \mathbf{si} \rangle \langle \eta_t^H, \mathbf{s}(f_t^{\mathcal{I}} - f_t) \rangle dt + \beta_G \int_0^T \frac{i_H(t)}{n_H} \langle \eta_t^H, \mathbf{s}(f_t^{\mathcal{I}} - f_t) \rangle dt. \end{aligned}$$

Recall that  $i_H(t) \leq n_{\max}$  and  $\frac{i_H(t)}{n_H} \leq 1$  since for any  $x \in E$ ,  $\mathbf{i}(x) \leq \mathbf{n}(x)$ . We may notice that the following inequalities hold, as for any  $t$ ,  $\|f_t\|_\infty \leq 1$ :

$$\begin{aligned} \langle \eta_t^H, \mathbf{si}(f_t^{\mathcal{I}} - f_t) \rangle &\leq 2(n_{\max})^2, & \langle \eta_t^H, \mathbf{s}(f_t^{\mathcal{I}} - f_t) \rangle &\leq 2n_{\max}, \\ \frac{1}{s_W(t)} \langle \eta_t^W, \mathbf{si} \rangle &\leq n_{\max}, & \frac{i_H(t)}{n_H} \langle \eta_t^H, \mathbf{s}(f_t^{\mathcal{I}} - f_t) \rangle &\leq 2n_{\max}. \end{aligned} \quad (3.10)$$

Let  $C = 2n_{\max}(\lambda_H n_{\max} + \lambda_W n_{\max} + \beta_G)$  and let  $\epsilon \in \mathbb{R}$ . It then follows from Inequalities (3.10) that

$$|\langle \eta_T^H - \eta_{T+\epsilon}^H, g \rangle| \leq C|T - (T + \epsilon)| = C|\epsilon|. \quad (3.11)$$

Consider now  $h \in \mathcal{C}_b(E, \mathbb{R})$  such that  $\|h\|_\infty \leq 1$ . Lemma 3.4.1 ensures that there exists a sequence  $(g_k)_{k \geq 1}$  taking values in  $\mathcal{C}_b^1(E, \mathbb{R})$  which converges simply to  $h$  and such that  $\|g_k\|_\infty \leq 1$ . By dominated convergence, this implies that

$$|\langle \eta_T^H - \eta_{T+\epsilon}^H, h \rangle| \leq C|T - (T + \epsilon)| = C|\epsilon|.$$

As  $E$  is a Polish space, it follows from Proposition A.6.1 of Tran, 2006 that

$$\|\eta_T^H - \eta_{T+\epsilon}^H\|_{TV} = \sup_{h \in \mathcal{C}_b(E, \mathbb{R}) : \|h\|_\infty \leq 1} |\langle \eta_T^H - \eta_{T+\epsilon}^H, h \rangle| \leq C|\epsilon|. \quad (3.12)$$

As this holds for any  $\epsilon$  and any  $T \geq 0$ , the strong continuity of  $\eta^H$  is established.

It remains to establish uniqueness of the solution  $\eta$  of Equation (3.6) with initial condition  $\eta_0 = \eta_*$ . We will once more establish uniqueness component-wise, and focus on  $\eta^H$  as  $\eta^W$  is treated in a similar fashion.

Let  $\eta, \bar{\eta}$  be two solutions of Equation (3.6) with initial condition  $\eta_*$ . Let  $\bar{i}_H(t) = \langle \bar{\eta}_t^H, \mathbf{i} \rangle$  and define  $\bar{s}_W(t)$  in analogous manner. As before, let  $T \geq 0$ . Consider again  $g \in \mathcal{C}_b^1(E, \mathbb{R})$  such that  $\|g\|_\infty \leq 1$ , and define  $f_t$  and  $f_t^{\mathcal{I}}$  as previously. Then

$$\begin{aligned} |\langle \eta_T^H - \bar{\eta}_T^H, g \rangle| &\leq \lambda_H \int_0^T |\langle \eta_t^H - \bar{\eta}_t^H, \mathbf{si}(f_t^{\mathcal{I}} - f_t) \rangle| dt \\ &+ \lambda_W \int_0^T \left| \frac{1}{s_W(t)} \langle \eta_t^W, \mathbf{si} \rangle \langle \eta_t^H, \mathbf{s}(f_t^{\mathcal{I}} - f_t) \rangle - \frac{1}{\bar{s}_W(t)} \langle \bar{\eta}_t^W, \mathbf{si} \rangle \langle \bar{\eta}_t^H, \mathbf{s}(f_t^{\mathcal{I}} - f_t) \rangle \right| dt \\ &+ \frac{\beta_G}{m_H} \int_0^T |i_H(t) \langle \eta_t^H, \mathbf{s}(f_t^{\mathcal{I}} - f_t) \rangle - \bar{i}_H(t) \langle \bar{\eta}_t^H, \mathbf{s}(f_t^{\mathcal{I}} - f_t) \rangle| dt. \end{aligned}$$

Proceeding similarly as in Inequalities (3.10), we obtain that

$$\begin{aligned} |\langle \eta_T^H - \bar{\eta}_T^H, g \rangle| &\leq \frac{C}{2} \int_0^T |\langle \eta_t^H - \bar{\eta}_t^H, f_t^{\mathcal{I}} - f_t \rangle| dt \\ &\leq \frac{C}{2} \left( \int_0^T |\langle \eta_t^H - \bar{\eta}_t^H, f_t^{\mathcal{I}} \rangle| dt + \int_0^T |\langle \eta_t^H - \bar{\eta}_t^H, f_t \rangle| dt \right). \end{aligned}$$

On the one hand, by definition,  $f_t \in \mathcal{C}_b(E, \mathbb{R})$  and  $\|f_t\|_\infty \leq 1$ . Thus  $|\langle \eta_t^H - \bar{\eta}_t^H, f_t \rangle| \leq \|\eta_t^H - \bar{\eta}_t^H\|_{TV}$ . On the other hand, it follows from the usual criterion of continuity for parametric integrals that  $f_t^{\mathcal{I}}$  is continuous on  $E$  and  $\|f_t^{\mathcal{I}}\|_\infty \leq \|f_t\|_\infty \leq 1$ . As a consequence,  $|\langle \eta_t^H - \bar{\eta}_t^H, f_t^{\mathcal{I}} \rangle| \leq \|\eta_t^H - \bar{\eta}_t^H\|_{TV}$ . Hence

$$|\langle \eta_T^H - \bar{\eta}_T^H, g \rangle| \leq C \int_0^T \|\eta_t^H - \bar{\eta}_t^H\|_{TV} dt.$$

We then may follow the same steps that allowed to establish Equation (3.12) from Equation (3.11), and obtain that

$$\|\eta_T^H - \bar{\eta}_T^H\|_{TV} \leq C \int_0^T \|\eta_t^H - \bar{\eta}_t^H\|_{TV} dt.$$

Gronwall's lemma then assures that

$$\forall t \in [0, T], \quad \|\eta_t^H - \bar{\eta}_t^H\|_{TV} = 0.$$

One obtains the analogous result for  $\eta^W$  in the same manner. As  $T \geq 0$  is arbitrary, this concludes the proof.  $\square$

**Tightness of  $(\zeta^K)_{K \geq 1}$  in  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), w))^2$** 

Let us now turn to the tightness of  $(\zeta^K)_{K \geq 1}$  in  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), w))^2$ , where  $w$  designates the weak topology on  $\mathcal{M}_F(E)$ . We start by establishing the following preliminary result, whose proof relies on the chain rule.

**Lemma 3.4.3.** *Let  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$ . Then for any  $T \geq t_0 \geq 0$ , for any  $x \in E$ ,*

$$f(T, \Psi(x, T, t_0)) = f(t_0, x) + \int_{t_0}^T \mathcal{A}f(t, \Psi(x, t, t_0)) dt.$$

*Proof.* For  $x = (n, s, \tau) \in E$  and  $t_0 \in \mathbb{R}_+$ , define

$$g_{t_0, x} : [t_0, +\infty) \rightarrow \mathbb{R}, \quad T \mapsto f(T, \Psi(x, T, t_0)).$$

Let us start by noticing that for any  $(t_0, x) \in \mathbb{R}_+ \times E$ ,  $g_{t_0, x} \in \mathcal{C}^1(\mathbb{R}_+)$ . Indeed,  $g_{t_0, x} = f_{n, s} \circ h_{t_0, x}$ , where

$$f_{n, s} : \mathbb{R}_+ \times \mathbb{R}^{n_{\max}} \rightarrow \mathbb{R}, \quad (u, \mathbf{v}) \mapsto f(u, (n, s, \mathbf{v}))$$

and  $h_{t_0, x} : \mathbb{R}_+ \rightarrow \mathbb{R}^{1+n_{\max}}, \quad t \mapsto \left( t, \tau - \sum_{k=1}^{n-s} (t - t_0) e_k \right).$

The chain rule and a quick computation of the differentiable of  $h_{t_0, x}$  yields that for every  $t \geq t_0$ ,

$$\frac{d}{dt} g_{t_0, x}(t) = \partial_1 f_{n, s}(h_{t_0, x}(t)) - \sum_{k=1}^{n-s} \partial_{k+1} f_{n, s}(h_{t_0, x}(t)).$$

Notice that on the one hand,  $\partial_1 f_{n, s}(u, \mathbf{v}) = \partial_t f(u, (n, s, \mathbf{v}))$  and on the other,  $k \geq 2$ ,  $\partial_k f_{n, s}(u, \mathbf{v}) = \partial_{\tau_{k-1}} f(u, (n, s, \mathbf{v}))$ . As a consequence, we have shown that for any  $(t_0, x) \in \mathbb{R}_+ \times E$ , for every  $t \geq t_0$ ,  $g_{t_0, x}$  is differentiable at  $t$  and satisfies

$$\frac{d}{dt} g_{t_0, x}(t) = \mathcal{A}f(t, \Psi(x, t, t_0)).$$

This concludes the proof. □

Throughout the section, we use the notation  $\mathcal{S} = \{H, W, G\}$ . Also, for any  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$ , let  $f_t(x) = f(t, x)$  for any  $(t, x) \in \mathbb{R}_+ \times E$ . Finally, we define for any continuous bounded function  $g : \mathbb{R}_+ \times E \rightarrow \mathbb{R}$ , for any  $t \geq 0$  and  $u = (\boldsymbol{\theta}, k, \ell, \sigma) \in \bigcup_{Y \in \mathcal{S}} U_Y$ :

$$g_{t, u}^H = g(t, j(x_k^H(t), \sigma)) - g(t, x_k^H(t)) \quad \text{and} \quad g_{t, u}^W = g(t, j(x_\ell^W(t), \sigma)) - g(t, x_\ell^W(t)).$$

**Proposition 3.4.4.** *Consider  $\zeta^K$  as introduced in Proposition 3.2.1. For any  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$ ,  $T \geq 0$  and  $X \in \{H, W\}$ ,*

$$\langle \zeta_T^{X|K}, f_T \rangle = \langle \zeta_0^{X|K}, f_0 \rangle + \int_0^T \langle \zeta_t^{X|K}, \mathcal{A}f_t \rangle dt + \frac{1}{K_X} \sum_{Y \in \mathcal{S}} \int_0^T \int_{U_Y} \mathcal{I}_Y(t-, u) f_{t-, u}^X Q_Y(dt, du).$$

*Proof.* Let  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$  and  $X \in \{H, W\}$ . Recall that, by definition, for any bounded function  $g : E \rightarrow \mathbb{R}$ , for any  $T \geq t \geq 0$  and  $u = (\boldsymbol{\theta}, k, \ell, \sigma) \in \bigcup_{Y \in \mathcal{S}} U_Y$ :

$$\langle \Delta_H(u, T, t), g \rangle = g(\Psi(j(x_k^H(t-), \sigma), T, t)) - g(\Psi(x_k^H(t-), T, t)),$$

and  $\langle \Delta_W(u, T, t), g \rangle$  is defined analogously, by replacing  $H$  by  $W$  and  $k$  by  $\ell$ .

From Equation (3.5), it follows that

$$\langle \zeta_T^{X|K}, f_T \rangle = \frac{1}{K_X} \left( \sum_{j=1}^{K_X} f_T(\Psi(x_j^X(0), T, 0)) + \sum_{Y \in \mathcal{S}} \int_0^T \int_{U_Y} \mathcal{I}_Y(t-, u) \langle \Delta_X(u, T, t), f_T \rangle Q_Y(dt, du) \right).$$

Using the result from Lemma 3.4.3, this becomes:

$$\begin{aligned} \langle \zeta_T^{X|K}, f_T \rangle &= \frac{1}{K_X} \sum_{j=1}^{K_X} \left( f_0(x_j^X(0)) + \int_0^T \mathcal{A}f_t(\Psi(x_j^X(0), t, 0)) dt \right) \\ &\quad + \frac{1}{K_X} \sum_{Y \in \mathcal{S}} \int_0^T \int_{U_Y} \mathcal{I}_Y(t-, u) \left( \int_t^T \langle \Delta_X(u, z, t), \mathcal{A}f_z \rangle dz \right) Q_Y(dt, du) \\ &\quad + \frac{1}{K_X} \sum_{Y \in \mathcal{S}} \int_0^T \int_{U_Y} \mathcal{I}_Y(t-, u) f_{t-,u}^X Q_Y(dt, du). \end{aligned}$$

It follows from the definition of  $\mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$  that both  $f$  and  $\mathcal{A}f$  are bounded, hence we may apply Fubini's theorem to obtain that

$$\begin{aligned} \langle \zeta_T^{X|K}, f_T \rangle &= \frac{1}{K_X} \sum_{j=1}^{K_X} f_0(x_j^X(0)) + \frac{1}{K_X} \sum_{Y \in \mathcal{S}} \int_0^T \int_{U_Y} \mathcal{I}_Y(t-, u) f_{t-,u}^X Q_Y(dt, du) \\ &\quad + \frac{1}{K_X} \int_0^T \left( \sum_{j=1}^{K_X} \mathcal{A}f_z(\Psi(x_j^X(0), z, 0)) + \sum_{Y \in \mathcal{S}} \int_0^z \int_{U_Y} \mathcal{I}_Y(t-, u) \langle \Delta_X(u, z, t), \mathcal{A}f_z \rangle Q_Y(dt, du) \right) dz. \end{aligned}$$

The first sum on the right-hand side equals  $\langle \zeta_0^{X|K}, f_0 \rangle$ . From the second line, one recognizes in the integrand the definition of  $\langle \zeta_z^{X|K}, \mathcal{A}f_z \rangle$  from Equation (3.5). This yields the desired result.  $\square$

For  $Y \in \mathcal{S}$ , let

$$\tilde{Q}_Y(dt, du) = Q_Y(dt, du) - dt\mu_Y(du)$$

be the compensated martingale-measure associated to  $Q_Y$ .

It follows that, for  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$  and  $X \in \{H, W\}$ ,

$$\langle \zeta_T^{X|K}, f_T \rangle = M_T^{X|K}(f) + V_T^{X|K}(f),$$

where we define

$$M_T^{X|K}(f) = \frac{1}{K_X} \sum_{Y \in \mathcal{S}} \int_0^T \int_{U_Y} \mathcal{I}_Y(t-, u) f_{t-,u}^X \tilde{Q}_Y(dt, du)$$

and

$$V_T^{X|K}(f) = \langle \zeta_0^{X|K}, f_0 \rangle + \int_0^T \langle \zeta_t^{X|K}, \mathcal{A}f_t \rangle dt + \frac{1}{K_X} \sum_{Y \in \mathcal{S}} \int_0^T \int_{U_Y} \mathcal{I}_Y(t, u) f_{t,u}^X \mu_Y(du) dt.$$

**Proposition 3.4.5.** *Let  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$  and  $X \in \{H, W\}$ . Then  $(M_T^{X|K}(f))_{T \geq 0}$  is a square integrable martingale. Using the same notations as in Theorem 3.3.2, its quadratic variation is given by*

$$\langle M^{X|K}(f) \rangle_T = \frac{1}{K_X} \int_0^T \langle \zeta_t^{X|K}, \mathcal{H}_t^X((f_t^2)^{\mathcal{I}} - 2f_t^{\mathcal{I}} f_t + f_t^2) \rangle dt,$$

where for any  $t \geq 0$  and  $x \in E$ ,

$$\mathcal{H}_t^X(x) = \beta_G \frac{I_H(t)}{N_H} \mathbf{s}(x) + \lambda_X \mathbf{s}(x) \mathbf{i}(x) + \lambda_{\bar{X}} \frac{\langle \zeta_t^{\bar{X}|K}, \mathbf{s} \mathbf{i} \rangle}{S_{\bar{X}}(t)} \mathbf{s}(x).$$

*Proof.* Let  $f \in \mathcal{C}_b^1(E, \mathbb{R})$  and  $X \in \{H, W\}$ . Consider  $M_T^{X|K}(f)$ , which can be written as

$$M_T^{X|K}(f) = M_{X,T}^{X|K}(f) + M_{\bar{X},T}^{X|K}(f) + M_{G,T}^{X|K}(f),$$

where, for  $Y \in \mathcal{S}$  and  $T \geq 0$ ,

$$M_{Y,T}^{X|K}(f) = \int_0^T \int_{U_Y} \frac{1}{K_X} \mathcal{I}_Y(t-, u) f_{t,u}^X \tilde{Q}_Y(dt, du).$$

Suppose that for any  $Y \in \mathcal{S}$ ,

$$\mathbb{E} \left[ \int_0^T \int_{U_Y} \left( \frac{1}{K_X} \mathcal{I}_Y(t, u) f_{t,u}^X \right)^2 \mu_Y(du) dt \right] < \infty,$$

then for all  $Y \in \mathcal{S}$ ,  $(M_{Y,T}^{X|K}(f))_{T \geq 0}$  is a square integrable martingale (Meleard and Bansaye, 2015), implying that  $(M_T^{X|K}(f))_{T \geq 0}$  also is a square integrable martingale. As  $Q^{H|K}$ ,  $Q^{W|K}$  and  $Q^{G|K}$  are independent, it follows that

$$\langle M^{X|K}(f) \rangle_T = \sum_{Y \in \mathcal{S}} \langle M_Y^{X|K}(f) \rangle_T.$$

It thus is enough to study  $(M_{Y,T}^{X|K}(f))_{T \geq 0}$  for all  $Y \in \mathcal{S}$ . In the following, we will detail the necessary computations in the case  $X = H$ , the case  $X = W$  being similar.

Consider the case  $Y = H$ . Start by noticing that  $\sum_{\ell=1}^{K_W} s_\ell^W(t) = K_W S_W(t)$ , and that for any  $k \in \llbracket 1, K_H \rrbracket$  and  $t \in [0, T]$ ,  $s_k^H(t)$  and  $i_k^H(t)$  are less than  $n_{\max}$ , almost surely. Hence, replacing  $S(t)$  by  $K_W S_W(t)$  in  $\mathcal{I}_H$ ,

$$\begin{aligned} \mathbb{E} \left[ \langle M_H^{H|K}(f) \rangle_T \right] &= \mathbb{E} \left[ \int_0^T \int_{U_H} \left( \frac{1}{K_H} \mathcal{I}_H(t, u) f_{t,u}^H \right)^2 \mu_H(du) dt \right] \\ &= \mathbb{E} \left[ \int_0^T \frac{1}{K_H^2} \sum_{k=1}^{K_H} \lambda_H s_k^H(t) i(\tau_k^H(t)) \langle \nu, (f_t(j(x_k^H(t), \cdot)) - f_t(x_k^H(t)))^2 \rangle dt \right] \\ &\leq \frac{1}{K_H} \lambda_H (n_{\max})^2 4 \|f\|_\infty^2 T. \end{aligned} \quad (3.13)$$

Since further  $K_H \geq K/n_{\max}$  and  $\|f\|_\infty^2 < \infty$ , we obtain that

$$\mathbb{E} \left[ \langle M_H^{H|K}(f) \rangle_T \right] \leq \frac{4}{K} \lambda_X (n_{\max})^3 \|f\|_\infty^2 T < \infty. \quad (3.14)$$

Thus  $(M_{H,t}^{H|K}(f))_{t \geq 0}$  is a square integrable martingale whose quadratic variation is given by

$$\begin{aligned} \langle M_H^{H|K}(f) \rangle_T &= \int_0^T \int_{U_H} \left( \frac{1}{K_H} \mathcal{I}_H(t, u) f_{t,u}^H \right)^2 \mu_H(du) dt \\ &= \frac{\lambda_H}{K_H} \int_0^T \langle \zeta_t^{H|K}, \mathbf{s}((f_t^2)^{\mathcal{I}} - 2f_t^{\mathcal{I}} f_t + f_t^2) \rangle dt, \end{aligned}$$

using the computations from Equation (3.13).

Similarly,  $(M_{W,T}^{H|K}(f))_{T \geq 0}$  is a square integrable martingale of quadratic variation

$$\langle M_W^{H|K}(f) \rangle_T = \lambda_W \int_0^T \frac{\langle \zeta_t^{W|K}, \mathbf{s} \rangle}{S_W(t)} \langle \zeta_t^{H|K}, \mathbf{s}((f_t^2)^{\mathcal{I}} - 2f_t^{\mathcal{I}} f_t + f_t^2) \rangle dt.$$

Further, for the case  $Y = G$ , let us use the equalities  $S(t) = K_H S_H(t)$  and  $I(t) = K_H I_H(t)$ . As  $I(t)/K = I_H(t)/N_H \leq 1$  almost surely, we obtain that

$$\mathbb{E} \left[ \langle M_G^{H|K}(f) \rangle_T \right] = \mathbb{E} \left[ \int_0^T \int_{U_G} \left( \frac{1}{K_H} \mathcal{I}_G(t, u) f_{t,u}^H \right)^2 \mu_G(du) dt \right] \leq \frac{4}{K} \beta_G (n_{\max})^2 \|f\|_{\infty}^2 T.$$

As before,  $(M_{G,T}^{H|K}(f))_{T \geq 0}$  thus is a square integrable martingale of quadratic variation given by

$$\langle M_G^{H|K}(f) \rangle_T = \frac{1}{K_H} \beta_G \int_0^T \frac{I_H(t)}{N_H} \langle \zeta_t^{H|K}, \mathbf{s}((f_t^2)^{\mathcal{I}} - 2f_t^{\mathcal{I}} f_t + f_t^2) \rangle dt.$$

This yields the desired result for  $(M_T^{H|K}(f))_{T \geq 0}$ . The proof is concluded by proceeding in the same way for  $(M_T^{W|K}(f))_{T \geq 0}$ .  $\square$

We are now ready to focus on the tightness of  $(\zeta^K)_{K \geq 1}$ , endowing  $\mathcal{M}_F(E)$  with the vague topology  $v$  as a first step.

**Proposition 3.4.6.** *Under the assumptions of Theorem 3.3.2, the sequence  $(\zeta^K)_{K \geq 1}$  is tight in  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), v))^2$ .*

The proof relies on the fact that in order to establish tightness of  $(\zeta^K)_{K \geq 1}$ , it is enough to show that for any  $X \in \{H, W\}$ ,  $(\langle \zeta^{X|K}, f \rangle)_{K \geq 1}$  is tight for a large enough set of test functions  $f$  (Roelly-Coppoletta, 1986). This in turn is ensured using the Aldous (Aldous, 1978) and Rebolledo (Joffe and Metivier, 1986) criteria, whose application is straightforward thanks to the upper bounds established in the previous proof.

*Proof.* Once more, we will proceed component-wise and show that  $(\zeta^{H|K})_{K \geq 1}$  and  $(\zeta^{W|K})_{K \geq 1}$  are both tight in  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), v))$ .

Let us focus on  $(\zeta^{H|K})_{K \geq 1}$ . According to Theorem 2.1 of Roelly-Coppoletta, 1986, it is sufficient to show that for any function  $f$  belonging to a dense subset of

$$\mathcal{C}_0(E, \mathbb{R}) = \left\{ f : E \rightarrow \mathbb{R} \text{ continuous s.t. } \lim_{\|x\|_{\infty} \rightarrow \infty} |f(x)| = 0 \right\},$$

the sequence  $(\langle \zeta^{H|K}, f \rangle)_{K \geq 1}$  is tight in  $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ . Notice that by density of the set of smooth compactly supported functions in  $C_0(\mathbb{R}^{n_{\max}})$  endowed with the uniform norm, it

follows that  $\mathcal{C}_0(E, \mathbb{R}) \cap \mathcal{C}_b^1(E, \mathbb{R})$  is also dense in  $\mathcal{C}_0(E, \mathbb{R})$  endowed with the uniform norm. Thus, let us consider  $f \in \mathcal{C}_0(E, \mathbb{R}) \cap \mathcal{C}_b^1(E, \mathbb{R})$ .

According to the Aldous (Aldous, 1978) and Rebolledo (Joffe and Metivier, 1986) criteria, in order to prove the tightness of  $(\langle \zeta^{H|K}, f \rangle)_{K \geq 1}$ , it is enough to show that:

- (i) For any  $t$  belonging to a dense subset  $\mathcal{T}$  of  $\mathbb{R}^+$ ,  $(\langle M^{H|K}(f) \rangle_t)_{K \geq 0}$  and  $(V_t^{H|K}(f))_{K \geq 0}$  are tight in  $\mathbb{R}$ .
- (ii) For any  $T \geq 0$ , for any  $\epsilon, \alpha > 0$ , there exist  $\delta > 0$  and  $K_0 \in \mathbb{N}$  such that for any two sequences of stopping times  $(S_K)_{K \geq 1}$  and  $(T_K)_{K \geq 1}$  satisfying  $S_K \leq T_K \leq T$  for all integers  $K$ ,

$$\sup_{K \geq K_0} \mathbb{P} \left( |\langle M^{H|K}(f) \rangle_{S_K} - \langle M^{H|K}(f) \rangle_{T_K}| \geq \alpha, T_K \leq S_K + \delta \right) \leq \epsilon$$

$$\text{and } \sup_{K \geq K_0} \mathbb{P} \left( |V_{S_K}^{H|K}(f) - V_{T_K}^{H|K}(f)| \geq \alpha, T_K \leq S_K + \delta \right) \leq \epsilon.$$

Notice that, in order to establish (i), it is enough to show that for any  $t \geq 0$ ,

$$\sup_{K \geq 1} \mathbb{E} \left[ |\langle M^{H|K}(f) \rangle_t| \right] < \infty \text{ and } \sup_{K \geq 1} \mathbb{E} \left[ |V_t^{H|K}(f)| \right] < \infty.$$

Recalling that  $C = 2n_{\max}(\lambda_H n_{\max} + \lambda_W n_{\max} + \beta_G)$ , it follows from Equation (3.14) that

$$\mathbb{E}[|\langle M^{H|K}(f) \rangle_t|] \leq \frac{1}{K} 2n_{\max} C \|f\|_{\infty}^2 t.$$

Similar computations yield that

$$\mathbb{E}[|V_t^{H|K}(f)|] \leq \|f\|_{\infty} + \|\mathcal{A}f\|_{\infty} + C \|f\|_{\infty} t.$$

As  $f \in \mathcal{C}_b^1(E, \mathbb{R})$ , this implies that (i) holds.

It remains to check (ii). Let  $\epsilon, \alpha > 0$ , and consider two sequences of stopping times  $(S_K)_{K \geq 1}$  and  $(T_K)_{K \geq 1}$  satisfying  $S_K \leq T_K \leq T$  for all integers  $K$ . As previously, using Equation (3.14), we obtain the following upper bound:

$$\mathbb{E} \left[ |\langle M^{H|K}(f) \rangle_{S_K} - \langle M^{H|K}(f) \rangle_{T_K}| \mid T_K \leq S_K + \delta \right] \leq \mathbb{E} \left[ \int_{S_K}^{T_K} dt \mid T_K \leq S_K + \delta \right] \frac{2}{K} n_{\max} C \|f\|_{\infty}^2$$

$$\leq \frac{\delta}{K} 2n_{\max} C \|f\|_{\infty}^2.$$

Hence, using conditional Markov's inequality,

$$\mathbb{P} \left( |\langle M^{H|K}(f) \rangle_{S_K} - \langle M^{H|K}(f) \rangle_{T_K}| \geq \alpha, T_K \leq S_K + \delta \right) \leq \frac{\delta}{\alpha K} 2n_{\max} C \|f\|_{\infty}^2. \quad (3.15)$$

Proceeding similarly, we also obtain that

$$\mathbb{P} \left( |V_{S_K}^{H|K}(f) - V_{T_K}^{H|K}(f)| \geq \alpha, T_K \leq S_K + \delta \right) \leq \frac{\delta}{\alpha} (\|\mathcal{A}f\|_{\infty} + C \|f\|_{\infty}). \quad (3.16)$$

Equations (3.15) and (3.16) imply the existence of  $\delta$  and  $K_0$  such that (ii) is satisfied. Naturally,  $\zeta^{W|K}$  can be handled analogously. This concludes the proof.  $\square$

Finally, this result on the tightness of  $(\zeta^K)_{K \geq 1}$  in  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), v))^2$  lets us establish the main result of this subsection.

**Proposition 3.4.7.** *Under the assumptions of Theorem 3.3.2, the sequence  $(\zeta^K)_{K \geq 1}$  is tight in  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), w))^2$ .*

*Proof.* Let  $X \in \{H, W\}$ . Tightness in  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), w))$  of  $(\zeta^{X|K})_{K \geq 1}$  will be shown using Theorem 1.1.8 from Tran, 2014, which we state in our setting for the sake of completeness. Let  $\Phi : z \in \mathbb{R} \mapsto 6z^2 - 15z^4 + 10z^3$  and for  $N \geq 1$ , define smooth approximations of  $x \in E \mapsto \mathbf{1}_{\{\|\tau(x)\|_\infty \geq N\}}$  by:

$$\forall x \in E, \forall N \geq 1, \phi_N(x) = \Phi(0 \vee (\|\tau(x)\|_\infty - (N - 1)) \wedge 1).$$

Then in order to ensure the tightness of  $(\zeta^{X|K})_{K \geq 1}$  in  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), w))$ , it is sufficient to show that for any  $T \geq 0$ , the following conditions hold:

(i) There exists a family of functions  $F$  which is dense in  $\mathcal{C}_0(E, \mathbb{R})$  and stable under addition, such that for any  $f \in F \cup \{x \in E \mapsto 1\}$ , the sequence  $(\langle \zeta^{X|K}, f \rangle)_{K \geq 1}$  is tight in  $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ .

(ii)

$$\lim_{N \rightarrow \infty} \limsup_{K \rightarrow \infty} \mathbb{E}[\sup_{t \leq T} \langle \zeta_t^{X|K}, \phi_N \rangle] = 0.$$

(iii) Any limiting value of  $(\zeta^{X|K})_{K \geq 1}$ , if it exists, belongs to  $\mathcal{C}([0, T], (\mathcal{M}_F(E), w))$ .

The proof hence consists in checking those assumptions. Let  $T \geq 0$ , and consider  $X = H$ , as the case  $X = W$  can be treated similarly. We may see that (i) is satisfied, as we have shown in the proof of Proposition 3.4.6 that for any  $f \in \mathcal{C}_0(E, \mathbb{R}) \cap \mathcal{C}_b^1(E, \mathbb{R})$ ,  $(\langle \zeta^{H|K}, f \rangle)_{K \geq 1}$  is tight, and further for any  $K \geq 1$ , for any  $T \geq 0$ ,  $\langle \zeta_T^{H|K}, 1 \rangle = 1$  almost surely.

Let us now turn our attention to (ii). Start by noticing that for any  $N \geq 1$  and  $x \in E$ ,

$$\phi_N(x) \leq \mathbf{1}_{\{\|\tau(x)\|_\infty \geq N-1\}} \leq f_{N-1}(x) := \sum_{i=1}^{n_{\max}} f_{N-1,i}(x)$$

where  $f_{N-1,i}(x) = \mathbf{1}_{\{\mathbf{n}(x) - \mathbf{s}(x) \geq i, |\tau_i(x)| \geq N-1\}}$ .

Let  $t \in [0, T]$ . For any  $N-1 \geq t$ ,  $x \in E$ ,  $z \in [0, t]$  and  $\sigma \geq 0$ , it holds by definition that

$$f_{N-1,i}(\Psi(j(x, \sigma), t, z)) - f_{N-1,i}(\Psi(x, t, z)) = \mathbf{1}_{\{\mathbf{n}(x) - \mathbf{s}(x) = i-1, |\sigma - (t-z)| > N-1\}} \leq \mathbf{1}_{\{\sigma > N-1\}}.$$

Hence, using Proposition 3.2.1 and the above upper bounds, it follows that almost surely,

$$\langle \zeta_t^{H|K}, \phi_N \rangle \leq \frac{1}{K_H} \sum_{k=1}^{K_H} f_{N-1}(\Psi(x_k^H(0), t, 0)) + \frac{n_{\max}}{K_H} \sum_{Y \in \mathcal{S}} \int_0^t \int_{U_Y} \mathcal{I}_Y(z-, u) \mathbf{1}_{\{\sigma > N\}} Q_Y^K(dz, du).$$

Defining as previously  $C = 2n_{\max}(\lambda_H n_{\max} + \lambda_W n_{\max} + \beta_G)$ , this leads to the following upper bound:

$$\mathbb{E}[\sup_{t \leq T} \langle \zeta_t^{H|K}, \phi_N \rangle] \leq \mathbb{E}[\sup_{t \leq T} \frac{1}{K_H} \sum_{k=1}^{K_H} f_{N-1}(\Psi(x_k^H(0), t, 0))] + \frac{CT}{2} n_{\max} \nu((N-1, +\infty)).$$



As a consequence, Assumption 3.3.1 (i) ensures that (ii) is satisfied. Notice that this assumption could actually be a little bit relaxed here, as it would be enough if the supremum over  $K$  were replaced by the limit superior over  $K \rightarrow \infty$ .

In order to check that condition (iii) holds, we will follow the arguments presented in Jourdain et al., 2012. Suppose that  $\eta^H$  is a limiting value of  $(\zeta^{H|K})_{K \geq 1}$ . By definition,

$$\sup_{t \in [0, T]} \sup_{f \in L^\infty, \|f\|_\infty \leq 1} |\langle \zeta_t^{H|K}, f \rangle - \langle \zeta_{t-}^{H|K}, f \rangle| \leq \frac{1}{K_H}.$$

As the application  $\mu \mapsto \sup_{t \in [0, T]} |\langle \mu_t, f \rangle - \langle \mu_{t-}, f \rangle|$  is continuous on  $\mathbb{D}([0, T], (\mathcal{M}_F(E), v))$  for any  $f$  in a measure-determining countable set, it follows that  $\eta^H \in \mathcal{C}([0, T], (\mathcal{M}_F(E), v))$ .

Let us now introduce  $\phi_{N, M} = \phi_N(1 - \phi_M)$ , which serves as a smooth and compactly supported approximation of  $x \in E \mapsto \mathbf{1}_{\{N \leq \|\tau(x)\|_\infty \leq M\}}$ , for  $N \leq M$ . As, on the one hand,  $\mu \mapsto \sup_{t \in [0, T]} \langle \mu_t, \phi_{N, M} \rangle$  is continuous on  $\mathbb{D}([0, T], (\mathcal{M}_F(E), v))$ , and on the other hand, for any  $K \geq 1$ ,  $\sup_{t \in [0, T]} \langle \zeta_t^{X|K}, \phi_{N, M} \rangle \leq 1$ , it follows that:

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \langle \eta_t^H, \phi_{N, M} \rangle \right] = \lim_{K \rightarrow \infty} \mathbb{E} \left[ \sup_{t \in [0, T]} \langle \zeta_t^{X|K}, \phi_{N, M} \rangle \right] \leq \limsup_{K \rightarrow \infty} \mathbb{E} \left[ \sup_{t \in [0, T]} \langle \zeta_t^{X|K}, \phi_N \rangle \right].$$

Letting  $M$  go to infinity in the left hand side, dominated convergence ensures that

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \langle \eta_t^H, \phi_N \rangle \right] \leq \limsup_{K \rightarrow \infty} \mathbb{E} \left[ \sup_{t \in [0, T]} \langle \zeta_t^{X|K}, \phi_N \rangle \right] \xrightarrow{N \rightarrow \infty} 0,$$

where the convergence of the right hand side is achieved as in the proof of (ii). In particular, it thus is possible to extract a subsequence from  $(\sup_{t \in [0, T]} \langle \eta_t^H, \phi_N \rangle)_N$  which converges almost surely to zero. This implies that for any  $\epsilon$ , there exists  $N$  such that almost surely,

$$1 - \sup_{t \in [0, T]} \langle \eta_t^H, \mathbf{1}_{\{\|\tau(\cdot)\|_\infty \leq N\}} \rangle \leq \sup_{t \in [0, T]} \langle \eta_t^H, \phi_N \rangle < \epsilon$$

Thus  $(\eta_t^H)_{t \in [0, T]}$  is almost surely tight.

Let  $g \in \mathcal{C}_b(E)$ , and let  $g_N = g(1 - \phi_N)$ . It then holds that for  $h$  small so that  $t+h \in [0, T]$ ,

$$|\langle \eta_{t+h}^H, g \rangle - \langle \eta_t^H, g \rangle| \leq |\langle \eta_{t+h}^H, g - g_N \rangle| + |\langle \eta_{t+h}^H, g_N \rangle - \langle \eta_t^H, g_N \rangle| + |\langle \eta_t^H, g - g_N \rangle|.$$

Let  $\epsilon > 0$ . As  $|g - g_N| \leq \|g\|_\infty \phi_N$ , there exists  $N_0$  such that  $\sup_{t \in [0, T]} \langle \eta_t^H, g - g_{N_0} \rangle < \epsilon/3$ . Further, as  $\eta^H \in \mathcal{C}([0, T], (\mathcal{M}_F(E), v))$ , for  $h$  small enough,  $|\langle \eta_{t+h}^H, g_{N_0} \rangle - \langle \eta_t^H, g_{N_0} \rangle| < \epsilon/3$ . This allows to conclude that  $\eta^H \in \mathcal{C}([0, T], (\mathcal{M}_F(E), w))$ , establishing (iii) and finally tightness of  $\eta^H$  in  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), w))$ .  $\square$

### Identification of the limiting values of $(\zeta^K)_{K \geq 1}$

The tightness of  $(\zeta^K)_{K \geq 1}$  in the space  $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(E), w)^2$  ensures that from any subsequence of  $(\zeta^K)_{K \geq 1}$ , one may extract a subsubsequence which converges in this space. The limits of these subsubsequences may be characterized as follows.

**Proposition 3.4.8.** *Under the assumptions of Theorem 3.3.2, all limiting values of  $(\zeta^K)_{K \geq 1}$  in  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), w))^2$  are continuous with regard to the total variation norm, and solutions of Equation (3.6).*

Before proceeding to the proof of this proposition, let us emphasize that there is some technical difficulty due to infectiousness being a discontinuous function of an individual's remaining infectious period. Indeed, tightness of  $(\zeta^K)_{K \geq 1}$  in  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), w))^2$  allows us to extract a subsequence  $(\zeta^{\varphi(K)})_{K \geq 1}$  which converges in law in this space to some limiting value  $\eta$ , and our aim is to show that  $\eta$  satisfies Equation (3.6). However, convergence in law in  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), w))$  is not enough to ensure that  $\langle \zeta_t^{X|\varphi(K)}, \mathbf{i}f \rangle$  converges in law to  $\langle \eta_t^X, \mathbf{i}f \rangle$  for  $f \in \mathcal{C}_b^1(E)$ , as  $\mathbf{i}$  is discontinuous on  $E$ . This leads to forthcoming Proposition 3.4.9.

*Proof.* Consider a subsequence  $(\zeta^{\varphi(K)})_{K \geq 1}$  of  $(\zeta^K)_{K \geq 1}$  which converges in law in the space  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), w))^2$ , and let  $\eta$  be its limit.

Notice that it follows from the Proof of Proposition 3.4.7 that  $\eta \in C(\mathbb{R}_+, (\mathcal{M}_F(E), w))^2$  almost surely. Hence, following Proposition A.6.1 of Tran, 2006, for any  $X \in \{H, W\}$ ,

$$\|\eta_T^X - \eta_{T-}^X\|_{TV} = \sup_{f \in \mathcal{C}_b(E, \mathbb{R}): \|f\|_\infty \leq 1} |\langle \eta_T^X, f \rangle - \langle \eta_{T-}^X, f \rangle| = 0 \text{ almost surely.}$$

It remains to show that  $\eta$  satisfies Equation (3.6). Let  $T \geq 0$  and  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$ , and consider the application  $\psi_T^H$  defined by

$$\begin{aligned} \psi_T^H(\eta) &= \langle \eta_T^H, f_T \rangle - \langle \eta_0^H, f_0 \rangle - \int_0^T \langle \eta_t^H, \mathcal{A}f_t \rangle dt - \int_0^T \langle \eta_t^H, \lambda_H \mathbf{si}(f_t^I - f_t) \rangle dt \\ &\quad - \int_0^T \frac{1}{\langle \eta_t^W, \mathbf{s} \rangle} \langle \eta_t^W, \lambda_W \mathbf{si} \rangle \langle \eta_t^H, \mathbf{s}(f_t^I - f_t) \rangle dt - \beta_G \int_0^T \frac{\langle \eta_t^H, \mathbf{i} \rangle}{\langle \eta_0^H, \mathbf{n} \rangle} \langle \eta_t^H, \mathbf{s}(f_t^I - f_t) \rangle dt. \end{aligned} \quad (3.17)$$

Start by noticing that  $\psi_T^H(\zeta^{\varphi(K)}) = M_T^{H|\varphi(K)}(f)$ , as  $K_H S_H(t) = K_W S_W(t)$ . Using Jensen's inequality, it follows from Equation (3.14) that

$$\mathbb{E}[|\psi_T^H(\zeta^{\varphi(K)})|^2] \leq \mathbb{E}[|\psi_T^H(\zeta^{\varphi(K)})|^2] = \mathbb{E}[\langle M^{H|K}(f) \rangle_t] \leq \frac{1}{K} 2n_{\max} C \|f\|_\infty^2 T \xrightarrow{K \rightarrow \infty} 0.$$

Suppose that  $(\psi_T^H(\zeta^{\varphi(K)}))_{K \geq 1}$  converges in law to  $\psi_T^H(\zeta)$ . According to Theorem 3.5 of Billingsley, 1999, it then is enough to prove that  $(\psi_T^H(\zeta^{\varphi(K)}))_{K \geq 1}$  is uniformly integrable to obtain that its expectation converges to the expectation of  $\psi_T^H(\zeta)$ . In our case, uniform integrability is easily assured as the sequence  $(\psi_T^H(\zeta^{\varphi(K)}))_{K \geq 1}$  is bounded. Indeed, using the fact that for all  $T \geq 0$ ,  $\zeta_T^{\varphi(K)} \in \mathfrak{M}_1$ , we obtain from Equation (3.17) that

$$\left| \psi_T^H(\zeta^{\varphi(K)}) \right| \leq ((2 + C) \|f\|_\infty + \|\mathcal{A}f\|_\infty) T.$$

We may now conclude that

$$\mathbb{E}[|\psi_T^H(\zeta)|] = \lim_{K \rightarrow \infty} \mathbb{E}[|\psi_T^H(\zeta^{\varphi(K)})|] = 0,$$

which yields the desired result.

It thus suffices to show that  $(\psi_T^H(\zeta^{\varphi(K)}))_{K \geq 1}$  converges in law to  $\psi_T^H(\zeta)$ . According to Skorokhod's representation theorem, there exists a probability space  $\Omega$  on which one may define  $(\tilde{\zeta}^K)_{K \geq 1}$  and  $\tilde{\eta}$  equal in law to  $(\zeta^{\varphi(K)})_{K \geq 1}$  and  $\eta$ , respectively, such that  $(\tilde{\zeta}^K)_{K \geq 1}$  converges almost surely in  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), w))^2$  to  $\tilde{\eta}$  on  $\Omega$ . In particular, it holds that

$$\forall T \geq 0, \forall X \in \{H, W\}, \forall g \in \mathcal{C}_b(E), \quad \langle \tilde{\zeta}_T^{X|K}, g \rangle \xrightarrow{K \rightarrow \infty} \langle \tilde{\eta}_T^X, g \rangle \text{ almost surely.}$$

It follows immediately that for any  $t \in [0, T]$  and  $X \in \{H, W\}$ , almost surely,

$$\langle \langle \tilde{\zeta}_t^{X|K}, f_t \rangle, \langle \tilde{\zeta}_t^{X|K}, \mathcal{A}f_t \rangle \rangle \xrightarrow{K \rightarrow \infty} \langle \langle \tilde{\eta}_t^X, f_t \rangle, \langle \tilde{\eta}_t^X, \mathcal{A}f_t \rangle \rangle. \quad (3.18)$$

Since  $|\langle \tilde{\zeta}_t^{X|K}, \mathcal{A}f_t \rangle| \leq 2 \|Df\|_\infty$ , where  $Df$  designates the differential of  $f$ , dominated convergence ensures that

$$\int_0^T \langle \tilde{\zeta}_t^{X|K}, \mathcal{A}f_t \rangle dt \xrightarrow{K \rightarrow \infty} \int_0^T \langle \tilde{\eta}_t^X, \mathcal{A}f_t \rangle dt \text{ almost surely.} \quad (3.19)$$

In order to establish the desired convergence of the last three terms of  $\psi_T^H(\tilde{\zeta})$ , we will make use of the following proposition, whose proof is postponed. In this context, a  $d$ -dimensional rectangle is a set defined as the product of  $d$  intervals of  $\mathbb{R} \cup \{-\infty, +\infty\}$ .

**Proposition 3.4.9.** *For any  $n \in \llbracket 1, n_{\max} \rrbracket$ , for any  $s \in \llbracket 0, n \rrbracket$ , consider  $m = m(n, s) < \infty$ , a set  $(A_k^{n,s})_{k \leq m}$  of pairwise disjoint  $(n-s)$ -dimensional rectangles and a set  $(\varphi_k^{n,s})_{k \leq m}$  of functions belonging to  $C_b^1(\mathbb{R}^{n-s})$ . For any  $\tau \in \mathbb{R}^{n_{\max}}$ , let  $\tau_{1,n-s} = (\tau_1, \dots, \tau_{n-s})$ . Define the function  $\phi : E \rightarrow \mathbb{R}$  by*

$$\forall x = (n, s, \tau) \in E, \quad \phi(x) = \sum_{k=1}^{m(n,s)} \mathbf{1}_{A_k^{n,s}(\tau_{1,n-s})} \varphi_k^{n-s}(\tau_{1,n-s}).$$

Then for any  $X \in \{H, W\}$  and  $T \geq 0$ , it holds that

$$\langle \tilde{\zeta}_T^{X|K}, \phi \rangle \xrightarrow{K \rightarrow \infty} \langle \tilde{\eta}_T^X, \phi \rangle \text{ in } L^1.$$

Let us focus on the second-to-last term, representing infection events occurring within workplaces, as the other two can be treated similarly.

The application  $\phi(x) = \lambda_W s(x) \mathbf{i}(x)$  is of the form described in Proposition 3.4.9, hence for any  $t \in [0, T]$ ,  $\langle \tilde{\zeta}_t^{W|K}, \lambda_W \mathbf{s} \mathbf{i} \rangle$  converges in  $L^1$  to  $\langle \tilde{\eta}_t^W, \lambda_W \mathbf{s} \mathbf{i} \rangle$  as  $K$  tends to infinity. Also, notice that as  $f \in C_b^1(\mathbb{R}_+ \times E, \mathbb{R})$ , it follows that for any  $t \in [0, T]$ ,  $f_t^I \in C_b^1(E)$ . Thus Proposition 3.4.9 ensures that for any  $t \in [0, T]$ ,  $\langle \tilde{\zeta}_t^{H|K}, \mathbf{s}(f_t^I - f_t) \rangle$  converges in  $L^1$  to  $\langle \tilde{\eta}_t^H, \mathbf{s}(f_t^I - f_t) \rangle$  as  $K$  tends to infinity.

In particular, the following convergence holds in probability:

$$X_t^K := \langle \tilde{\zeta}_t^{W|K}, \lambda_W \mathbf{s} \mathbf{i} \rangle \langle \tilde{\zeta}_t^{H|K}, \mathbf{s}(f_t^I - f_t) \rangle \xrightarrow{K \rightarrow \infty} X_t := \langle \tilde{\eta}_t^W, \lambda_W \mathbf{s} \mathbf{i} \rangle \langle \tilde{\eta}_t^H, \mathbf{s}(f_t^I - f_t) \rangle.$$

Letting  $c = 2\lambda_W n_{\max}^2 \|f\|_\infty$  and  $D = \{(x, y) : |x| \leq cy^2\}$ , then  $(X_t^K, \langle \tilde{\zeta}_t^{W|K}, \mathbf{s} \rangle)$  and  $(X_t, \langle \tilde{\eta}_t^W, \mathbf{s} \rangle)$  belong almost surely to  $D$ , for any  $K \geq 1$ . As  $\langle \tilde{\zeta}_t^{W|K}, \mathbf{s} \rangle$  converges almost surely to  $\langle \tilde{\eta}_t^W, \mathbf{s} \rangle$ , and the application  $(x, y) \mapsto (x/y) \mathbf{1}_{\{y \neq 0\}}$  is continuous on  $D$ , we deduce the following convergence in probability:

$$Y_t^K := \frac{\langle \tilde{\zeta}_t^{W|K}, \lambda_W \mathbf{s} \mathbf{i} \rangle}{\langle \tilde{\zeta}_t^{W|K}, \mathbf{s} \rangle} \langle \tilde{\zeta}_t^{H|K}, \mathbf{s}(f_t^I - f_t) \rangle \xrightarrow{K \rightarrow \infty} Y_t := \frac{\langle \tilde{\eta}_t^W, \lambda_W \mathbf{s} \mathbf{i} \rangle}{\langle \tilde{\eta}_t^W, \mathbf{s} \rangle} \langle \tilde{\eta}_t^H, \mathbf{s}(f_t^I - f_t) \rangle.$$

In addition, for any  $K \geq 1$  and any  $t \in [0, T]$ ,  $|Y_t^K| \leq 2\lambda_W n_{\max} \|f\|_{\infty}$ . Thus using twice dominated convergence, we first obtain that the above convergence of  $(Y_t^K)_{K \geq 1}$  to  $Y_t$  also holds in  $L^1$ , and subsequently the following convergence holds in  $L^1$ :

$$\int_0^T \frac{\langle \tilde{\zeta}_t^{W|K}, \lambda_W \mathbf{si} \rangle}{\langle \tilde{\zeta}_t^{W|K}, \mathbf{s} \rangle} \langle \tilde{\zeta}_t^{H|K}, \mathbf{s}(f_t^{\mathcal{I}} - f_t) \rangle dt \xrightarrow{K \rightarrow \infty} \int_0^T \frac{\langle \tilde{\eta}_t^W, \lambda_W \mathbf{si} \rangle}{\langle \tilde{\eta}_t^W, \mathbf{s} \rangle} \langle \tilde{\eta}_t^H, \mathbf{s}(f_t^{\mathcal{I}} - f_t) \rangle dt. \quad (3.20)$$

Reasoning in a similar manner, one also obtains:

$$\begin{aligned} \int_0^T \langle \tilde{\zeta}_t^H, \lambda_H \mathbf{si}(f_t^{\mathcal{I}} - f_t) \rangle dt &\xrightarrow{K \rightarrow \infty} \int_0^T \langle \tilde{\eta}_t^H, \lambda_H \mathbf{si}(f_t^{\mathcal{I}} - f_t) \rangle dt \text{ in } L^1, \\ \int_0^T \frac{\langle \tilde{\zeta}_t^H, \mathbf{i} \rangle}{\langle \tilde{\zeta}_0^H, \mathbf{n} \rangle} \langle \tilde{\zeta}_t^H, \mathbf{s}(f_t^{\mathcal{I}} - f_t) \rangle dt &\xrightarrow{K \rightarrow \infty} \int_0^T \frac{\langle \tilde{\eta}_t^H, \mathbf{i} \rangle}{\langle \tilde{\eta}_0^H, \mathbf{n} \rangle} \langle \tilde{\eta}_t^H, \mathbf{s}(f_t^{\mathcal{I}} - f_t) \rangle dt \text{ in } L^1. \end{aligned} \quad (3.21)$$

Thus, Equations (3.18-3.21) imply that all the terms on the right hand side of the definition of  $\psi_T^H(\tilde{\zeta}^K)$  as stated in Equation (3.17) converge in probability, and thus their linear combination converges in probability to the linear combination of their limits. In other words,  $\psi_T^H(\tilde{\zeta}^K)$  converges in probability to  $\psi_T^H(\tilde{\eta})$ , which ensures as desired that  $\psi_T^H(\tilde{\zeta}^K)$  converges in law to  $\psi_T^H(\tilde{\eta})$ . This concludes the proof.  $\square$

In order to conclude, we only need to show that Proposition 3.4.9 holds.

*Proof of Proposition 3.4.9. Step 1.* Recall that a  $d$ -dimensional rectangle is a set  $A$  defined as the product of  $d$  intervals of  $\mathbb{R} \cup \{-\infty, +\infty\}$ . If all  $d$  intervals are included in  $\mathbb{R}$ , the rectangle will further be said finite.

Let  $X \in \{H, W\}$ ,  $n \in \llbracket 1, n_{\max} \rrbracket$  and  $s \in \llbracket 0, n \rrbracket$ . We start by showing that for any  $T \geq 0$  and any finite  $(n-s)$ -dimensional rectangle  $B$ ,

$$\langle \tilde{\zeta}_T^{X|K}, \mathbf{1}_{\{(n,s)\} \times B} \rangle \xrightarrow{K \rightarrow \infty} \langle \tilde{\eta}_T^X, \mathbf{1}_{\{(n,s)\} \times B} \rangle \text{ almost surely.} \quad (3.22)$$

As  $\tilde{\eta}^X \in \mathcal{C}(\mathbb{R}_+, (\mathcal{M}_F(E), w))$ , it follows that for any  $T$ ,  $\tilde{\zeta}_T^{X|K}$  converges almost surely to  $\tilde{\eta}_T^X$  in  $(\mathcal{M}_F(E), w)$ . Thus, in order to establish the desired result, it is sufficient to show that  $\mathbf{B} = \{(n, s)\} \times B$  is a  $\tilde{\eta}_T^X$ -continuity set, in which case the Portmanteau theorem allows to conclude.

For any set  $A$ , let  $\partial A$  be the boundary of  $A$ . Then  $\partial \mathbf{B} = \{(n, s)\} \times \partial B$ . As  $B$  is a  $(n-s)$ -dimensional rectangle, there exist  $a_i < b_i \in \mathbb{R}$  for  $i \in \llbracket 1, n-s \rrbracket$  such that  $B$  can be written as the product of intervals (potentially open, closed or half-open) delimited by  $a_i < b_i$ , for  $i \in \llbracket 1, n-s \rrbracket$ . Thus

$$\partial B = \bigcup_{i=1}^{n-s} \bigcup_{c \in \{a_i, b_i\}} \left( \prod_{j=1}^{i-1} [a_j, b_j] \times \{c\} \times \prod_{k=i+1}^{n-s} [a_k, b_k] \right).$$

Consider any  $i \in \llbracket 1, n-s \rrbracket$  and  $c \in \mathbb{R}$ . We are going to prove that

$$\langle \tilde{\eta}_T^X, \mathbf{1}_{\{(n,s)\} \times (\prod_{j=1}^{i-1} [a_j, b_j] \times \{c\} \times \prod_{k=i+1}^{n-s} [a_k, b_k])} \rangle = 0, \quad (3.23)$$

which will be enough to conclude. In order to achieve this, let us introduce a mollifier  $\psi \in C^\infty(\mathbb{R})$  in the same sense as in the proof of Lemma 3.4.1, with compact support in  $[-1, 1]$ . For  $\varepsilon > 0$ , define the function  $\varphi_\varepsilon : x \mapsto \varepsilon^{-1}\psi(x/\varepsilon)$ , whose support lies in  $[-\varepsilon, \varepsilon]$  and which converges to  $\delta_0$  in the sense of distributions, when  $\varepsilon$  goes to zero. For any  $x = (n, s, \tau) \in E$ , let

$$\phi_\varepsilon(x) = \mathbf{1}_{\{\mathbf{n}(x)=n, \mathbf{s}(x)=s\}} \left( \prod_{\substack{j=1 \\ j \neq i}}^{n-s} \mathbf{1}_{[a_j, b_j]} * \varphi_\varepsilon(\tau_j) \right) \mathbf{1}_c * \varphi_\varepsilon(\tau_i).$$

As  $\phi_\varepsilon \in C_b^1(E)$  and  $\tilde{\zeta}_T^{X|K}$  converges almost surely to  $\tilde{\eta}_T^X$  in  $(\mathcal{M}_F(E), w)$ , dominated convergence implies that

$$\mathbb{E}[\langle \tilde{\eta}_T^X, \phi_\varepsilon \rangle] = \lim_{K \rightarrow \infty} \mathbb{E}[\langle \tilde{\zeta}_T^{X|K}, \phi_\varepsilon \rangle].$$

Notice that

$$\phi_\varepsilon(x) \leq \mathbf{1}_{\{\mathbf{n}(x)-\mathbf{s}(x) > i\}} \mathbf{1}_c * \varphi_\varepsilon(\tau_i(x)).$$

Hence proceeding as in the proof of Proposition 3.4.7, it follows that

$$\begin{aligned} \mathbb{E}[\langle \tilde{\zeta}_T^{X|K}, \phi_\varepsilon \rangle] &\leq \mathbb{E} \left[ \frac{1}{K_X} \sum_{k=1}^{K_X} \mathbf{1}_{\{n_k^X - s_k^X(0) \geq i, |(\tau_{k,i}^X(0) - T) - c| \leq \varepsilon\}} \right] \\ &\quad + \frac{C}{2} \int_0^T \nu([c + (T-t) - \varepsilon, c + (T-t) + \varepsilon]) dt. \end{aligned}$$

Absolute continuity of  $\nu$  with regard to the Lebesgue measure and Assumption 3.3.1 ensure that the right hand side is dominated by a function  $c(\varepsilon)$  which does not depend on  $K$ , and which goes to zero with  $\varepsilon$ . Thus  $\mathbb{E}[\langle \tilde{\eta}_T^X, \phi_\varepsilon \rangle] \leq c(\varepsilon)$ . In particular, one may construct a sequence  $(\varepsilon_n)_{n \geq 1}$  which converges to 0 and satisfies  $\sum_{n \geq 0} c(\varepsilon_n) < \infty$ . Then on the one hand, the Borel-Cantelli lemma ensures that  $\langle \tilde{\eta}_T^X, \phi_\varepsilon \rangle$  converges almost surely to 0 as  $n$  tends to infinity. On the other hand, by dominated convergence,  $\langle \tilde{\eta}_T^X, \phi_\varepsilon \rangle$  converges almost surely to the left-hand side of Equation (3.23) as  $\varepsilon$  tends to zero, hence Equation (3.23) is proven to be true.

As a consequence, we conclude that  $\langle \tilde{\eta}_T^X, \partial \mathbf{B} \rangle = 0$ , and thus Equation (3.22) holds.

**Step 2.** Consider now a function  $\phi$  as described in the proposition. For any  $N$ , let us introduce a partition of  $\mathbb{R}^{n-s}$  whose elements consist in (partially open) hypercubes of side length  $2^{-N}$ . For every  $k \leq m(n, s)$ , a partition  $(B_{k,j}^N)_{j \geq 1}$  of  $A_k^{n,s}$  is obtained by taking the intersection of  $A_k^{n,s}$  with those hypercubes. As  $A_k^{n,s}$  is a rectangle itself, the family  $(B_{k,j}^N)_{j \geq 1}$  consists of rectangles of side length at most  $2^{-N}$ . For every  $j$ , consider a point  $z_{k,j}^N$  belonging to  $B_{k,j}^N$ . Finally, define the set  $J_N(k) = \{j \geq 0 : \sup\{\|x\|_\infty : x \in B_{k,j}^N\} \leq N\}$ , which contains only a finite number of elements. Then we can define the following approximation of  $\phi$ :

$$\forall x \in E, \quad \phi_N(x) = \sum_{\substack{1 \leq n \leq n_{\max} \\ 0 \leq s \leq n}} \mathbf{1}_{\{\mathbf{n}(x)=n, \mathbf{s}(x)=s\}} \sum_{k=1}^{m(n,s)} \sum_{j \in J_N(k)} \varphi_k^{n-s}(z_{k,j}^N) \mathbf{1}_{B_{k,j}^N}(\tau_{1, n-s}(x)).$$

Using our result from the first step, for every  $(n, s)$  such that  $1 \leq n \leq n_{\max}$  and  $0 \leq s \leq n$ , for every  $k \leq m(n, s)$  and  $j \in J_N(k)$ , we obtain that

$$\lim_{K \rightarrow \infty} \langle \tilde{\zeta}_T^{X|K}, \phi_N \rangle = \langle \tilde{\eta}_T^X, \phi_N \rangle \text{ almost surely.} \quad (3.24)$$

Notice that for any  $x \in E$  such that  $\|\tau(x)\|_\infty > N$ ,  $\phi_N(x) = 0$ . Hence for any  $x = (n, s, \tau)$ , one obtains the following inequality:

$$|\phi_N(x) - \phi(x)| \leq |\phi(x)| \mathbf{1}_{\{\|\tau\|_\infty > N\}} + \sum_{k=1}^{m(n,s)} \sum_{j \in J_N(k)} |\varphi_k^{n-s}(z_{k,j}^N) - \varphi_k^{n-s}(\tau_{1,n-s})| \mathbf{1}_{B_{k,j}^N}(\tau_{1,n-s}). \quad (3.25)$$

Notice that there exists at most one  $(k, j) \in \llbracket 1, m(n, s) \rrbracket \times J_N(k)$  such that  $\tau_{1,n-s} \in B_{k,j}^N$ . As  $\varphi_k^{n,s} \in \mathcal{C}_b^1(\mathbb{R}^{n-s})$ , the mean value inequality further implies that

$$\forall k \leq m(n, s), \forall j \in J_N(k), \forall z \in B_{k,j}^N, |\varphi_k^{n-s}(z_{k,j}^N) - \varphi_k^{n-s}(z)| \leq \|D\varphi_k^{n,s}\|_\infty d_N,$$

where  $d_N$  denotes the maximum of the diameters of  $d$ -dimensional hypercubes of side length  $2^{-N}$ , for  $d \leq n_{\max}$ . Letting  $M = \max_{n,s,k} \|D\varphi_k^{n,s}\|_\infty$ , it follows that :

$$\forall x \in E, |\phi_N(x) - \phi(x)| \leq \|\phi\|_\infty \mathbf{1}_{\{\|\tau\|_\infty > N\}} + M d_N \quad (3.26)$$

Hence  $\phi_N$  converges point-wise to  $\phi$ . Thus, by dominated convergence,

$$\langle \tilde{\eta}_T^X, \phi_N \rangle \xrightarrow{N \rightarrow \infty} \langle \tilde{\eta}_T^X, \phi \rangle. \quad (3.27)$$

Furthermore, it follows from Equation (3.26) that for any  $K \geq 1$ ,

$$\mathbb{E}[|\langle \tilde{\zeta}_T^{X|K}, \phi_N \rangle - \langle \tilde{\zeta}_T^{X|K}, \phi \rangle|] \leq \|\phi\|_\infty \sup_{K \geq 1} \mathbb{E}[\langle \tilde{\zeta}_T^{X|K}, \mathbf{1}_{\{\|\tau(\cdot)\|_\infty > N\}} \rangle] + M d_N.$$

Reasoning as in the proof of Proposition 3.4.7, and using Assumption 3.3.1, we obtain that

$$\lim_{N \rightarrow \infty} \sup_{K \geq 1} \mathbb{E}[\langle \tilde{\zeta}_T^{X|K}, \mathbf{1}_{\{\|\tau(\cdot)\|_\infty > N\}} \rangle] = 0,$$

and as a consequence,

$$\lim_{N \rightarrow \infty} \sup_{K \geq 1} \mathbb{E}[|\langle \tilde{\zeta}_T^{X|K}, \phi_N \rangle - \langle \tilde{\zeta}_T^{X|K}, \phi \rangle|] = 0. \quad (3.28)$$

Noticing that

$$\mathbb{E}[|\langle \tilde{\zeta}_T^{X|K}, \phi \rangle - \langle \tilde{\eta}_T^X, \phi \rangle|] \leq \mathbb{E}[|\langle \tilde{\zeta}_T^{X|K}, \phi - \phi_N \rangle|] + \mathbb{E}[|\langle \tilde{\zeta}_T^{X|K}, \phi_N \rangle - \langle \tilde{\eta}_T^X, \phi_N \rangle|] + |\langle \tilde{\eta}_T^X, \phi_N - \phi \rangle|$$

together with Equations (3.28), (3.24) and (3.27) finally yields that

$$\lim_{K \rightarrow \infty} \mathbb{E}[|\langle \tilde{\zeta}_T^{X|K}, \phi \rangle - \langle \tilde{\eta}_T^X, \phi \rangle|] = 0.$$

This concludes the proof.  $\square$

### Proof of Theorem 3.3.2

The previous results are sufficient to establish Theorem 3.3.2. Indeed, it follows from Propositions 3.4.7 and 3.4.8 that from every subsequence of  $(\zeta^K)_{K \geq 1}$ , one may extract a subsubsequence converging in  $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(E), w))^2$  to a solution of Equation (3.6) which is continuous with respect to the total variation norm. As by assumption,  $\zeta_0^K$  converges in law to  $\eta_0 \in \mathfrak{M}_1$ , Proposition 3.4.2 implies that all of these subsubsequences converge to the same limit  $\eta$ , which is the unique solution of Equation (3.6) with initial condition  $\eta_0$ . As  $\eta_0 \in \mathfrak{M}_1$ , Proposition 3.4.2 further ensures that  $\eta \in \mathbb{D}(\mathbb{R}_+, \mathfrak{M}_1)$ . This establishes the convergence of  $(\zeta^K)_{K \geq 1}$  to  $\eta$  in  $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_1(E))^2$ .

### 3.4.2 Proof of Theorem 3.3.3

This section is devoted to extracting dynamical system (3.8a–c) from the measure-valued integral equation (3.6), under the assumption that  $\nu$  is the exponential distribution of parameter  $\gamma$ .

#### Preliminary study of the dynamical system

Before establishing Theorem 3.3.3 itself, let us start by showing that dynamical system (3.8a–c) endowed with initial condition (3.9) admits at most a unique solution. Existence will follow from the proofs of the forthcoming subsections, since they construct a solution to the dynamical system.

For this section, let us rewrite dynamical system (3.8a–c) as follows, in order to emphasize the associated Cauchy problem. Recall that the dynamical system is of dimension  $d = 2 + 2\#\mathbb{S} = n_{\max}(n_{\max} + 1)$ .

Let  $y \in \mathcal{C}^1(\mathbb{R}_+, \mathbb{R}^d)$  and  $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$  be defined such that dynamical system (3.8a–c) amounts to

$$y'(t) = f(y(t)) \quad \forall t \geq 0. \quad (3.29)$$

The components of  $y$  (and resp.  $f$ ) will be called  $s, i$  and  $n_{S,I}^X$  (resp.  $f_s, f_i$  and  $f_{X,S,I}$ ) for  $X \in \{H, W\}$  and  $(S, I) \in \mathbb{S}$ , in order to simplify their identification with the unknowns of the corresponding dynamical system. More precisely, consider the applications

$$\tau_X(y) = -\frac{\lambda_X}{m_X} \sum_{(S,I) \in \mathbb{S}} SI n_{S,I}^X \text{ for } X \in \{H, W\}, \text{ and } \tau_G(y) = \beta Gi.$$

Then  $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is defined as follows, for any  $y = (s, i, n_{S,I}^X : X \in \{H, W\}, (S, I) \in \mathbb{S}) \in \mathbb{R}^d$ :

$$f_s(y) = -(\tau_H(y) + \tau_W(y) + \tau_G(y)s) \text{ and } f_i(y) = -f_s(y) - \gamma i,$$

while for all  $X \in \{H, W\}$  and  $(S, I) \in \mathbb{S}$

$$\begin{aligned} f_{X,S,I}(y) = & - \left[ \left( \lambda_X I + \frac{\tau_X(y)}{s} + \tau_G(y) \right) S - \gamma I \right] n_{S,I}^X + \gamma (I+1) n_{S,I+1}^X \mathbf{1}_{\{S+I < n_{\max}\}} \\ & + \left( \lambda_X (I-1) + \frac{\tau_X(y)}{s} + \tau_G(y) \right) (S+1) n_{S+1,I-1}^X \mathbf{1}_{\{I \geq 1\}}. \end{aligned}$$

Also, notice that there are some natural constraints that we expect the solution of dynamical system (3.8a–c) to satisfy. Clearly,  $s, i$  and  $n_{S,I}^X$  should belong to  $[0, 1]$ . Also, as the population is partitioned into susceptible, infected and removed individuals, it follows that  $s+i \leq 1$ . Similarly, as all individuals belong to exactly one household and one workplace, and as  $n_{S,I}^X$  corresponds to the proportion of structures of type  $X$  which contain  $S$  susceptible and  $I$  infected individuals, we expect that for  $X \in \{H, W\}$ ,

$$\sum_{(S,I) \in \mathbb{S}} n_{S,I}^X \leq 1, \text{ and } \sum_{(S,I) \in \mathbb{S}} S n_{S,I}^X \leq m_X s. \quad (3.30)$$

We thus define the following set  $V \subset \mathbb{R}^d$ , which formalizes these constraints:

$$V = \left\{ y \in [0, 1]^d : s + i \leq 1, \sum_{(S,I) \in \mathbb{S}} n_{S,I}^X \leq 1 \text{ and } m_X s - \sum_{(S,I) \in \mathbb{S}} S n_{S,I}^X \geq 0 \quad \forall X \in \{H, W\} \right\}.$$

**Proposition 3.4.10.** *Let  $y^* \in V$ . Then the following assertions hold:*

- (i) *Suppose that there exists a solution  $y$  of the Cauchy problem (3.29) with initial condition  $y(0) = y^*$ . Then  $y(t) \in V$  for any  $t \geq 0$  for which  $y$  is well defined.*
- (ii) *For any  $T \geq 0$ , this problem admits at most a unique solution  $y$  on  $[0, T]$ .*
- (iii) *In particular, for any  $\varepsilon > 0$ , the dynamical system (3.8a–c) endowed with initial condition (3.9) admits at most a unique solution.*

The proof of this proposition is available in Appendix 3.D. It relies on establishing the Lipschitz continuity of  $f$  on  $V$ , from which uniqueness is deduced using Gronwall's lemma.

### Some properties of the limiting measure $\eta$

The results of this section focus on the limiting measure  $\eta$  given by Theorem 3.3.2, and will be useful for establishing Theorem 3.3.3.

Let us introduce the following notations. For  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$  and  $T \geq t \geq 0$ , define

$$f_{T,t} : x \in E \mapsto f(T, \Psi(x, T, t)) \text{ and } f_{T,t}^{\mathcal{I}} : x \in E \mapsto \langle \nu, f_{T,t}(j(x, \cdot)) \rangle. \quad (3.31)$$

We further define, for  $X \in \{H, W\}$  and  $t \geq 0$ , the following quantity which relates to the infectious pressure exerted on susceptibles outside of their structure of type  $X$ :

$$\Lambda_X(t) = \frac{\lambda_{\overline{X}}}{s_{\overline{X}}(t)} \langle \eta_t^{\overline{X}}, \mathbf{si} \rangle + \beta_G \frac{i_H(t)}{n_H}.$$

We can now state a result which is similar in spirit to Proposition 3.4.4. Notice that it holds under the same Assumptions as Theorem 3.3.2, and is not restricted to the Markovian case.

**Proposition 3.4.11.** *Let  $\eta$  be the unique solution in  $\mathcal{C}(\mathbb{R}_+, \mathfrak{M}_1)$  of Equation (3.6). Then for any  $T \geq 0$  and  $t \in [0, T]$ , for any measurable bounded function  $f : \mathbb{R}_+ \times E \rightarrow \mathbb{R}$ , it holds for  $X \in \{H, W\}$  that*

$$\langle \eta_t^X, f_{T,t} \rangle = \langle \eta_0^X, f_{T,0} \rangle + \lambda_X \int_0^t \langle \eta_u^X, \mathbf{si}(f_{T,u}^{\mathcal{I}} - f_{T,u}) \rangle dt + \int_0^t \Lambda_X(u) \langle \eta_u^X, \mathbf{s}(f_{T,u}^{\mathcal{I}} - f_{T,u}) \rangle dt. \quad (3.32)$$

*In particular, the application  $t \mapsto \langle \eta_t^X, f_{T,t} \rangle$  is continuous.*

*Proof.* Start by noticing that for any  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E, \mathbb{R})$ , for  $X \in \{H, W\}$ ,

$$\langle \eta_t^X, f_t \rangle = \langle \eta_0^X, f_{t,0} \rangle + \lambda_X \int_0^t \langle \eta_u^X, \mathbf{si}(f_{t,u}^{\mathcal{I}} - f_{t,u}) \rangle dt + \int_0^t \Lambda_X(u) \langle \eta_u^X, \mathbf{s}(f_{t,u}^{\mathcal{I}} - f_{t,u}) \rangle du. \quad (3.33)$$

Indeed, the proof of Equation (3.33) follows the exact same lines as the proof of Proposition 3.4.4, showing that for any  $f \in \mathcal{C}_b^1(\mathbb{R}_+ \times E)$ , Equation (3.33) leads to Equation (3.6) using Lemma 3.4.3.

Consider now a measurable bounded function  $f : \mathbb{R}_+ \times E \rightarrow \mathbb{R}$ . Proceeding as in the proof of Lemma 3.4.1, we consider a mollifier  $\psi$  on  $\mathbb{R}^{1+n_{\max}}$  and let  $\psi_k(t, \tau) = k^{n_{\max}+1} \psi(kt, k\tau)$  for any  $(t, \tau) \in \mathbb{R} \times \mathbb{R}^{n_{\max}}$ . Letting  $f_k(t, (n, s, \tau)) = f(\cdot, (n, s, \cdot)) * \psi_k(t, \tau)$ , we obtain by convolution a sequence of smooth functions  $(f_k)_{k \geq 1}$  which converges point-wise to  $f$ .



Then on the one hand, for any  $t \in [0, T]$  and  $x \in E$ , as  $k$  tends to infinity,  $(f_k)_{T,t}(x)$  converges to  $f_{T,t}(x)$ , and further  $(f_k)_{T,t}^{\mathcal{I}}(x)$  converges to  $f_{T,t}^{\mathcal{I}}(x)$  by dominated convergence. Define  $g_k \in C_b^1(E, \mathbb{R})$  by  $g_k(t, x) = (f_k)_{T,t}(x)$ , it then holds that  $(g_k)_{t,u} = (f_k)_{T,u}$  and  $(g_k)_{t,u}^{\mathcal{I}} = (f_k)_{T,u}^{\mathcal{I}}$ . Thus applying Equation (3.33) to  $g_k$  and using dominated convergence as  $k$  goes to infinity yields the desired result.

Finally, the continuity of  $t \mapsto \langle \eta_t^X, f_{T,t} \rangle$  on  $[0, T]$  is a consequence of Equation (3.32), as the integrands of the right-hand-side are bounded.  $\square$

Proposition 3.4.11 allows us to establish the following result under the assumption that  $\nu$  is the exponential distribution. In particular, it implies that within each structure, at any time, the remaining infectious periods of currently infectious individuals are independent and identically distributed, of common law  $\nu$ .

**Proposition 3.4.12.** *Assume that  $\nu$  is the exponential law of parameter  $\gamma$ , and consider  $\eta$  as defined in Theorem 3.3.2 with initial condition  $\eta_0 = \eta_{0,\varepsilon}$ . Let  $(X_n)_{n \geq 0}$  be a sequence of independent identically distributed random variables of common law  $\nu$ . Let  $T \geq 0$ ,  $n \in \llbracket 1, n_{\max} \rrbracket$  and  $s \in \llbracket 0, n-1 \rrbracket$ . For any  $m \in \llbracket 0, n-s \rrbracket$ , any functions  $f \in \mathcal{B}_b(\mathbb{R}_+ \times \mathbb{R}^m)$ ,  $g_1, \dots, g_{n-s-m} \in \mathcal{B}_b(\mathbb{R}_+ \times \mathbb{R})$  and any  $j_1 < \dots < j_m$  and  $k_1 < \dots < k_{n-s-m}$  such that  $\{k_1, \dots, k_{n-s-m}\} \cup \{j_1, \dots, j_m\} = \llbracket 1, n-s \rrbracket$ , define*

$$F(t, x) = \mathbf{1}_{\left\{ \begin{array}{l} \mathbf{n}(x)=n, \mathbf{s}(x)=s, \\ \tau_{j_\ell}(x) > 0 \forall 1 \leq \ell \leq m \end{array} \right\}} \left( f(t, \tau_{j_1}(x), \dots, \tau_{j_m}(x)) - \mathbb{E}[f(t, X_1, \dots, X_m)] \right) \prod_{\ell=1}^{n-s-m} g_\ell(t, \tau_{k_\ell}(x)).$$

Then

$$\forall t \in [0, T], \quad \langle \eta_t^X, F_{T,t} \rangle = 0. \quad (3.34)$$

*Proof.* Let  $n \in \llbracket 1, n_{\max} \rrbracket$ . Consider any  $m, f, g, \{j_1, \dots, j_m\}$  and  $\{k_1, \dots, k_m\}$  satisfying the constraints given in the proposition, and define  $F$  as above. Throughout the proof, we let  $E_m(t) = \mathbb{E}[f(t, X_1, \dots, X_m)]$ .

Notice that  $F$  satisfies the assumptions of Proposition 3.4.11. Letting  $C = (\lambda_X n_{\max} + \lambda_{\bar{X}} n_{\max} + \beta_G) n_{\max}$ , it follows from Equation (3.32) that:

$$|\langle \eta_t^X, F_{T,t} \rangle| \leq |\langle \eta_0^X, F_{T,0} \rangle| + C \int_0^t (|\langle \eta_u^X, F_{T,u} \rangle| + |\langle \eta_u^X, F_{T,u}^{\mathcal{I}} \rangle|) du. \quad (3.35)$$

Let  $(Y_k)_{k \geq 1}$  be a sequence of independent identically distributed random variables of common law  $\nu$  which is independent from  $(X_k)_{k \geq 1}$ , and write  $p_{n,s}^X = \pi_n^X \binom{n}{s} (1-\varepsilon)^s \varepsilon^{n-s}$ . Then on the one hand, Equation (3.7) ensures:

$$\begin{aligned} \langle \eta_0^X, F_{T,0} \rangle &= p_{n,s}^X \left( \mathbb{E}[f(T, Y_{j_1} - T, \dots, Y_{j_m} - T) \mathbf{1}_{\{Y_{j_1} - T > 0, \dots, Y_{j_m} - T > 0\}}] \right. \\ &\quad \left. - \mathbb{E}[f(T, X_1, \dots, X_m)] \mathbb{E}[\mathbf{1}_{\{Y_{j_1} - T > 0, \dots, Y_{j_m} - T > 0\}}] \right) \prod_{\ell=1}^{n-s-m} \mathbb{E}[g_\ell(T, Y_{k_\ell} - T)]. \end{aligned}$$

Usual properties of the exponential distribution thus lead to  $\langle \eta_0^X, F_{T,0} \rangle = 0$ .

On the other hand, let us compute  $F_{T,u}^{\mathcal{I}}(x)$  for any  $x \in E$ . Distinguishing the cases  $n-s \in \mathbb{J}_m = \{j_1, \dots, j_m\}$  and  $n-s \notin \mathbb{J}_m$  leads to:

$$\begin{aligned} F_{T,u}^{\mathcal{I}}(x) &= \mathbf{1}_{\left\{ \begin{array}{l} \mathbf{n}(x)=n, \mathbf{s}(x)=s+1, \\ \tau_j(x) > T-u \forall j \in \mathbb{J}_m \setminus \{n-s\} \end{array} \right\}} \left( \mathbf{1}_{\{n-s \in \mathbb{J}_m\}} e^{-\gamma(T-u)} a_{T,u}(x) \right. \\ &\quad \left. + \mathbf{1}_{\{n-s \notin \mathbb{J}_m\}} \mathbb{E}[g_{n-s}(X_{n-s} - (T-u))] b_{T,u}(x) \right) \end{aligned}$$

where

$$a(t, x) = (\mathbb{E}[f(t, \tau_{j_1}(x), \dots, \tau_{j_{m-1}}(x), X_{n-s})] - E_m(t)) \prod_{j \in \llbracket 1, n-s \rrbracket \setminus \mathbb{J}_m} g_j(t, \tau_{k_j}(x)),$$

$$b(t, x) = (f(t, \tau_{j_1}(x), \dots, \tau_{j_m}(x)) - E_m(t)) \prod_{j \in \llbracket 1, n-s-1 \rrbracket \setminus \mathbb{J}_m} g_j(t, \tau_{k_j}(x)).$$

We are now ready to proceed by induction on  $s \in \llbracket 0, n-1 \rrbracket$ . First, consider the case  $s = n-1$ . Then either  $m = 0$  and the result is immediate, or  $m = 1$  and  $\mathbb{J}_1 = \{1\}$ . Hence necessarily  $n-s \in \mathbb{J}_1$  and  $a(x) = 0$  as  $\mathbb{E}[f(X_0 - (T-t))\mathbf{1}_{\{X_0 > T-t\}}] = \mathbb{E}[f(X_1)]$ . Thus  $F_{T,t}^T(x) = 0$  for any  $x \in E$ , and Equation (3.35) reduces to

$$|\langle \eta_T^X, F \rangle| = |\langle \eta_T^X, F_{T,T} \rangle| \leq C \int_0^T |\langle \eta_t^X, F_{T,t} \rangle| dt.$$

Recalling that  $t \mapsto \langle \eta_t^X, F_{T,t} \rangle$  is continuous on  $[0, T]$  according to Proposition 3.4.11, Gronwall's inequality ensures that  $|\langle \eta_t^X, F_{T,t} \rangle| = 0$  for every  $t \in [0, T]$ .

Suppose now that  $s < n-1$ , and that the result holds for  $s+1$ . It then follows from the induction hypothesis that for any  $u \in [0, T]$ ,

$$\langle \eta_u^X, \mathbf{1}_{\left\{ \begin{array}{l} \mathbf{n}(x)=n, \mathbf{s}(x)=s+1, \\ \tau_j(x) > T-u \ \forall j \in \mathbb{J}_m \setminus \{n-s\} \end{array} \right\}} a_{T,u}(\cdot) \rangle = \langle \eta_u^X, \mathbf{1}_{\left\{ \begin{array}{l} \mathbf{n}(x)=n, \mathbf{s}(x)=s+1, \\ \tau_j(x) > T-u \ \forall j \in \mathbb{J}_m \setminus \{n-s\} \end{array} \right\}} b_{T,u}(\cdot) \rangle = 0.$$

Hence  $\langle \eta_t^X, F_{T,u}^T \rangle = 0$  for any  $u \in [0, T]$ , and Gronwall's inequality allows to conclude as previously.  $\square$

### Proof of Theorem 3.3.3

From now on, we assume that  $\eta_0 = \eta_{0,\varepsilon}$  as given by (3.7), and that  $\nu$  is the exponential distribution with parameter  $\gamma$ . Throughout this section, let  $h$  be the Heaviside step function, i.e.  $h(z) = \mathbf{1}_{\{z > 0\}}$  for any real number  $z$ .

Let us establish the following proposition, which serves as a starting point of the proof of Theorem 3.3.3.

**Proposition 3.4.13.** *Under the assumptions of Theorem 3.3.3, it holds that for any  $X \in \{H, W\}$ ,  $(s_X(t))_{t \geq 0}$  and  $(i_X(t))_{t \geq 0}$  satisfy*

$$\begin{aligned} \frac{d}{dt} s_X(t) &= -n_X \left( \frac{\lambda_X}{n_X} \langle \eta_t^X, \mathbf{si} \rangle + \frac{\lambda_{\bar{X}}}{n_{\bar{X}}} \langle \eta_t^{\bar{X}}, \mathbf{si} \rangle + \beta_G \frac{i_H(t)}{n_H} \frac{s_X(t)}{n_X} \right), \\ \frac{d}{dt} i_X(t) &= -\frac{d}{dt} s_X(t) + \gamma i_X(t). \end{aligned} \quad (3.36)$$

Further,

$$s_X(0) = (1 - \varepsilon)n_X \text{ and } i_X(0) = \varepsilon n^X. \quad (3.37)$$

*Proof.* Notice that  $\mathbf{s} \in \mathcal{C}_b^1(E, \mathbb{R})$  is such that  $\mathcal{A}\mathbf{s} = 0$  and  $\mathbf{s}^T(x) - \mathbf{s}(x) = -1$  for all  $x \in E$ . It thus follows immediately from Equation (3.6) that, for any  $X \in \{H, W\}$ ,

$$s_X(T) = s_X(0) - \int_0^T \left( \lambda_X \langle \eta_t^X, \mathbf{si} \rangle + \lambda_{\bar{X}} \frac{s_X(t)}{s_{\bar{X}}(t)} \langle \eta_t^{\bar{X}}, \mathbf{si} \rangle + \beta_G \frac{i_H(t)}{n_H} s_X(t) \right) dt.$$

Further, since  $\eta \in \mathcal{C}(\mathbb{R}_+, (\mathfrak{M}_1, \|\cdot\|_{TV}))$ , it follows that for any  $t \geq 0$ ,  $(\zeta_t^K)_{K \geq 1}$  converges in law to  $\eta_t$ . As  $\mathbf{s}$  is continuous and bounded on  $E$ , this implies that  $\langle \zeta_t^{X|K}, \mathbf{s} \rangle \rightarrow \langle \eta_t^X, \mathbf{s} \rangle$  when  $K$  tends to infinity. The analogous result holds for  $\mathbf{n}$ . Hence Lemma 3.2.2 ensures that, for any  $t \geq 0$ ,  $s_X(t)/s_{\bar{X}}(t) = n^X/n_{\bar{X}}$ . In other words,

$$s_X(T) = s_X(0) - n_X \int_0^T \left( \frac{\lambda_X}{n_X} \langle \eta_t^X, \mathbf{si} \rangle + \frac{\lambda_{\bar{X}}}{n_{\bar{X}}} \langle \eta_t^{\bar{X}}, \mathbf{si} \rangle + \beta_G \frac{i_H(t)}{n_H} \frac{s_X(t)}{n_X} \right) dt.$$

As  $\eta \in \mathcal{C}(\mathbb{R}_+, (\mathfrak{M}_1, \|\cdot\|_{TV}))$ , and  $\mathbf{s}$  and  $\mathbf{i}$  are bounded measurable functions, it follows that the integrand is continuous with regard to  $t$ . Thus, the first line of Equation (3.36) comes from the fundamental theorem of calculus.

Recall that  $\eta_0 = \eta_{0,\varepsilon}$  as defined in Equation (3.7). In particular, we now have  $n_X = m_X$ , hence

$$s_X(0) = \langle \eta_{0,\varepsilon}^X, \mathbf{s} \rangle = \sum_{n=1}^{n_{\max}} \pi_n^X \sum_{s=0}^n s \binom{n}{s} (1-\varepsilon)^s \varepsilon^{n-s} = (1-\varepsilon) \sum_{n=1}^{n_{\max}} n \pi_n^X = (1-\varepsilon) n^X. \quad (3.38)$$

This yields the first part of Equation (3.37).

It remains to take an interest in  $i_X(t)$ . As  $\mathbf{i}$  does not belong to  $\mathcal{C}_b^1(E, \mathbb{R})$  we cannot proceed in the same way. Remember that  $\mathbf{i}(x) = \sum_{j=1}^{n_{\max}} h(\tau_j)$  for  $x = (n, s, \tau) \in E$ . Thus, we may apply Proposition 3.4.11 and obtain that

$$\begin{aligned} \langle \eta_T^X, \mathbf{i} \rangle &= \langle \eta_0^X, \mathbf{i}_{T,0} \rangle + \lambda_X \int_0^T \langle \eta_t^X, \mathbf{si}(\mathbf{i}_{T,t}^{\bar{I}} - \mathbf{i}_{T,t}) \rangle dt \\ &\quad + \lambda_{\bar{X}} \int_0^T \frac{1}{s_{\bar{X}}(t)} \langle \eta_t^{\bar{X}}, \mathbf{si} \rangle \langle \eta_t^X, \mathbf{s}(\mathbf{i}_{T,t}^{\bar{I}} - \mathbf{i}_{T,t}) \rangle dt + \beta_G \int_0^T \frac{i_H(t)}{n_H} \langle \eta_t^X, \mathbf{s}(\mathbf{i}_{T,t}^{\bar{I}} - \mathbf{i}_{T,t}) \rangle dt. \end{aligned} \quad (3.39)$$

Further, for any  $\sigma \geq 0$ ,  $T \geq t \geq 0$  and  $x \in E$ ,

$$i(\Psi(j(x, \sigma), T, t)) - i(\Psi(x, T, t)) = \mathbf{1}_{\{\sigma > (T-t)\}},$$

hence

$$\mathbf{i}_{T,t}^{\bar{I}}(x) - \mathbf{i}_{T,t}(x) = \nu([T-t, \infty)) = e^{-\gamma(T-t)}.$$

Injecting this into Equation (3.39) and using as before that  $s_X(t)/s_{\bar{X}}(t) = n_X/n_{\bar{X}}$  yields

$$i_X(T) = \langle \eta_0^X, \mathbf{i}_{T,0} \rangle + n_X e^{-\gamma T} \int_0^T e^{\gamma t} \left( \frac{\lambda_X}{n_X} \langle \eta_t^X, \mathbf{si} \rangle + \frac{\lambda_{\bar{X}}}{n_{\bar{X}}} \langle \eta_t^{\bar{X}}, \mathbf{si} \rangle + \beta_G \frac{i_H(t)}{n_H} \frac{s_X(t)}{n_X} \right) dt. \quad (3.40)$$

As  $\eta_0 = \eta_{0,\varepsilon}$ , we may compute the first term of the right-hand side of this equation and obtain that

$$\langle \eta_0^X, \mathbf{i}_{T,0} \rangle = \langle \eta_{0,\varepsilon}^X, \mathbf{i}_{T,0} \rangle = e^{-\gamma T} \sum_{n=1}^{n_{\max}} \pi_n^X \sum_{s=0}^n \binom{n}{s} (1-\varepsilon)^s \varepsilon^{n-s} (n-s) = e^{-\gamma T} \varepsilon n^X. \quad (3.41)$$

Using the continuity of the integrand in Equation (3.40), we may now differentiate it with regard to  $T$ :

$$\frac{d}{dT} i_X(T) = n_X \left( \frac{\lambda_X}{n_X} \langle \eta_T^X, \mathbf{si} \rangle + \frac{\lambda_{\bar{X}}}{n_{\bar{X}}} \langle \eta_T^{\bar{X}}, \mathbf{si} \rangle + \beta_G \frac{i_T^H}{n_H} \frac{s_X(T)}{n_X} \right) - \gamma i_X(T) = -\frac{d}{dT} s_X(T) - \gamma i_X(T).$$

We thus have recovered the second line of Equation (3.36). Finally, the second half of Equation (3.37) is obtained by a computation analogous to (3.38). This concludes the proof.  $\square$

For  $(S, I) \in \mathbb{S}$ , define the function  $f^{S,I} : E \rightarrow \{0, 1\}$  by

$$f^{S,I}(x) = \mathbf{1}_{\{\mathbf{s}(x)=S, \mathbf{i}(x)=I\}}.$$

For  $t \geq 0$ , let  $n_{S,I}^X(t) = \langle \eta_t^X, f^{S,I} \rangle$ , which defines a continuous function on  $\mathbb{R}_+$  as  $f^{S,I}$  is bounded and measurable and  $\eta \in \mathcal{C}(\mathbb{R}_+, (\mathfrak{M}_1, \|\cdot\|_{TV}))$ . In words, this corresponds to the proportion of structures of type  $X$  which contain exactly  $S$  susceptible and  $I$  infected individuals. Notice that

$$\{x \in E : \mathbf{s}(x)\mathbf{i}(x) > 0\} = \{x \in E : \mathbf{s}(x)\mathbf{i}(x) > 0, (\mathbf{s}(x), \mathbf{i}(x)) \in \mathbb{S}\}.$$

We may thus rewrite the first line of Equation (3.36) as follows:

$$\frac{d}{dt}s_X(t) = -n_X \left( \frac{\lambda_X}{n_X} \sum_{(S,I) \in \mathbb{S}} SIn_{S,I}^X(t) + \frac{\lambda_{\bar{X}}}{n_{\bar{X}}} \sum_{(S,I) \in \mathbb{S}} SIn_{(S,I)}^{\bar{X}}(t) + \beta_G \frac{i_H(t)}{n_H} \frac{s_X(t)}{n_X} \right).$$

Similarly, it holds that

$$\Lambda_X(t) = \frac{\lambda_{\bar{X}}}{s_{\bar{X}}(t)} \langle \eta_t^{\bar{X}}, \mathbf{si} \rangle + \beta_G \frac{i_H(t)}{n_H} = \frac{\lambda_{\bar{X}}}{s_{\bar{X}}(t)} \sum_{(S,I) \in \mathbb{S}} SIn_{(S,I)}^{\bar{X}}(t) + \beta_G \frac{i_H(t)}{n_H},$$

which may also be written in terms of the notations of Equation (3.8a–c):

$$\Lambda_X(t) = \left( \frac{s_{\bar{X}}(t)}{n_X} \right)^{-1} \tau_{\bar{X}}(t) + \beta_G \frac{i_H(t)}{n_H}.$$

This motivates a closer study of the functions  $n_{S,I}^X$  for  $(S, I) \in \mathbb{S}$ .

**Proposition 3.4.14.** *Let  $X \in \{H, W\}$  and  $(S, I) \in \mathbb{S}$ . Under the assumptions of Theorem 3.3.3, it holds that*

$$\begin{aligned} \frac{d}{dt}n_{S,I}^X(t) &= \gamma \left( (I+1)n_{S,I+1}^X(t)\mathbf{1}_{\{S+I < n_{\max}\}} - In_{S,I}^X(t) \right) \\ &+ \lambda_X \left( (S+1)(I-1)n_{S+1,I-1}^X(t)\mathbf{1}_{\{I \geq 1\}} - SIn_{S,I}^X(t) \right) \\ &+ \Lambda_X(t) \left( (S+1)n_{S+1,I-1}^X(t)\mathbf{1}_{\{I \geq 1\}} - Sn_{S,I}^X(t) \right). \end{aligned} \quad (3.42)$$

Further

$$n_{S,I}^X(0) = \binom{S+I}{I} \pi_{S+I}^X (1-\varepsilon)^S \varepsilon^I. \quad (3.43)$$

*Proof.* Let us start by establishing the initial condition of Equation (3.43). Let  $(S, I) \in \mathbb{S}$ . We will make use of another expression of  $f^{S,I}$ , which will actually be of use throughout the proof. For two integers  $j \leq n$ , let  $\mathbb{B}(n, j)$  be the set of unordered subsets of  $j$  elements chosen in  $\llbracket 1, n \rrbracket$ . It then holds that for any  $x \in E$ ,

$$f^{S,I}(x) = \sum_{n=S+I}^{n_{\max}} \mathbf{1}_{\left\{ \begin{array}{l} \mathbf{n}(x)=n \\ \mathbf{s}(x)=S \end{array} \right\}} \sum_{\mathbf{v} \in \mathbb{B}(n-S, I)} H^{n, S, \mathbf{v}}(\tau(x)),$$

where, for  $\mathbf{v} \in \mathbb{B}(n - S, I)$  and any  $\tau \in \mathbb{R}^{n_{\max}}$ ,

$$H^{n,S,\mathbf{v}}(\tau) = \prod_{j \in \mathbf{v}} h(\tau_j) \prod_{j \in \llbracket 1, n-S \rrbracket \setminus \mathbf{v}} (1 - h(\tau_j)).$$

The idea behind this expression is that a structure of type  $(n, s, \tau)$  contains  $S$  susceptible and  $I$  infected members if and only if  $s = S$ , and further exactly  $I$  out of the  $n - S$  first components of  $\tau$  are positive. In other words, there exists at most one element  $\mathbf{v} \in \mathbb{B}(n - S, I)$  for which the term in the sum is not equal to zero, in which case the set  $\mathbf{v}$  corresponds to the indexes of infectious members, while  $\llbracket 1, n - S \rrbracket \setminus \mathbf{v}$  is the set of removed members.

Let  $\varepsilon > 0$ . Throughout the following, for  $X \in \{H, W\}$ ,  $n \in \llbracket 1, n_{\max} \rrbracket$  and  $s \in \llbracket 0, n \rrbracket$ , let  $p_{n,s}^X = \pi_n^X \binom{n}{s} (1 - \varepsilon)^s \varepsilon^{n-s}$ . Notice that by definition of  $\eta_{0,\varepsilon}$ ,

$$n_{S,I}^X(0) = \sum_{n=S+I}^{n_{\max}} p_{n,S}^X \sum_{\mathbf{v} \in \mathbb{B}(n-S,I)} \langle \nu, h \rangle^{\#\mathbf{v}} \langle \nu, 1 - h \rangle^{(n-S) - \#\mathbf{v}}.$$

Whenever  $(n - S) - \#\mathbf{v} > 0$ , the term vanishes as  $\langle \nu, 1 - h \rangle = 0$ . Hence only the case  $\mathbf{v} = \llbracket 1, n - S \rrbracket$  remains, which in turn corresponds to  $n = S + I$  and leads to Equation (3.43).

Next, let us apply Proposition 3.4.11 to  $f^{S,I}$ . A brief computation, based on distinguishing the cases where  $n - s \in \mathbf{v}$  or  $n - s \notin \mathbf{v}$  for any  $\mathbf{v} \in \mathbb{B}(n - S, I)$ , yields

$$(f^{S,I})_{T,t}^{\mathcal{I}} = \mathbf{1}_{\{I \geq 1\}} e^{-\gamma(T-t)} f_{T,t}^{S+1,I-1} + \mathbf{1}_{\{S+1 < n_{\max}\}} (1 - e^{-\gamma(T-t)}) f_{T,t}^{S+1,I}.$$

As a consequence, Proposition 3.4.11 ensures that

$$\begin{aligned} n_{S,I}^X(T) &= \langle \eta_0^X, f_{T,0}^{S,I} \rangle - \int_0^T \left( \lambda_X \langle \eta_t^X, \mathbf{si} f_{T,t}^{S,I} \rangle + \Lambda_X(t) \langle \eta_t^X, \mathbf{s} f_{T,t}^{S,I} \rangle \right) dt \\ &+ \mathbf{1}_{\{I \geq 1\}} \int_0^T e^{-\gamma(T-t)} \left( \lambda_X \langle \eta_t^X, \mathbf{si} f_{T,t}^{S+1,I-1} \rangle + \Lambda_X(t) \langle \eta_t^X, \mathbf{s} f_{T,t}^{S+1,I-1} \rangle \right) dt \\ &+ \mathbf{1}_{\{S+1 < n_{\max}\}} \int_0^T (1 - e^{-\gamma(T-t)}) \left( \lambda_X \langle \eta_t^X, \mathbf{si} f_{T,t}^{S+1,I} \rangle + \Lambda_X(t) \langle \eta_t^X, \mathbf{s} f_{T,t}^{S+1,I} \rangle \right) dt. \end{aligned} \quad (3.44)$$

We will thus focus on differentiating with respect to  $T$  the different expressions composing the right-hand-side.

Let us start with the term  $\langle \eta_0^X, f_{T,0}^{S,I} \rangle$ . By definition of  $f_{T,0}^{S,I}$  and  $\eta_{0,\varepsilon}^X$ , it holds that

$$\langle \eta_0^X, f_{T,0}^{S,I} \rangle = \sum_{n=S+I}^{n_{\max}} p_{n,S}^X \binom{I}{n-S} e^{-\gamma IT} (1 - e^{-\gamma T})^{n-S-I}.$$

Hence, using the Equality  $(n - S - I) \binom{I}{n-S} = (I + 1) \binom{n-S}{I}$ , we obtain that

$$\frac{d}{dT} \langle \eta_0^X, f_{T,0}^{S,I} \rangle = -\gamma I \langle \eta_0^X, f_{T,0}^{S,I} \rangle + \gamma (I + 1) \sum_{n=S+I+1}^{n_{\max}} p_{n,S}^X \binom{n-S}{I} e^{-\gamma (I+1)T} (1 - e^{-\gamma T})^{n-S-I-1}$$

and we thus recognise that

$$\frac{d}{dT} \langle \eta_0^X, f_{T,0}^{S,I} \rangle = -\gamma I \langle \eta_0^X, f_{T,0}^{S,I} \rangle + \gamma (I + 1) \langle \eta_0^X, f_{T,0}^{S,I+1} \rangle. \quad (3.45)$$

Let us now focus on the remaining terms of the right-hand side of Equation (3.44). This motivates a closer study of  $f_{T,t}^{S,I}$  for any  $(S, I) \in \mathbb{S}$ . By definition, for  $x \in E$ ,

$$f_{T,t}^{S,I}(x) = \sum_{n=S+I}^{n_{\max}} \mathbf{1}_{\left\{ \begin{array}{l} \mathbf{n}(x)=n \\ \mathbf{s}(x)=S \end{array} \right\}} \sum_{\mathbf{v} \in \mathbb{B}(n-S, I)} \prod_{j \in \mathbf{v}} h(\tau_j(x) - (T-t)) \prod_{j \in \llbracket 1, n-S \rrbracket \setminus \mathbf{v}} (1 - h(\tau_j(x) - (T-t))).$$

In particular,  $f_{T,t}^{S,I}(x) > 0$  requires that at time  $t$ , in a structure of type  $x$ , there are  $J \geq I$  infected individuals, out of which exactly  $I$  must have a remaining infectious period exceeding  $(T-t)$ . Hence, consider  $(S, I)$  and  $T$  to be fixed and define, for any  $J \in \llbracket I, n_{\max} - S \rrbracket$  and  $x \in E$ ,

$$g_J(t, x) = \sum_{n=S+J}^{n_{\max}} \mathbf{1}_{\left\{ \begin{array}{l} \mathbf{n}(x)=n \\ \mathbf{s}(x)=S \end{array} \right\}} \sum_{\mathbf{v}_0 \in \mathbb{B}(n-S, J)} \sum_{\substack{\mathbf{v} \subseteq \mathbf{v}_0 \\ \#\mathbf{v}=I}} G^{n, \mathbf{v}_0, \mathbf{v}}(t, \tau(x)),$$

where, for  $t \geq 0$  and  $\tau \in \mathbb{R}^{n_{\max}}$ ,

$$G^{n, \mathbf{v}_0, \mathbf{v}}(t, \tau) = \prod_{j \in \mathbf{v}} h(\tau_j - (T-t)) \prod_{j \in \mathbf{v}_0 \setminus \mathbf{v}} \mathbf{1}_{\{0 < \tau_j \leq T-t\}} \prod_{j \in \llbracket 1, n-S \rrbracket \setminus \mathbf{v}_0} (1 - h(\tau_j)).$$

Dependence of  $g_J$  and  $G^{n, \mathbf{v}_0, \mathbf{v}}$  on  $(S, I)$  and  $T$  is omitted in these notations for readability. It then holds that

$$\forall x \in E, \quad f_{T,t}^{S,I}(x) = \sum_{J=I}^{n_{\max}-S} g_J(t, x).$$

Let  $J \in \llbracket I, n_{\max} - S \rrbracket$ ,  $\mathbf{v}_0 = \{j_1, \dots, j_J\} \in \mathbb{B}(n-s, J)$  and  $\mathbf{v} \subseteq \mathbf{v}_0$  such that  $\#\mathbf{v} = I$ . Proposition 3.4.12 applied to the functions  $f : \mathbb{R} \times \mathbb{R}^J \rightarrow \mathbb{R}$  and  $g_k : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  defined by

$$f(t, \tau_{j_1}, \dots, \tau_{j_J}) = \prod_{j \in \mathbf{v}} h(\tau_j - (T-t)) \prod_{j \in \mathbf{v}_0 \setminus \mathbf{v}} \mathbf{1}_{\{0 \leq \tau_j \leq T-t\}},$$

$$g_k(t, \tau) = 1 - h(\tau) \quad \forall k \in \llbracket 1, n-S-J \rrbracket,$$

leads to the following equality, for all  $t \in [0, T]$ :

$$\begin{aligned} \langle \eta_t^X, g_J(t, \cdot) \rangle &= \sum_{n=S+J}^{n_{\max}} \sum_{\substack{\mathbf{v}_0 \in \mathbb{B}(n-S, J) \\ \mathbf{v} \subseteq \mathbf{v}_0: \#\mathbf{v}=I}} e^{-\gamma(T-t)I} (1 - e^{-\gamma(T-t)})^{J-I} \langle \eta_t^X, \mathbf{1}_{\left\{ \begin{array}{l} \mathbf{n}(\cdot)=n \\ \mathbf{s}(\cdot)=S \end{array} \right\}} H^{n, S, \mathbf{v}_0}(\tau(\cdot)) \rangle \\ &= \binom{J}{I} e^{-\gamma(T-t)I} (1 - e^{-\gamma(T-t)})^{J-I} n_{S, J}^X(t). \end{aligned}$$

As a consequence, we obtain in particular that

$$\langle \eta_t^X, \mathbf{si}f_{T,t}^{S,I} \rangle = \sum_{J=I}^{n_{\max}-S} \binom{J}{I} e^{-\gamma(T-t)I} (1 - e^{-\gamma(T-t)})^{J-I} S J n_{S, J}^X(t).$$

Differentiating with respect to  $T$  yields

$$\partial_T \langle \eta_t^X, \mathbf{si}f_{T,t}^{S,I} \rangle = -\gamma I \langle \eta_t^X, \mathbf{si}f_{T,t}^{S,I} \rangle + \mathbf{1}_{\{S+I < n_{\max}\}} \gamma (I+1) \langle \eta_t^X, \mathbf{si}f_{T,t}^{S, I+1} \rangle.$$

In particular,  $(T, t) \mapsto \partial_T \langle \eta_t^X, \mathbf{si}f_{T,t}^{S,I} \rangle$  is continuous, as for any  $(S, J) \in \mathbb{S}$ ,  $n_{S, J}^X$  is continuous thanks to the continuity of  $t \mapsto \eta_t^X$  with regard to the total variation norm. Let  $g : \mathbb{R} \rightarrow \mathbb{R}_+$

be such that  $t \mapsto g(t)\langle \eta_t^X, \mathbf{si}_{T,t}^{S,I} \rangle$  is continuous on  $\mathbb{R}_+$  for any  $(S, I) \in \mathbb{S}$ . It then holds that

$$\begin{aligned} \frac{d}{dT} \int_0^T g(t)\langle \eta_t^X, \mathbf{si}_{T,t}^{S,I} \rangle dt &= g(T)SIn_{S,I}^X(T) - \gamma I \int_0^T g(t)\langle \eta_t^X, \mathbf{si}_{T,t}^{S,I} \rangle dt \\ &\quad + \mathbf{1}_{\{S+I < n_{\max}\}} \gamma(I+1) \int_0^T g(t)\langle \eta_t^X, \mathbf{si}_{T,t}^{S,I+1} \rangle dt. \end{aligned} \quad (3.46)$$

Similarly, for  $g : \mathbb{R} \rightarrow \mathbb{R}_+$  such that  $t \mapsto g(t)\langle \eta_t^X, \mathbf{sf}_{T,t}^{S,I} \rangle$  is continuous on  $\mathbb{R}_+$  for any  $(S, I) \in \mathbb{S}$ ,

$$\begin{aligned} \frac{d}{dT} \int_0^T g(t)\langle \eta_t^X, \mathbf{sf}_{T,t}^{S,I} \rangle dt &= g(T)Sn_{S,I}^X(T) - \gamma I \int_0^T g(t)\langle \eta_t^X, \mathbf{sf}_{T,t}^{S,I} \rangle dt \\ &\quad + \mathbf{1}_{\{S+I < n_{\max}\}} \gamma(I+1) \int_0^T g(t)\langle \eta_t^X, \mathbf{sf}_{T,t}^{S,I+1} \rangle dt. \end{aligned} \quad (3.47)$$

In particular, we may apply Equations (3.46) and (3.47) to  $g(t) = 1$ , as well as  $g(t) = \Lambda_X(t)$  and  $g(t) = e^{-\gamma t} \Lambda_X(t)$ . Indeed,  $t \mapsto \Lambda_X(t)Sn_{S,I}^X(t)$  is continuous and well defined for any  $t \geq 0$ , thanks to the inequality  $\langle \eta_t^{\bar{X}}, \mathbf{si} \rangle Sn_{S,I}^X(t) \leq (n_{\max} n_X / n_{\bar{X}}) s_{\bar{X}}(t)^2$  which ensures that despite the division by  $s_{\bar{X}}$  in the definition of  $\Lambda_X$ , there are no singularities.

In conclusion, Equation (3.45), together with Equations (3.46) and (3.47), allows to differentiate the right-hand-side of Equation (3.44). In particular, regrouping the terms factorised by  $-\gamma I$  and  $\gamma(I+1)$  allows to distinguish  $-\gamma In_{S,I}^X(T)$  and  $\gamma(I+1)n_{S,I+1}^X(T)$ , using Equation (3.44). This computation finally leads to Equation (3.42).  $\square$

We may finally focus on the main result of this section, namely Theorem 3.3.3.

*Proof of Theorem 3.3.3.* Before concluding, we need to emphasize that it would have been possible to chose  $X = W$  when replacing  $S$  and  $I$  by  $K_X S_X$  and  $K_X I_X$  for  $\mathcal{I}_G$  in Proposition 3.4.5. All of the subsequent results still hold, simply replacing the household-related quantities in the definition of the rate for mean-field infections by their workplace-related counterparts.

As a consequence, Propositions 3.4.13 and 3.4.14 show that for any  $X \in \{H, W\}$ ,

$$y_X = \left( \frac{s_X}{m_X}, \frac{i_X}{m_X}, n_{S,I}^X : (S, I) \in \mathbb{S}, n_{(S,I)}^{\bar{X}} : (S, I) \in \mathbb{S} \right)$$

satisfies the Cauchy problem (3.29) with initial condition (3.9). However, Proposition 3.4.10 ensures uniqueness of the solutions to this Cauchy problem. It hence is sensible to define, for  $t \geq 0$ ,

$$s(t) = \frac{s_H(t)}{m_H} = \frac{s_W(t)}{m_W} \quad \text{and} \quad i(t) = \frac{i_H(t)}{m_H} = \frac{i_W(t)}{m_W}.$$

This leads to dynamical system (3.8a–c) with initial conditions (3.9), and concludes the proof.  $\square$

## Discussion

This paper has focused on proposing a new reduction for an *SIR* model with two levels of mixing, which explicitly includes households and workplaces. This reduced model was obtained as its large population limit, and the associated convergence of the stochastic model was established.

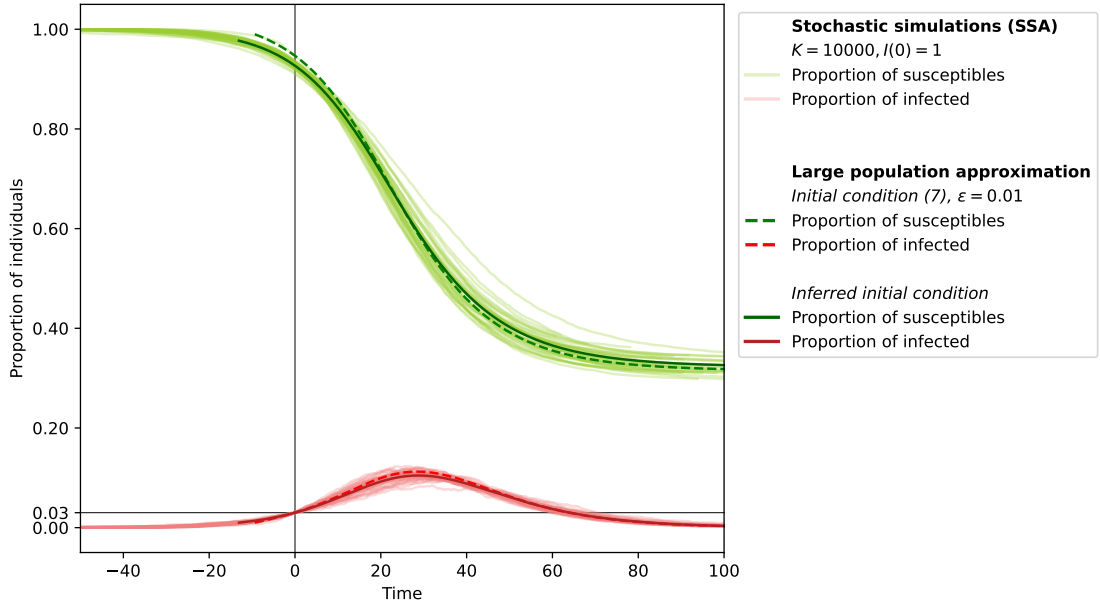
A possible model extension would be to consider a local level of mixing containing an arbitrary, yet finite, number of layers. As long as within each layer, each node is part of exactly one clique, and as long as cliques within each layer are constituted independently from one another as in the case for households and workplaces, the adaptation of the aforementioned results is expected to be straightforward. In particular, for exponentially distributed infectious period lengths, the dimension of the corresponding dynamical system should still be of order  $O(n_{\max}^2)$ , implying that the model should remain tractable.

Furthermore, we have compared the reduced model obtained in this work with the corresponding EBCM in the line of Volz et al., 2011. In the case of our household-workplace model with two levels of mixing, the EBCM seems the less appropriate choice, as it is less parsimonious and only approaches the epidemic well if the initial proportion of infected is very small. However, this may change if a more general contact structure within layers is considered, such as a configuration model for the global level, in which case it seems sensible to assume that EBCM-like equations will appear.

Finally, let us emphasize that by essence, the large population limit obtained here corresponds to a situation where the number of infected individuals is of the same order as the population size. In a realistic scenario, however, an epidemic is initiated by very few infected individuals. In the case of a large epidemic outbreak, the number of infected subsequently grows until it no longer is negligible when compared to the population size, at which point the large population limit correctly captures the dynamics of the outbreak. This raises the question: which initial condition is pertinent for the large population approximation? For uniformly mixing population, this is rather straightforward, for two main reasons. On the one hand, at each time, infected individuals are interchangeable in terms of infectious pressure exerted on susceptibles. On the other hand, the presence of recovered individuals at time  $t$  can be neglected in the study of the epidemic dynamics over the time interval  $[t, \infty)$ , simply by restricting the study to all other individuals, which still constitute a uniformly mixing population. As a consequence, in such a setting, it makes sense to suppose that at time zero, there are only infected and susceptible individuals, and that infected individuals are chosen uniformly at random in the population, with independent and identically distributed infectious period lengths.

This idea can of course be extended to our setting, and corresponds to the initial condition proposed in Theorem 3.3.3, while similar initial conditions have also been used in the literature in related settings (Volz et al., 2011; Di Lauro et al., 2021). In our model however, one actually needs to know how infected and recovered individuals are distributed among households and workplaces, meaning that neither can recovered be ignored, nor is there any reason to believe that infected individuals are distributed uniformly at random in the population. Indeed, Figure 3.3 illustrates that when compared to stochastic simulations starting from a single infected, the large population approximation with initial condition given by Equation (3.9) fails to reproduce the epidemic dynamics, while they are correctly captured when using an initial condition which is inferred from stochastic simulations. As a





**Figure 3.3:** Comparison of the stochastic model starting from a single infected, with its large population approximation for two different choices of initial condition. The first initial condition is given by Equation (3.9) for  $\varepsilon = 0.01$ . The second initial condition is obtained by simulating a large number ( $> 2000$ ) of stochastic epidemic trajectories, starting from a single infected until the proportion of infected reaches one percent. The initial condition corresponds to the average of the initial conditions observed in each simulation. Regarding the stochastic model, for this figure, 100 epidemics starting from a single infected were simulated. Similarly to Figure 3.2, only those reaching a threshold of 3% of infected are represented, and a time shift is applied to ease comparison between model outputs. Structure size distributions are those of Figure 3.1. Epidemic parameters:  $(\beta_G, \lambda_H, \lambda_W, \gamma) = (0.085, 0.1, 0.001, 0.125)$ ,  $R_I = 1.7$ .

consequence, it seems of interest to get a better understanding of this realistic initial condition, which arises from an epidemic started by a single infected. This may be achieved using a branching process approximation of the epidemic, which is designed to approach the initial, stochastic phase of the epidemic, and hence would represent a reduced model which complements the large population limit obtained in the present work.

## Appendix

### 3.A Absolute continuity of $\eta$ and related PDE system

In this appendix, we show that under the Assumptions of Theorem 3.3.2, the measure-valued process  $\eta$  obtained in the large population limit is connected to a PDE-system.

Consider any  $n \in \llbracket 1, n_{\max} \rrbracket$  and  $s \in \llbracket 0, n \rrbracket$ . We start by establishing that conditionally on a structure being of size  $n$  and containing  $s$  susceptibles, the distribution of the remaining infectious periods of its  $n - s$  members who have contracted the disease at some previous time, is absolutely continuous with respect to the Lebesgue measure on  $\mathbb{R}^{n-s}$ . More precisely, we will see that at any time  $t \geq 0$ , there exists a function  $\rho_{X,n,s}(t, \cdot) : \mathbb{R}^{n-s} \rightarrow \mathbb{R}_+$  such

that for any non-negative measurable function  $f : E \rightarrow \mathbb{R}$ , for any  $X \in \{H, W\}$  and  $t \geq 0$ ,

$$\langle \eta_t^X, \mathbf{1}_{\{\mathbf{n}(\cdot)=n, \mathbf{s}(\cdot)=s\}} f(\cdot) \rangle = \int_{\mathbb{R}^{n-s}} f(n, s, \tau) \rho_{X,n,s}(t, \tau) d\tau.$$

In the following, for any  $n \geq 1$ , let  $0_n$  be the zero of  $\mathbb{R}^n$ . Finally,  $(e_k)_{1 \leq k \leq n_{\max}}$  is the canonical basis of  $\mathbb{R}^{n_{\max}}$ .

**Lemma 3.A.1.** *For each  $X \in \{H, W\}$ , there exists a family of real-valued non negative measurable functions indexed by  $n \in \llbracket 1, n_{\max} \rrbracket$  and  $s \in \llbracket 0, n \rrbracket$  such that*

$$\forall T \geq 0, \quad \rho_{X,n,s} \in L^1([0, T] \times \mathbb{R}^{n-s}),$$

which verifies that for any  $f : E \rightarrow \mathbb{R}$  non-negative measurable function, for any  $t \geq 0$ ,

$$\langle \eta_t^X, f \rangle = \sum_{n=1}^{n_{\max}} \left( f(n, n, 0_{n_{\max}}) \rho_{X,n,n}(t) + \sum_{s=0}^{n-1} \int_{\mathbb{R}^{n-s}} f(n, s, \sum_{k=1}^{n-s} \tau_k e_k) \rho_{X,n,s}(t, \tau) d\tau \right).$$

The proof relies on studying the distribution  $\eta^X$  conditionally on structures being of given size and number of susceptibles. Absolute continuity with regard to the Lebesgue measure is then established by domination with another, absolutely continuous measure.

*Proof.* Let  $X \in \{H, W\}$ . To begin with, let us introduce the family of sets

$$E_{(n,s)} = \{x \in E : \mathbf{n}(x) = n, \mathbf{s}(x) = s\}, \quad \forall (n, s) \in \llbracket 1, n_{\max} \rrbracket \times \llbracket 0, n \rrbracket$$

which constitutes a partition of  $E$ . In the following, let  $(n(t), s(t), \tau(t))$  be distributed according to  $\eta_t^X$ , for any  $t \geq 0$ . Let  $f$  be a non-negative measurable function on  $E$ . Then

$$\begin{aligned} \langle \eta_t^X, f \rangle &= \mathbb{E}[f(n(t), s(t), \tau(t))] \\ &= \sum_{n=1}^{n_{\max}} \sum_{s=0}^n \mathbb{E}[f(n(t), s(t), \tau(t)) | n(t) = n, s(t) = s] \mathbb{P}(n(t) = n, s(t) = s). \end{aligned}$$

Let  $\tau(t) = (\tau_1(t), \dots, \tau_{n_{\max}}(t))$ . Conditionally on  $\{n(t) = n, s(t) = s\}$ , almost surely  $(n(t), s(t), \tau(t)) \in E_{(n,s)}$  and  $\tau_k(t) = 0$  for any  $k > n - s$ . Thus

$$\begin{aligned} \langle \eta_t^X, f \rangle &= \sum_{n=1}^{n_{\max}} f(n, n, 0_{n_{\max}}) \mathbb{P}(n(t) = n, s(t) = s) \\ &\quad + \sum_{n=1}^{n_{\max}} \sum_{s=0}^{n-1} \mathbb{E} \left[ f \left( n, s, \sum_{k=1}^{n-s} \tau_k(t) e_k \right) \middle| n(t) = n, s(t) = s \right] \mathbb{P}(n(t) = n, s(t) = s). \end{aligned} \tag{3.48}$$

Let  $p_X(t, n, s) = \mathbb{P}(n(t) = n, s(t) = s)$ . Then on the one hand, define for any  $n \in \llbracket 1, n_{\max} \rrbracket$ ,

$$\rho_{X,n,n}(t) = p_X(t, n, n).$$

On the other hand, for  $s < n$ , let  $\tilde{\eta}_t^{X,n,s}$  be the distribution of  $(\tau_1(t), \dots, \tau_{n-s}(t))$  conditionally on  $\{n(t) = n, s(t) = s\}$ . If we manage to show that  $\tilde{\eta}_t^{X,n,s}$  is absolutely continuous

with regard to the Lebesgue measure on  $\mathbb{R}^{n-s}$ , then letting  $\tilde{\rho}_{X,n,s}(t, \cdot)$  be its density, the result follows with

$$\forall \tau \in \mathbb{R}^{n-s}, \rho_{X,n,s}(t, \tau) = p_X(t, n, s) \tilde{\rho}_{X,n,s}(t, \tau).$$

In order to show that for any  $n \in \llbracket 1, n_{\max} \rrbracket$  and  $s \in \llbracket 1, n-1 \rrbracket$ ,  $\tilde{\eta}_t^{X,n,s}$  is absolutely continuous with regard to the Lebesgue measure on  $\mathbb{R}^{n-s}$ , we will proceed by induction.

Let us first consider the case  $s = n-1$ . We want to apply Equation (3.6) to test functions  $\varphi^* \in \mathcal{C}_b^1(E, \mathbb{R}_+)$  which are equal to zero outside of the set  $\{x \in E : \mathbf{n}(x) = n, \mathbf{s}(x) = n-1\}$ . Hence, let  $\varphi \in \mathcal{C}_b^1(\mathbb{R})$  be non-negative, and define  $\varphi^* \in \mathcal{C}_b^1(E, \mathbb{R}_+)$  by

$$\forall x \in E, \quad \varphi^*(x) = \varphi(\tau_1(x)) \mathbf{1}_{\{\mathbf{n}(x), \mathbf{s}(x) = (n, s)\}}.$$

Let  $T > 0$  and for  $t \in [0, T]$ , define  $f_t(x) = \varphi^*(\Psi(x, T, t))$  for  $x \in E$ . Using Equation (3.48) and applying Equation (3.6) to  $f_t$  after noticing that  $\mathcal{A}f_t(x) = 0$  for any  $(t, x) \in \mathbb{R}_+ \times E$ , we obtain that

$$\begin{aligned} p_X(t, n, n-1) \langle \tilde{\eta}_T^{X,n,n-1}, \varphi \rangle &= \langle \eta_T^X, \varphi^* \rangle = \langle \eta_T^X, f_T \rangle \\ &= \langle \eta_0^X, f_0 \rangle + \lambda_H \int_0^T \langle \eta_t^X, \mathbf{si}(f_t^{\mathcal{I}} - f_t) \rangle dt + \int_0^T \Lambda_X(t) \langle \eta_t^X, \mathbf{s}(f_t^{\mathcal{I}} - f_t) \rangle dt. \end{aligned}$$

As  $\varphi \geq 0$  and letting  $C = n_{\max}(\lambda_H n_{\max} + \lambda_W n_{\max} + \beta_G)$ , it follows that

$$p_X(t, n, n-1) \langle \tilde{\eta}_T^{X,n,n-1}, \varphi \rangle \leq \langle \eta_0^X, f_0 \rangle + C \int_0^T \langle \eta_t^X, f_t^{\mathcal{I}} \rangle dt. \quad (3.49)$$

Let  $g_\nu : \mathbb{R} \rightarrow \mathbb{R}_+$  be the density of  $\nu$  with respect to the Lebesgue measure. Notice that Equation (3.7) with a change of variables  $z = \sigma - T$  implies that

$$\langle \eta_0^X, f_0 \rangle = n \pi_n^X (1 - \varepsilon)^{n-1} \varepsilon \int_{\mathbb{R}} \varphi(z) g_\nu(z + T) dz.$$

Furthermore, by definition, for any  $x \in E$ ,

$$f_t^{\mathcal{I}}(x) = \mathbf{1}_{\{\mathbf{n}(x)=n, \mathbf{s}(x)=n\}} \int_0^{+\infty} \varphi(\sigma - (T-t)) g_\nu(\sigma) d\sigma.$$

Thus Equation (3.48), Fubini's theorem and a change of variables  $z = \sigma - (T-t)$  lead to

$$\int_0^T \langle \eta_t^X, f_t^{\mathcal{I}} \rangle dt = \int_{\mathbb{R}} \varphi(z) \int_0^T p_X(t, n, n) g_\nu(z + (T-t)) dt dz.$$

Let

$$\begin{aligned} H_T^{X,n,n-1} : z \in \mathbb{R} &\mapsto (n \pi_n^X (1 - \varepsilon)^{n-1} \varepsilon g_\nu(z + T)) \\ &\quad + \int_0^T p_X(t, n, n) g_\nu(z + (T-t)) dt. \end{aligned}$$

Then  $H_T^{X,n,n-1}$  is integrable on  $\mathbb{R}$  with respect to the Lebesgue measure, as its integral is equal to the right-hand side of Equation (3.49) which is finite. It further satisfies

$$p_X(T, n, s) \langle \tilde{\eta}_T^{X,n,n-1}, \varphi \rangle \leq \int_{\mathbb{R}} H_T^{X,n,n-1}(z) \varphi(z) dz.$$

In particular, let  $B$  be a Borel set which is null for the Lebesgue measure. Consider a mollifier  $\psi$  on  $\mathbb{R}$  in the same sense as for the Proof of Lemma 3.4.1. For  $\alpha > 0$ , let  $\psi_\alpha : z \in \mathbb{R} \mapsto \alpha^{-1}\psi(z/\alpha)$ . Then for any  $\alpha > 0$ , we may define  $\varphi_\alpha = \mathbf{1}_B * \psi_\alpha$ , which is an element of  $\mathcal{C}_b^1(\mathbb{R})$ . Thus for every  $\alpha > 0$ ,

$$p_X(T, n, s) \langle \tilde{\eta}_T^{X, n, n-1}, \varphi_\alpha \rangle \leq \int_{\mathbb{R}} H_T^{X, n, n-1}(z) \varphi_\alpha(z) dz. \quad (3.50)$$

As  $\varphi_\alpha$  further is bounded by 1 for every  $\alpha > 0$ , and as  $\tilde{\eta}_T^{X, n, n-1}$  is a finite measure and  $H_T^{X, n, n-1}$  is integrable on  $\mathbb{R}$  with respect to the Lebesgue measure, dominated convergence allows to let  $\alpha$  go to 0 on both sides of inequality (3.50). As  $B$  is a null set for the Lebesgue measure, the right-hand side goes to 0. This demonstrates that  $\tilde{\eta}_T^{X, n, n-1}$  is absolutely continuous with respect to the Lebesgue measure on  $\mathbb{R}$ .

Let us now suppose that  $\tilde{\eta}_T^{X, n, s+1}$  is absolutely continuous with respect to the Lebesgue measure on  $\mathbb{R}^{n-s-1}$  for some  $n' \in \llbracket 1, n_{\max} \rrbracket$  and some  $s \in \llbracket 0, n-2 \rrbracket$ . This time, let  $\varphi \in \mathcal{C}_b^1(\mathbb{R}^{n-s})$  be non-negative and define  $\varphi^* \in \mathcal{C}_b^1(E, \mathbb{R}_+)$  by

$$\forall x \in E, \quad \varphi^*(x) = \mathbf{1}_{\{\mathbf{n}(x)=n, \mathbf{s}(x)=s\}} \varphi(\tau_1(x), \dots, \tau_{n-s}(x)).$$

For any  $T \geq 0$  and  $t \in [0, T]$ , let  $f_t : x \in E \mapsto \varphi^*(\Psi(x, T, t))$ . Proceeding like before, we obtain that

$$p_X(T, n, s) \langle \tilde{\eta}_T^{X, n, s}, \varphi \rangle \leq \langle \eta_0^X, f_0 \rangle + C \int_0^T \langle \eta_t^X, f_t^T \rangle dt. \quad (3.51)$$

In order to simplify notations in the following, define for  $T \geq t \geq 0$  and  $(\tau_1, \dots, \tau_{n-s}) \in \mathbb{R}^{n-s}$  the function

$$\varphi_{T,t}(\tau_1, \dots, \tau_{n-s}) = \varphi(\tau_1 - (T-t), \dots, \tau_{n-s} - (T-t)).$$

Then similarly to the case  $(n, n-1)$ , Equation (3.7) leads to

$$\langle \eta_0^X, f_0 \rangle = \pi_n^X \binom{n}{s} (1-\varepsilon)^s \varepsilon^{n-s} \int_{\mathbb{R}^{n-s}} \varphi_{T,0}(\sigma_1, \dots, \sigma_{n-s}) \prod_{j=1}^{n-s} g_\nu(\sigma_j) d\sigma_1 \dots d\sigma_{n-s}. \quad (3.52)$$

Notice here that for any  $T > 0$  and  $t \in [0, T]$ , the application  $z \in \mathbb{R}^{n-s} \mapsto z - \sum_{k=1}^{n-s} (T-t) e_k$ , where  $e_k$  is the  $k$ -th vector of the canonical basis of  $\mathbb{R}^{n-s}$ , defines a  $\mathcal{C}^1$ -diffeomorphism from  $\mathbb{R}^{n-s}$  into itself, whose Jacobian matrix is the identity matrix of dimension  $n-s$ . Using the case  $t=0$ , Equation (3.52) becomes by change of variables:

$$\langle \eta_0^X, f_0 \rangle = \pi_n^X \binom{n}{s} (1-\varepsilon)^s \varepsilon^{n-s} \int_{\mathbb{R}^{n-s}} \varphi(z_1, \dots, z_{n-s}) \prod_{j=1}^{n-s} g_\nu(z_j + T) dz_1 \dots dz_{n-s}.$$

Moreover, notice that this time, for any  $x \in E$  and  $\sigma > 0$ ,

$$f_t^T(x) = \mathbf{1}_{\{\mathbf{n}(x)=n, \mathbf{s}(x)-1=s\}} \int_0^{+\infty} \varphi_{T,t}(\tau_1(x), \dots, \tau_{n-s-1}(x), \sigma) g_\nu(\sigma) d\sigma.$$

Thus, for  $t \in [0, T]$ ,

$$\langle \eta_t^X, f_t^T \rangle = p_X(t, n, s+1) \int_{\mathbb{R}^{n-s-1}} \left( \int_0^{+\infty} \varphi_{T,t}(\tau, \sigma) g_\nu(\sigma) d\sigma \right) \tilde{\eta}_t^{X, n, s+1}(d\tau).$$

Fubini's theorem together with the induction hypothesis yield

$$\langle \eta_t^X, f_t^T \rangle = p_X(t, n, s+1) \int_{\mathbb{R}_+} \left( \int_{\mathbb{R}^{n-s-1}} \varphi_{T,t}(\tau, \sigma) \tilde{\rho}_{X,(n,s+1)}(t, \tau) d\tau \right) g_\nu(\sigma) d\sigma.$$

Throughout the following, for  $z = (z_1, \dots, z_{n-s-1}) \in \mathbb{R}^{n-s-1}$  and  $u \in \mathbb{R}$ , we let  $\varphi(z, u) = \varphi(z_1, \dots, z_{n-s-1}, u)$ . The previously introduced family of  $\mathcal{C}^1$ -diffeomorphisms may serve again for a change of variables, allowing to obtain that

$$\langle \eta_t^X, f_t^T \rangle = p_X(t, n, s+1) \int_{\mathbb{R}} \left( \int_{\mathbb{R}^{n-s-1}} \varphi(z, u) g_\nu(u + (T-t)) \rho_{X,n,s+1}(t, z + (T-t) \sum_{k=1}^{n-s-1} e_k) dz \right) du.$$

For  $z = (z_1, \dots, z_{n-s}) \in \mathbb{R}^{n-s}$ , define

$$\begin{aligned} H_T^{X,n,s}(z) &= \pi_n^X \binom{n}{s} (1-\varepsilon)^s \varepsilon^{n-s} \prod_{j=1}^{n-s} g_\nu(z_j + T) \\ &+ \int_0^T \rho_{X,n,s+1}(t, z_1 + (T-t), \dots, z_{n-s-1} + (T-t)) g_\nu(z_{n-s} + (T-t)) dt. \end{aligned}$$

It then follows like before from Equation (3.51) that  $H_T^{X,n,s}$  is integrable on  $\mathbb{R}^{n-s}$  with respect to the Lebesgue measure, and

$$p_X(T, n, s) \langle \tilde{\eta}_T^{X,n,s}, \varphi \rangle \leq \int_{\mathbb{R}^{n-s}} H_T^{X,n,s}(z) dz.$$

The absolute continuity of  $\tilde{\eta}_T^{X,n,s}$  with respect to the Lebesgue measure on  $\mathbb{R}^{n-s}$  follows using the same arguments as previously.

Finally, the fact that for any  $n \in \llbracket 1, n_{\max} \rrbracket$  and  $s \in \llbracket 0, n \rrbracket$ ,  $\rho_{X,n,s}$  is Lebesgue-almost everywhere non-negative follows from  $\eta_t^X$  being a probability measure on  $E$  for any  $t \geq 0$ . Hence for any  $T \geq 0$ ,

$$\|\rho_{X,n,s}\|_{L^1([0,T] \times \mathbb{R}^{n-s})} = \int_0^T \langle \eta_t^X, \mathbf{1}_{\{n(\cdot)=n, s(\cdot)=s\}} \rangle \leq T.$$

This concludes the proof.  $\square$

Notice that in the case  $n = s$ ,  $\rho_{X,n,n}$  depends on  $t$  alone, whereas for  $s < n$ ,  $\rho_{X,n,s}$  depends both on  $t$  and on  $\tau \in \mathbb{R}^{n-s}$ . Throughout the following, with some abuse of notation, this distinction will mostly be implicit. The adaptation to the case  $s = n$  is generally straightforward. For instance, forthcoming Equation (3.53) reduces to

$$\frac{d}{dt} \rho_{X,n,n}(t) = -\Lambda_X(t) n \rho_{X,n,n}(t),$$

where

$$\Lambda_X(t) = \lambda_{\bar{X}} \frac{\langle \eta_t^{\bar{X}}, \mathbf{s} \mathbf{i} \rangle}{\langle \eta_t^{\bar{X}}, \mathbf{s} \rangle} + \beta_G \frac{\langle \eta_t^H, \mathbf{i} \rangle}{\langle \eta_0^H, \mathbf{n} \rangle}.$$

When integrating over  $\tau$ , the case  $n = s$  becomes

$$\int_{\mathbb{R}^0} \rho_{X,n,n}(t, \tau) d\tau = \rho_{X,n,n}(t).$$

Similarly, symbolically, for any  $\tau \in \mathbb{R}^{n-s}$ ,  $i(\tau) = 0$  whenever  $n - s = 0$ . This makes sense, as in a structure of size  $n$  containing  $n$  susceptibles, there are no infected. This leads for example to

$$\int_{\mathbb{R}^0} \mathbf{1}_{\{i(\tau)=0\}} \rho_{X,n,n}(t, \tau) d\tau = \rho_{X,n,n}(t).$$

When the adaptation is less clear, the case  $n = s$  will be treated separately.

Let  $g_\nu : \mathbb{R} \rightarrow \mathbb{R}_+$  be the density of  $\nu$  with respect to the Lebesgue measure. Further, we will consider here the example of initial condition  $\eta_{0,\varepsilon}$  defined by Equation (3.7).

**Proposition 3.A.2.** *Let  $T \geq 0$ . The family  $(\rho_{X,n,s})_{X \in \{H,W\}, 1 \leq n \leq n_{\max}, 0 \leq s \leq n}$  is a weak solution to the following system of partial differential equations. For any  $(n, s) \in \llbracket 1, n_{\max} \rrbracket \times \llbracket 0, n \rrbracket$ , for any  $\tau \in \mathbb{R}^{n-s}$  and  $t \in (0, T)$ ,*

$$\begin{aligned} \partial_t \rho_{X,n,s}(t, \tau) - \sum_{k=1}^{n-s} \partial_{\tau_k} \rho_{X,n,s}(t, \tau) &= -s(\lambda_X i(\tau) + \Lambda_X(t)) \rho_{X,n,s}(t, \tau) \\ &+ \mathbf{1}_{\{s+1 \leq n\}} (s+1) (\lambda_X i(\tau_{1,n-s-1}) + \Lambda_X(t)) \rho_{X,n,s+1}(t, \tau_{1,n-s-1}) g_\nu(\tau_{n-s}), \end{aligned} \quad (3.53)$$

with initial condition given by

$$\rho_{X,n,n}(0) = \pi_n^X \varepsilon^n \text{ and } \forall s < n, \tau \in \mathbb{R}^{n-s}, \rho_{X,n,s}(0, \tau) = \pi_n^X \binom{n}{s} \varepsilon^s (1 - \varepsilon)^{n-s} \prod_{k=1}^{n-s} g_\nu(\tau_k).$$

*Proof.* Let  $T \geq 0$  and  $(n, s) \in \llbracket 1, n_{\max} \rrbracket \times \llbracket 0, n \rrbracket$ . In the case  $n = s$ , consider a function defined for  $t \geq 0$  and  $x \in E$  by  $f(t, x) = \mathbf{1}_{\{\mathbf{n}(x)=\mathbf{s}(x)=n\}} f(t)$ , where  $f \in C^\infty([0, T])$  is compactly supported. Notice that  $\mathcal{A}f(t, x) = \frac{d}{dt} f(t)$  and  $f^\mathcal{I} = 0$ . As  $f(T) = 0$ , Equation (3.6) leads to

$$-\int_0^{+\infty} \left( \frac{d}{dt} f(t) \right) \rho_{X,n,n}(t) dt = -\int_0^{+\infty} f(t) \Lambda_X(t) n \rho_{X,n,n}(t) dt + f(0) \pi_n^X \varepsilon^n,$$

from which the desired conclusion follows.

Throughout the following, for  $k \in \llbracket 1, n_{\max} \rrbracket$  and  $\tau \in \mathbb{R}^{n_{\max}}$ , let  $\tau_{1,k} = (\tau_1, \dots, \tau_k)$ . Suppose now that  $s < n$ . Consider  $f$  measurable on  $\mathbb{R}_+ \times E$  such that  $f(t, x) = 0$  if  $(\mathbf{n}(x), \mathbf{s}(x)) \neq (n, s)$  and  $f(t, x) = f_{n,s}(t, \tau_{1,n-s}(x))$  otherwise, where  $f_{n,s} \in C^\infty([0, T] \times \mathbb{R}^{n-s})$  is compactly supported. In particular, for any  $\sigma > 0$  and  $(t, x) \in (0, +\infty) \times E$  such that  $\mathbf{s}(x) \geq 1$ , notice that  $f(t, \mathbf{j}(x, \sigma))$  is zero, unless  $s+1 \leq n$  and  $(\mathbf{n}(x), \mathbf{s}(x)) = (n, s+1)$  in which case it is equal to  $f_{n,s}(t, \tau_{1,n-s-1}(x) + \sigma e_{n-s})$ . As previously, it holds that  $f(T, \cdot) = 0$ .

Thus Equation (3.6) yields

$$\begin{aligned}
& - \int_0^{+\infty} \int_{\mathbb{R}^{n-s}} \partial_t f_{n,s}(\tau) \rho_{X,n,s}(t, \tau) d\tau dt = - \sum_{k=1}^{n-s} \int_0^{+\infty} \int_{\mathbb{R}^{n-s}} \partial_{\tau_k} f_{n,s}(\tau) \rho_{X,n,s}(t, \tau) d\tau dt \\
& - \int_0^{+\infty} \int_{\mathbb{R}^{n-s}} (\lambda_X s i(\tau) + \Lambda_X(t) s) f_{n,s}(\tau) \rho_{X,n,s}(t, \tau) d\tau dt \\
& + \mathbf{1}_{\{s=n-1\}} \int_0^{+\infty} \int_{\mathbb{R}} \Lambda_X(t) n f_{n,s}(\sigma) g_\nu(\sigma) d\sigma \rho_{X,n,n}(t) dt \\
& + \mathbf{1}_{\{s+1 < n\}} \int_0^{+\infty} \int_{\mathbb{R}^{n-s}} (\lambda_X i(\tau_{1,n-s-1}) + \Lambda_X(t)) (s+1) f_{n,s}(\tau) g_\nu(\tau_{n-s}) \rho_{X,n,s+1}(t, \tau_{1,n-s-1}) d\tau dt \\
& + \pi_n^X \binom{n}{s} \varepsilon^s (1-\varepsilon)^{n-s} \int_{\mathbb{R}^{n-s}} f(0, \tau_1, \dots, \tau_{n-s}) \prod_{k=1}^{n-s} g_\nu(\tau_k) d\tau_k.
\end{aligned}$$

This establishes that  $\rho_{X,n,s}$  is a weak solution to Equation (3.53), and concludes the proof.  $\square$

## 3.B Implementation of the large population limit

### 3.B.1 Automatic implementation of the dynamical system

It is possible to implement dynamical system (3.8a–c) in an automated way, in the sense that equations do not need to be written individually. The key lies in the fact that the set  $\mathbb{S}$  can be constructed automatically, with an intrinsic organization of the states  $(S, I)$  it contains. For example, one may arrange them by growing number  $n$  of susceptible and infected members of the structure, and for each  $n$ , by growing number  $i$  of infected, leading to

$$\mathbb{S} = \{(2, 0), (1, 1), \dots, (n_{\max}, 0), (n_{\max} - 1, 1), \dots, (1, n_{\max} - 1)\}.$$

This in turn allows to make an explicit correspondence between any state  $(S, I) \in \mathbb{S}$  and e.g. some position in a vector containing all functions of our dynamical system of interest. A similar idea was already employed in Pellis et al., 2011, for another purpose. With the previous structure of  $\mathbb{S}$ , one may for instance notice that for any  $n \in \llbracket 2, n_{\max} \rrbracket$  and  $i \in \llbracket 0, n-1 \rrbracket$ , the state  $(n-i, i)$  is the  $c(n-i, i)$ -th state enumerated in  $\mathbb{S}$ , where  $c(n-i, i) = (n-1)n/2 + i$ . As a consequence, the general expression of Equation (3.8c) may be used to handle all the dynamics of the functions  $n_{S,I}^X$ , for  $(S, I) \in \mathbb{S}$  and  $X \in \{H, W\}$ .

Also, notice that in practice, household sizes tend not to be as big as workplace sizes. It thus makes sense to distinguish explicitly a maximal size for each type of structure. This allows to avoid implementing unnecessary equations corresponding e.g. to household sizes that are not actually observed, and which thus artificially increase the dimension of the system.

### 3.B.2 Computational performance

The aim of this section is to numerically assess the computational cost associated to solving the large dimensional dynamic system (3.8a–c) in comparison to stochastic simulations using Gillespie's algorithm, also referred to as SSA (stochastic simulation algorithm). In order to

**Table 3.1:** Considered values of the contact rates and final times, grouped by value of  $R_I$  and proportions of infections per layer ( $p_G, p_H, p_W$ ) characterizing the scenarios.

	$R_I = 1.2$	$R_I = 1.4$	$R_I = 1.7$	$R_I = 2.0$	$R_I = 2.5$	$(p_G, p_H, p_W)$
$\beta_G$	0.03	0.035	0.045	0.05	0.06	
$\lambda_H$	0.05	0.07	0.09	0.15	0.2	
$\lambda_W$	0.0015	0.0016	0.0018	0.002	0.0022	(0.2, 0.4, 0.4)
$T$	130	130	105	85	75	
$\beta_G$	0.06	0.07	0.085	0.1	0.125	
$\lambda_H$	0.06	0.07	0.1	0.15	1.5	
$\lambda_W$	0.00075	0.0008	0.001	0.0011	0.00115	(0.4, 0.4, 0.2)
$T$	145	130	95	80	55	

do so, the average execution times of one stochastic simulation (SSA) and of one resolution of the associated dynamic system using the ODE solver `odeint` from the `scipy.integrate` library are compared.

Let us start by describing the general procedure. Each of the two scripts (stochastic simulation or reduced model) is executed one hundred times, all runs being independent from one another. For each run and each script, the computation time of the script of interest is measured, as well as the computation time of a reference function (summing all integers up to one billion with a simple for-loop). The ratios of the runtimes of both the script of interest and the reference function are computed. Comparison of the computation times for the stochastic and the reduced model is then based on the comparison of the averages of those normalised runtimes.

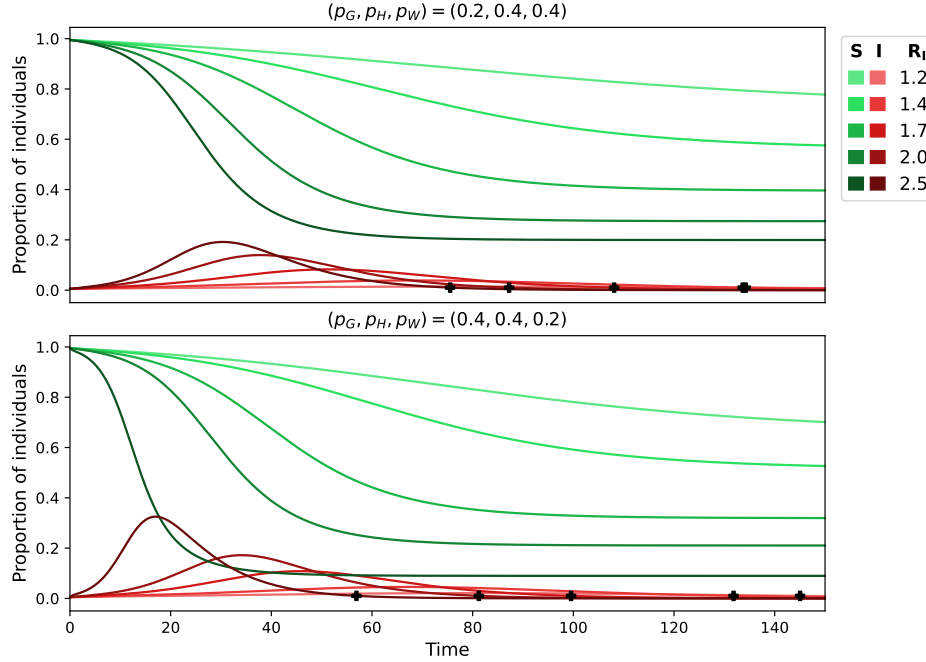
It remains to take an interest in the choice of the model parameters, namely the structure size distributions, the epidemic parameters *i.e.* the contact rates  $\beta_G$ ,  $\lambda_H$ ,  $\lambda_W$  and the removal rate  $\gamma$ , as well as the initial proportion of infected  $\varepsilon$  and the time interval  $[0, T]$  on which the epidemic is simulated. For the stochastic model, the population size  $K$  will be fixed to ten thousand individuals. For all scenarios considered here, the structure size distributions will be those of Figure 3.1, and the initial proportion of individuals will be set to  $\varepsilon = 0.005$ . Different values of the epidemic parameters will be considered, as to obtain scenarios that differ both in terms of  $R_I$  and in terms of the proportions of infections occurring within the general population, within households or within workplaces, respectively referred to as  $p_G$ ,  $p_H$  and  $p_W$ . The removal rate  $\gamma$  will be fixed at 0.125, and only the contact rates will effectively vary. In total, ten different scenarios will be used, characterized by their values of  $R_I \in \{1.2, 1.4, 1.7, 2.0, 2.5\}$  and  $(p_G, p_H, p_W) \in \{(0.2, 0.4, 0.4), (0.4, 0.4, 0.2)\}$ .

Finally, parameter  $T$  will be chosen as follows. For each set of epidemic parameters detailed above, the reduced model is used to compute the time  $T_*$  at which the epidemic falls below one percent of infected individuals in the population, after the epidemic peak.  $T$  then is determined by rounding down  $T_*$  to the closest multiple of five.

Figure 3.4 uses the reduced model to plot the trajectories of the proportion of susceptible and infected individuals in the population, for each scenario. The corresponding parameters are summarized in Table 3.1. Notice that in particular, this includes the parameters of Figure 3.2.

Let us now turn to the results. For each scenario of Table 3.1, measurement of average

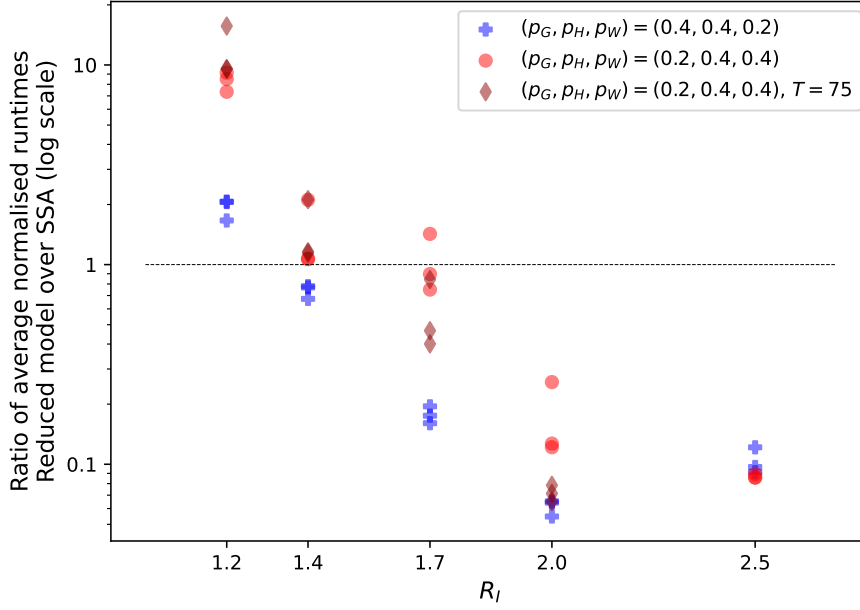




**Figure 3.4:** Proportion of susceptible ( $S$ ) and infected ( $I$ ) in the population, for each scenario detailed in Table 3.1, as given by dynamical system (3.8a–c). Scenarios are separated by values of  $(p_G, p_H, p_W)$  of infections per layer, namely  $(0.2, 0.4, 0.4)$  and  $(0.4, 0.4, 0.2)$  for the top and bottom panels, respectively. The corresponding values of  $R_I$  are indicated by the color shades, as shown in the legend. The black crosses indicate for each curve that the proportion of infected falls below the threshold of one percent.

normalised computation times was repeated three times. The results are shown in Figure 3.5, which indicates for each scenario the ratio of the average normalised runtime for one resolution of dynamical system (3.8a–c) over the average normalised runtime of one stochastic simulation. Let us first take an interest in the datasets labeled  $(p_G, p_H, p_W) = (0.2, 0.4, 0.4)$  and  $(p_G, p_H, p_W) = (0.4, 0.4, 0.2)$ . One may notice first that for each scenario, the results of all three repeats are close to one another, indicating that the results are reproducible. Further, for both possible values of  $(p_G, p_H, p_W)$ , the results indicate a shared general trend. Indeed, for values of  $R_I$  close to the critical case  $R_I = 1$ , the ratio exceeds one, and diminishes subsequently, falling below one between  $R_I = 1.4$  and  $R_I = 1.7$  and attaining values of order  $10^{-1}$ . This behavior suggests that solving dynamic system (3.8a–c) is advantageous in terms of computation time for intermediate or high values of  $R_I$ , being up to one order of magnitude faster than one stochastic simulation. As the time interval  $[0, T]$  on which the epidemic is studied originally depends on the scenario and is significantly shorter for larger values of  $R_I$ , one may wonder whether this difference influences the results. As a consequence, we have repeated the same procedure for all of the scenarios characterised by  $(p_G, p_H, p_W) = (0.2, 0.4, 0.4)$ , with fixed  $T = 75$ . Figure 3.5 shows that the associated results are very similar to those obtained previously, pleading against this hypothesis.

Of course, this comparison could be pushed further. For instance, the most basic version of the SSA algorithm was used, and more advanced methods such as  $\tau$ -leaping are expected to accelerate stochastic simulations. Also, a more thorough exploration of the parameter space



**Figure 3.5:** Ratio of the average normalised computation time for solving once dynamical system (3.8a–c) over the average normalised computation time for one stochastic simulation (SSA). This ratio was computed three times for each scenario of Table 3.1. The results are presented as a function of  $R_I$ , while colors indicate the value of  $(\rho_G, \rho_H, \rho_W)$ . Unless stated otherwise, the parameter  $T$  from Table 3.1 was used. The dotted line indicates the threshold of one.

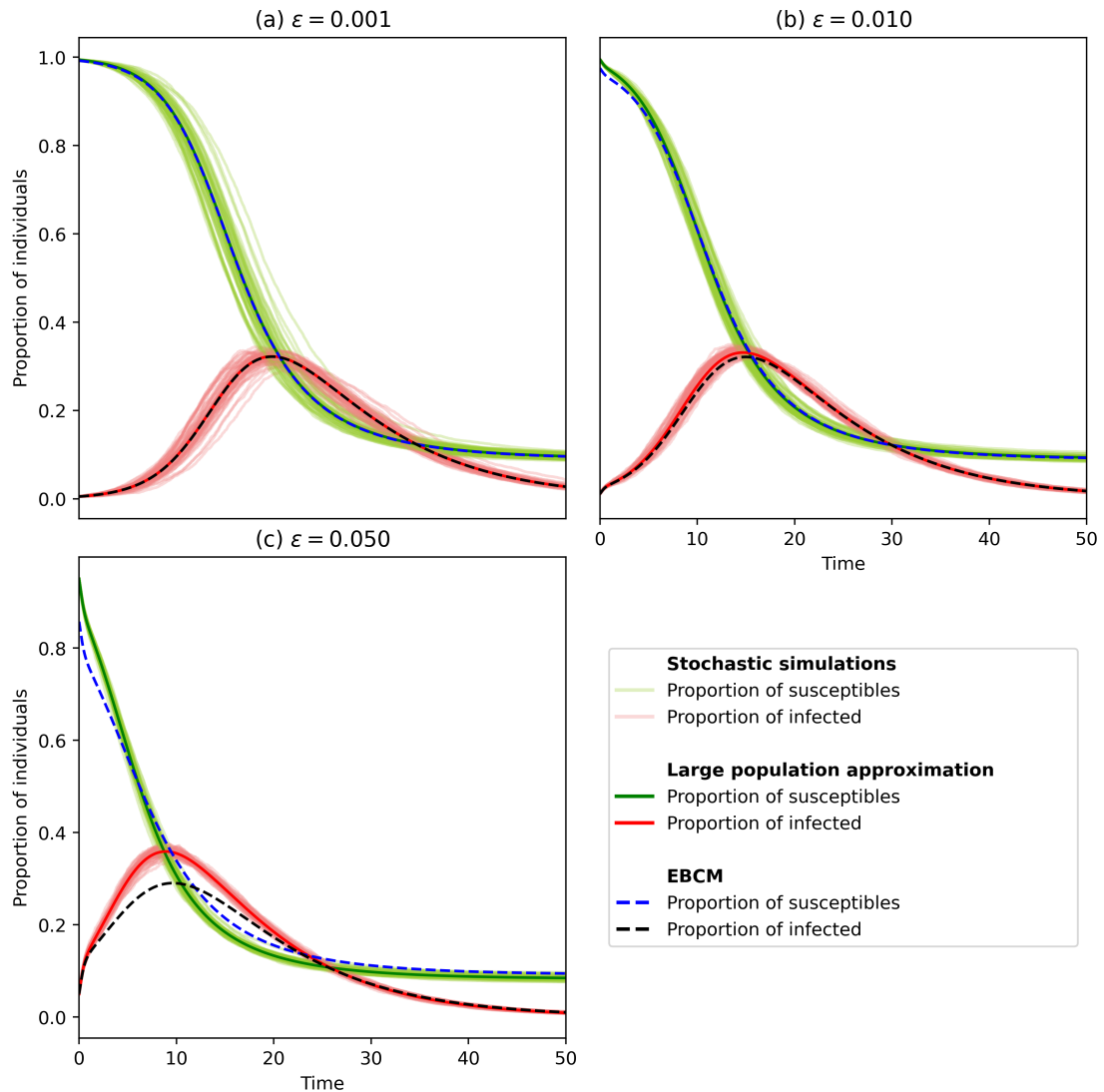
would be pertinent, assessing for instance the influence of the structure size distributions.

### 3.C Edge-based compartmental model

#### 3.C.1 Presentation of the EBCM

Let us start by describing how to obtain the population structure of the local level of mixture described in Section 3.2.1 using a clique configuration model (CCM). In our case, each node belongs to exactly one clique within each layer (one household and one workplace, respectively). Let us briefly notice that whenever a node is picked uniformly at random, the probability of it belonging to a structure of type  $X$  and size  $n$  is given by  $\hat{\pi}_n^X = n\pi_n^X/m_X$ , for any  $n \in \llbracket 1, n_{\max} \rrbracket$  and  $X \in \{H, W\}$ . As a consequence, the layer corresponding to structures of type  $X \in \{H, W\}$  is obtained by the following two steps. First, associate to each node a structure size distributed according to the size-biased law  $\hat{\pi}^X$ . This is done independently for each node. Second, for  $k \in \llbracket 1, n_{\max} \rrbracket$ , form cliques of size  $k$  by drawing uniformly without replacement  $k$ -tuples in the set of nodes of associated structure size  $k$ . This step stops when all nodes of associated clique size  $k$  belong to a clique. This procedure is repeated independently for each layer, allowing to assemble households and workplaces.

Let us now turn to deriving the EBCM. Consider  $s$  and  $i$  the proportions of susceptible and infected individuals in the population, respectively. Let  $\theta_n^X(t)$  for  $X \in \{H, W\}$  and  $n \in \llbracket 1, n_{\max} \rrbracket$  be the chance of a susceptible belonging to a structure of type  $X$  and size  $n$  to escape infection within this structure, and  $\theta^G(t)$  the chance of escaping infection through



**Figure 3.6:** Comparison of the stochastic model with the large population approximation given by dynamical system (3.8a–c) and the corresponding EBCM. Household and workplace distributions are those of Figure 3.1. Epidemic parameters are set to  $(\beta_G, \lambda_H, \lambda_W, \gamma) = (0.125, 1.5, 0.00115, 0.125)$ . Initial conditions correspond to  $\epsilon \in \{0.001, 0.01, 0.05\}$  as indicated for each panel. For each of these scenarios, Gillespie’s algorithm is used to simulate 50 trajectories of the stochastic model defined in Proposition 3.2.1 in a population of  $K = 10000$  individuals (faint lines). For Panel (a), only trajectories reaching a threshold proportion of 0.005 infected are kept, and time is shifted so that time 0 corresponds to the moment when this threshold is reached. Finally, the deterministic solution  $(s, i)$  of both dynamical system (3.8a–c) (thick lines) and the EBCM (dashed lines) are represented for each scenario. For Panel (a), the same time shifting procedure as for simulations is applied.

the mean-field level, up to time  $t \geq 0$ . The key idea is that a node is susceptible at time  $t$  if and only if it has escaped infection up to time  $t$ , and the risks of infection within each layer are independent from one another. This makes use of properties of the CCM, which heuristically explain the decoupling of the risk of infection in the two local layers from one another. Further, the fact that in an infinite population, each individual structure has a negligible impact on the proportion of infected yields the intuition behind the decorrelation of the risks of infection at the local and global level. This leads to

$$s = \theta^G \prod_{X \in \{H, W\}} \left( \sum_{n=1}^{n_{\max}} \hat{\pi}_n^X \theta_n^X \right).$$

As we are considering an *SIR* model, it follows that  $i'(t) = -s'(t) + \gamma i(t)$ , so that the difficulty resides in understanding the dynamics of  $\theta^G$  and  $\theta_n^X(t)$ , for  $X \in \{H, W\}$  and  $n \in \llbracket 1, n_{\max} \rrbracket$ .

Define for  $X \in \{H, W\}$  and  $n \in \llbracket 2, n_{\max} \rrbracket$ :

$$m_n^X = \theta^G \hat{\pi}_n^X \theta_n^X \sum_{k=1}^{n_{\max}} \hat{\pi}_k^X \theta_k^X,$$

which corresponds to the proportion of individuals who are susceptible and belong to a structure of type  $X$  and size  $n$ . Also, let  $n_{(S,I,R)}^X$  be the proportion of susceptibles belonging to a structure of type  $X$  containing exactly  $S$  susceptibles,  $I$  infected and  $R$  removed individuals. This allows us to introduce the following quantities, which participate in the rates at which a member of a structure of type  $X$  and size  $n$  is infected, either within the considered structure or outside of it, respectively:

$$T_n^X = \lambda_X \sum_{\substack{(S,I,R) \in \mathbb{N}^3 \\ S+I+R=n}} SI n_{(S,I,R)}^X \quad \text{and} \quad \tau_n^X = \left( \beta_G i + \sum_{k=1}^{n_{\max}} T_k^X \right) \frac{\hat{\pi}_n^X \theta_n^X}{\sum_{k=1}^{n_{\max}} \hat{\pi}_k^X \theta_k^X}.$$

One obtains the following dynamics:

$$\frac{d}{dt} \theta^G = -\beta_G i \theta^G, \quad \text{and} \quad \forall X \in \{H, W\}, \forall n \in \llbracket 2, n_{\max} \rrbracket, \quad \frac{d}{dt} \theta_n^X = -\frac{T_n^X}{m_n^X} \theta_n^X.$$

Further, for any  $X \in \{H, W\}$ ,  $n \in \llbracket 2, n_{\max} \rrbracket$  and  $(S, I, R) \in \mathbb{N}^3$  such that  $S + I + R = n$  and either  $S \geq 2$  or  $SI \geq 1$ :

$$\begin{aligned} \frac{d}{dt} n_{(S,I,R)}^X &= - \left( \lambda_X SI + \frac{\tau_n^X}{m_n^X} S + \gamma I \right) n_{(S,I,R)}^X \\ &\quad + \gamma (I + 1) n_{(S,I+1,R-1)}^X \mathbf{1}_{\{R \geq 1\}} \\ &\quad + \left( \lambda_X (S + 1)(I - 1) + \frac{\tau_n^X}{m_n^X} (S + 1) \right) n_{(S+1,I-1,R)}^X \mathbf{1}_{\{I > 1\}}. \end{aligned}$$

Additionally, as in a structure of size one, no infection may occur within the structure itself,  $\theta_1^H$  and  $\theta_1^W$  are constant over time. Finally, it remains to define the initial conditions.

**Table 3.2:** Numerical assessment of the computation time needed to solve either dynamical system (3.8a–c) or the EBCM introduced in Appendix 3.C. Model parameters: household and workplace size distribution from Figure 3.1;  $(\beta_G, \lambda_H, \lambda_W, \gamma) = (0.125, 1.5, 0.00115, 0.125)$ ; initial proportion of infected  $\varepsilon = 0.005$ ; resolution of the numerical system over the time interval  $[0, 30]$ .

	Runs	Normalised runtimes		
		Average	Minimum	Maximum
Dynamical system (3.8a–c)	50	0.15	0.14	0.17
EBCM	10	2076	1887	2254

Following Volz et al., 2011, we consider the case  $\varepsilon \ll 1$ . Then the only quantities which are not null at time zero are: for any  $X \in \{H, W\}$ ,  $n \in \llbracket 1, n_{\max} \rrbracket$  and  $I \in \llbracket 1, n - 1 \rrbracket$ ,

$$\begin{aligned}
 i(0) &= \varepsilon, \\
 \theta^G(0) &= \theta_n^X(0) = 1 - \varepsilon, \\
 n_{(n,0,0)}^X &= \frac{1}{n} \hat{\pi}_n^X (1 - \varepsilon)^n, \\
 n_{(n-I,I,0)}^X &= \hat{\pi}_n^X (1 - \varepsilon)^{n-I} \varepsilon^I.
 \end{aligned} \tag{3.54}$$

The proportions of susceptible and infected as predicted by both the EBCM and dynamical system (3.8a–c) are shown in Figure 3.6, for different values of  $\varepsilon$ . Let us first notice that in the case  $\varepsilon = 0.001$ , corresponding to Panel (a) of Figure 3.6, the solutions  $(s, i)$  of both the EBCM and dynamical system (3.8a–c) are in perfect accordance, emphasizing the fact that for very small values of  $\varepsilon$ , the EBCM seems to yield the correct asymptotic population dynamics. However, for larger values of  $\varepsilon$ , the EBCM struggles to reproduce these dynamics. The problem for capturing the epidemic dynamics for higher values of  $\varepsilon$  lies in the fact that defining the proper initial condition for the EBCM is not straightforward, leading to initial conditions consisting in an approximation which is only sensible whenever  $\varepsilon$  is very small.

### 3.C.2 Computational performance

In order to compare the computation times needed to solve either dynamical system (3.8a–c) or the dynamical system associated to the EBCM which has been introduced above, we will proceed similarly as in Appendix 3.B.2, making use of the ODE solver `odeint` from the `scipy.integrate` library in both cases. However, this time, only one parameter set will be used, corresponding to the parameters chosen for Panel (a) of Figure 3.2. Further, the average normalised computation time is only computed once, instead of having three repeats as in Appendix 3.B.2. Considering the relatively small fluctuations between repeats for all scenarios in Figure 3.5, this is not expected to significantly affect the qualitative result.

The model parameters and the associated average runtimes are shown in Table 3.2. Due to the excessive computation needed to solve the EBCM, only 10 runs of this script were performed. However, considering that the average normalised runtime for solving the EBCM is several orders of magnitude higher than the average normalised runtime for solving dynamical system (3.8a–c), this again is not expected to significantly alter the results. Finally, the computation times necessary for solving the EBCM are relatively homogeneous over all runs, indicating that the average computation time is not biased by an outlier.

### 3.D Proof of Proposition 3.4.10

Let us start with the following lemma, which will be needed afterwards.

**Lemma 3.D.1.** Consider a solution  $y$  of dynamical system (3.8a–c) and let  $\Delta(t) = m_X s(t) - \sum_{(S,I) \in \mathbb{S}} S n_{S,I}^X(t)$ . Then

$$\frac{d}{dt} \Delta(t) = \gamma n_{1,1}^X(t) - \left( \tau_G(t) + \frac{1}{s(t)} \tau_{\bar{X}}(t) \right) \Delta(t).$$

*Proof of Lemma 3.D.1.* Let  $X \in \{H, W\}$ . First, notice that

$$\{(S+1, I-1) : (S, I) \in \mathbb{S}, I \geq 1\} = \{(S, I) \in \mathbb{S} : S > 1\}.$$

As  $S-1 = 0$  whenever  $S = 1$ , we thus obtain that

$$\begin{aligned} & \sum_{(S,I) \in \mathbb{S}} S^2 In_{S,I}^X - \sum_{(S,I) \in \mathbb{S}} S(S+1)(I-1) n_{S+1,I-1}^X \mathbf{1}_{\{I \geq 1\}} \\ &= \sum_{(S,I) \in \mathbb{S}} S^2 In_{S,I}^X - \sum_{(S,I) \in \mathbb{S}} (S-1) S In_{S,I}^X = \sum_{(S,I) \in \mathbb{S}} S In_{S,I}^X. \end{aligned} \quad (3.55)$$

Similarly,  $\{(S, I+1) : (S, I) \in \mathbb{S}, S+I < n_{\max}\} = \mathbb{S} \setminus \{(S, I) \in \mathbb{S} : I = 0\} \cup \{(1, 1)\}$ . As  $SI = 0$  whenever  $I = 0$ , it follows that

$$\sum_{(S,I) \in \mathbb{S}} \gamma S In_{S,I}^X - \sum_{(S,I) \in \mathbb{S}} \gamma S(I+1) n_{S,I+1}^X \mathbf{1}_{\{S+I < n_{\max}\}} = \gamma n_{1,1}^X. \quad (3.56)$$

The desired conclusion then results directly from Equations (3.8a–c), regrouping the terms of the form of Equations (3.55) and (3.56) in order to simplify the expression.  $\square$

We are now ready to focus on the desired result.

*Proof of Proposition 3.4.10.* (i) By assumption,  $y(0) \in V$ . Let us start by checking that all components of  $y$ , as well as  $\Delta$ , stay non-negative over time.

Let  $t_0 \geq 0$  be such that  $y(t_0) \in V$ . If  $i(t_0) = 0$ , then  $i'(t_0) = -s'(t_0) \geq 0$  by assumption, which ensures that  $i$  will not become negative on a neighbourhood of  $t_0$ . Similar arguments hold for the lower bounds of  $\Delta$  and  $s$ , using Lemma 3.D.1 and Inequality (3.30), respectively.

Let us now turn our attention to  $n_{S,I}^X$  for  $X \in \{H, W\}$  and  $(S, I) \in \mathbb{S}$ . Recall from Equation (3.8c) that its derivative may be ill defined, due to the division by  $s(t)$ . However, inequality (3.30) ensures that the ratios  $\tau_{\bar{X}} n_{S,I}^X / s$  are well defined at all time, for any  $(S, I) \in \mathbb{S}$ . As a consequence, we may now notice as previously that, if  $n_{S,I}^X(t_0) = 0$ , Equation (3.8c) ensures that  $\frac{d}{dt} n_{S,I}^X(t_0) \geq 0$ .

The desired conclusion follows: whenever either of the quantities of interest reach zero, their derivatives are non-negative which ensures that they do not become negative shortly thereafter.

Next, let us have a look at the upper bounds. For  $X \in \{H, W\}$ , a brief computation yields

$$\frac{d}{dt} \sum_{(S,I) \in \mathbb{S}} n_{S,I}^X(t) = -\gamma n_{1,1}^X(t) \leq 0.$$

This assures that starting from  $y^* \in V$ , the inequality  $\sum_{(S,I) \in \mathbb{S}} n_{S,I}^X(t) \leq 1$  holds. For  $X \in \{H, W\}$  and  $(S, I) \in \mathbb{S}$ , it follows that if  $n_{S,I}^X(t_0) = 1$ , then for any  $(S', I') \in \mathbb{S} \setminus (S, I)$ ,  $n_{S',I'}^X(t_0) = 0$ . Thus

$$\frac{d}{dt} n_{S,I}^X(t_0) = - \left( \lambda_X SI + \tau_{\bar{X}}(t_0) \frac{S}{s(t_0)} + \tau_G(t_0) S + \gamma I \right) \leq 0,$$

from which one may deduce that  $n_{S,I}^X$  remains less than or equal to one. The remaining upper bounds on  $s$ ,  $i$  and  $s + i$  may be obtained using similar arguments.

We thus have established that if  $y^* \in V$ , then  $y(t) \in V$  for all  $t$  such that  $y$  is well defined.

(ii) Consider any  $T \geq 0$ . Let  $y = (s, i, n_{S,I}^X : X \in \{H, W\}, (S, I) \in \mathbb{S})$  be a solution to the Cauchy problem (3.29) with initial condition  $y^* \in V$ . Then it follows from Inequality (3.30) that

$$s'(t) \geq -[(\lambda_H + \lambda_W)n_{\max} + \beta_G]s(t),$$

and as further  $s \in \mathcal{C}^1(\mathbb{R}_+)$  according to the Cauchy problem, we obtain by comparison that  $s(t) \geq s(0) \exp(-[(\lambda_H + \lambda_W)n_{\max} + \beta_G]t)$  for any  $t \in [0, T]$ . As a consequence, on  $[0, T]$ ,  $s$  is bounded from below by

$$\epsilon_T = s(0) \exp(-[(\lambda_H + \lambda_W)n_{\max} + \beta_G]T).$$

In order to prove that there exists at most a unique solution  $y$  for any initial condition  $y^* \in V$ , we will distinguish two cases.

First, if  $s(0) = 0$ , then it follows that  $s(t) = 0$  for any  $t \geq 0$ , and hence  $n_{S,I}^X(t) = 0$  for any  $t \geq 0$  and any  $X \in \{H, W\}$ ,  $(S, I) \in \mathbb{S}$ . Subsequently, the equation for  $i$  reduces to  $i'(t) = -\gamma i(t)$  on  $\mathbb{R}_+$ , ensuring uniqueness of  $y$  on  $\mathbb{R}_+$ .

Second, if  $s(0) > 0$ , it follows from (i) and from our lower bound on  $s$  over  $[0, T]$  that  $y(t) \in V_T = V \cap \{s \geq \epsilon_T\}$  for any  $t \in [0, T]$ . Our aim is to show that  $f$  is Lipschitz continuous on  $V_T$ . Let  $y = (s, i, n_{S,I}^X : X \in \{H, W\}, (S, I) \in \mathbb{S})$  and  $\hat{y} = (\hat{s}, \hat{i}, \hat{n}_{S,I}^X : X \in \{H, W\}, (S, I) \in \mathbb{S})$  be two elements of  $V$ . First, consider  $f_s$ . Let  $X \in \{H, W\}$ , and define  $c_X = \lambda_X \#\mathbb{S}(n_{\max})^2/m_X$ , then

$$|\tau_X(y) - \tau_X(\hat{y})| \leq \frac{\lambda_X}{m_X} \sum_{(S,I) \in \mathbb{S}} SI |n_{S,I}^X - \hat{n}_{S,I}^X| \leq c_X \|y - \hat{y}\|_{\infty}.$$

It further holds that

$$|\tau_G(y)s - \tau_G(\hat{y})\hat{s}| \leq \beta_G(|i||s - \hat{s}| + |i - \hat{i}||\hat{s}|) \leq 2\beta_G \|y - \hat{y}\|_{\infty}.$$

Thus, letting  $c_s = c_H + c_W + 2\beta_G$ , it follows that

$$|f_s(y) - f_s(\hat{y})| \leq c_s \|y - \hat{y}\|_{\infty}.$$

Similarly, letting  $c_i = c_s + \gamma$ ,

$$|f_i(y) - f_i(\hat{y})| \leq |f_s(y) - f_s(\hat{y})| + \gamma|i - \hat{i}| \leq c_i \|y - \hat{y}\|_{\infty}.$$

It remains to focus on  $f_{X,S,I}$ . Proceeding as above, letting  $c'_X = (2\lambda_X n_{\max} + 4\beta_G + \gamma)n_{\max}$ , we obtain that

$$\begin{aligned} |f_{X,S,I}(y) - f_{X,S,I}(\hat{y})| &\leq c'_X \|y - \hat{y}\|_\infty + \left| \frac{\tau_{\bar{X}}(y)}{s} S n_{S,I}^X - \frac{\tau_{\bar{X}}(\hat{y})}{\hat{s}} S \hat{n}_{S,I}^X \right| \\ &\quad + \left| \frac{\tau_{\bar{X}}(y)}{s} (S+1) n_{S+1,I-1}^X - \frac{\tau_{\bar{X}}(\hat{y})}{\hat{s}} (S+1) \hat{n}_{S+1,I-1}^X \right| \mathbf{1}_{\{I \geq 1\}}. \end{aligned}$$

Notice that as  $y$  belongs to  $V_T$ , it follows that for  $k_X = \lambda_X n_{\max} m_X$ ,

$$(\tau_{\bar{X}}(y) S n_{S,I}^X, s) \in D_T = \{(x, y) : \epsilon_T \leq y \leq 1, 0 \leq x \leq k_X y^2\}.$$

Let  $(x, y)$  and  $(u, v)$  be two elements of  $D_T$ . It then holds that

$$\left| \frac{x}{y} - \frac{u}{v} \right| \leq \frac{1}{v} \left( \frac{x}{y} |v - y| + |x - u| \right) \leq \epsilon_T^{-1} (1 \vee k_X) (|v - y| + |x - u|).$$

Letting  $k_{X,T} = \epsilon_T^{-1} (1 \vee k_X) (\lambda_X n_{\max}^2 / m_X + 1) n_{\max}$ , we obtain for any  $X \in \{H, W\}$  and  $(S, I) \in \mathbb{S}$ ,

$$\begin{aligned} \left| \frac{\tau_{\bar{X}}(y)}{s} S n_{S,I}^X - \frac{\tau_{\bar{X}}(\hat{y})}{\hat{s}} S \hat{n}_{S,I}^X \right| &\leq \epsilon_T^{-1} (1 \vee k_X) (|\tau_{\bar{X}}(y) S n_{S,I}^X - \tau_{\bar{X}}(\hat{y}) S \hat{n}_{S,I}^X| + |s - \hat{s}|) \\ &\leq k_{X,T} \|y - \hat{y}\|_\infty. \end{aligned}$$

As a consequence, we conclude that

$$|f_{X,S,I}(y) - f_{X,S,I}(\hat{y})| \leq (c'_X + k_{X,T}) \|y - \hat{y}\|_\infty.$$

This establishes the desired Lipschitz continuity of  $f$  on  $V_T$ , with associated Lipschitz constant  $c_T = \max(c_s, c_i, c'_H + k_{H,T}, c'_W + k_{W,T})$ .

Suppose now that there are two solutions  $y$  and  $\hat{y}$  of Equations (3.8a–c) such that  $y(0) = \hat{y}(0)$ . It then holds that

$$\|y(T) - \hat{y}(T)\|_\infty \leq \int_0^T \|f(y(t)) - f(\hat{y}(t))\|_\infty dt \leq c_T \int_0^T \|y(t) - \hat{y}(t)\|_\infty dt.$$

Thus Gronwall's lemma ensures that  $\|y(t) - \hat{y}(t)\|_\infty = 0$  for any  $t \leq T$ . The desired conclusion on uniqueness follows.

(iii) In order to establish (iii), it remains to show that the initial condition  $y^*$  defined by Equation (3.9) belongs to  $V$ . Let us start by noticing that, following Equation (3.9), for any  $X \in \{H, W\}$ ,

$$\sum_{(S,I) \in \mathbb{S}} n_{S,I}^X(0) = \sum_{n=2}^{n_{\max}} \pi_n^X \sum_{I=0}^{n-1} \binom{n}{I} \varepsilon^I (1 - \varepsilon)^{n-I} = \sum_{n=2}^{n_{\max}} \pi_n^X (1 - \varepsilon^n) \leq 1.$$

Similarly,

$$\begin{aligned} \sum_{(S,I) \in \mathbb{S}} S n_{S,I}^X(0) &= \sum_{n=2}^{n_{\max}} \pi_n^X \sum_{I=0}^{n-1} (n - I) \binom{n}{I} \varepsilon^I (1 - \varepsilon)^{n-I} = \sum_{n=2}^{n_{\max}} \pi_n^X n (1 - \varepsilon) \\ &= (m_X - \pi_1^X) (1 - \varepsilon) \leq m_X s(0), \end{aligned}$$

where we have used the fact that  $s(0) = 1 - \varepsilon$ , and recognizing  $\mathbb{E}[n - B]$  for  $B \sim \mathcal{B}(n, \varepsilon)$  to deduce the second equality. The other conditions following immediately from (3.9), we conclude that  $y^* \in V$ .  $\square$





# Sensitivity of the reduction accuracy to network and epidemic parameters

---

We investigate the impact of epidemic and/or network parameters on the capacity of the reduced models obtained in Chapters 2 and 3 to approximate epidemic dynamics of household-workplace models. First, we conduct a sensitivity analysis to quantify the influence of the epidemic parameters on the precision of the reduced model developed in Chapter 2. We show that while all parameters influence the accuracy of the epidemic peak size prediction, the general population contact rate has the strongest impact on the quality of the final size approximation. Second, for the reduced model derived in Chapter 3 as the large population limit of the household-workplace model, we show that it still yields a satisfying approximation of epidemic dynamics when considering a more general network model. More precisely, this generalized household-workplace model allows to take into account the correlation of the household and workplace choice between individuals. In addition, the within-workplace contact networks are not necessarily uniformly mixing. Our results indicate that this has a stronger impact on epidemic outcomes than the correlation of structure membership between individuals.

## Contents

---

<b>4.1</b>	<b>Introduction</b>	<b>165</b>
<b>4.2</b>	<b>Sensitivity analysis of the precision of the uniformly mixing reduced model</b>	<b>167</b>
4.2.1	General approach and quantities of interest	167
4.2.2	Experimental design	168
4.2.3	Results	169
<b>4.3</b>	<b>The epidemic impact of network model perturbations</b>	<b>170</b>
4.3.1	A generalization of the household-workplace model	170
4.3.2	Simulation study design	172
4.3.3	Results	173
<b>4.4</b>	<b>Discussion</b>	<b>176</b>

---

## 4.1 Introduction

A natural question arising in the context of reduced models is to gain a thorough understanding of the approximation error committed, when compared to the model which one seeks to approach. Notably, theoretical guarantees can be achieved in some contexts. For example, one may use central limit theorems to control the fluctuation of density-dependent Markov

jump processes around their deterministic large population limit (Ethier and Kurtz, 1986, Chapter 11). In this case, the fluctuations around the reduced model are explicit, opening the door to their detailed comprehension. However, depending on the way the reduced model has been derived, such results are not always in reach.

A fruitful and more widely applicable approach consists in capturing the reduced model's precision through a simulation study, allowing to identify on which parts of the parameter space the reduction performs well. In particular, we have performed such numerical explorations in Chapter 2. We have assessed the capacity of a well calibrated, deterministic, uniformly mixing model to predict key epidemic outcomes of the household-workplace model introduced in Pellis et al., 2009. While this procedure allows to highlight conditions on the model parameters under which the reduced model is most accurate or performs poorly, it does not suffice to properly quantify the impact of the model parameters on this precision.

The quantification of the influence of a model's parameters on its outcomes is at the heart of sensitivity analysis (Saltelli et al., 2002). Different methods have been developed in this setting, with complementary strength and weaknesses, as illustrated in detail for some epidemic models by Wu et al., 2013. In particular, variance-based methods, including Sobol's variance decomposition, have a clear, quantitative interpretation, and further allow to capture interactions among different model parameters. We refer to Chapter 4 of Saltelli et al., 2007 for an exhaustive introduction to variance-based methods.

In the context of the present thesis, we are interested in developing our understanding of the accuracy of the reduced models developed in Chapters 2 and 3.

On the one hand, for the uniformly mixing reduced model developed in Chapter 2, we aim to go beyond the simulation study carried out previously by complementing it with a variance-based global sensitivity analysis. Here, we study the impact of epidemic parameters on the quality of this approximation. We are particularly interested in factor prioritization, *i.e.* in identifying the most influential parameters. The motivation behind this goal is twofold. First, it would allow to test the robustness of the qualitative conclusions drawn in Chapter 2. Second, factor prioritization may lead to an intuitive understanding of the conditions under which the reduced model performs well.

On the other hand, for the large population limit of the household-workplace model obtained in Chapter 3, it would be possible to investigate gaussian fluctuations of the stochastic finite-population model around its deterministic limit, in the spirit of aforementioned central limit theorems. However, this approach only allows comparison to the stochastic household-workplace model formalized in Section 3.2.2.

Notably, in Chapter 3, we make two strong assumptions on the contact network underlying the epidemic spread. Indeed, we first suppose that individuals choose their household and workplace independently from one another and from other individuals. This does not hold for real-life settings, as for instance workplace romances are known to be frequent, including those leading to long-time partnerships (Bozon and Heran, 1989; Kalmijn and Flap, 2001; Wilson, 2015). Second, within-structure contact networks are simplified by assuming uniform mixing within all structures. While this arguably seems reasonable for households, it is not corroborated by data regarding workplace contacts (Contreras et al., 2022).

Thus, we set up a simulation study allowing to investigate the impact of these two phenomena on epidemic outcomes of the household-workplace model. In other words, we assess if the large population limit of Chapter 3 yields a satisfying approximation of the

epidemic for a more general class of contact networks than originally considered.

This chapter is structured as follows. Section 4.2 is devoted to the sensitivity analysis of the performance of the uniformly mixing reduced model proposed in Chapter 2. Next, In Section 4.3, we introduce an extension of the household-workplace model, which serves for numerical exploration of the robustness of the approximation by the large population limit of Chapter 3 to modifications of the contact network. We conclude with a discussion of the obtained results and perspectives in Section 4.4, as this chapter contains ongoing work.

## 4.2 Sensitivity analysis of the precision of the uniformly mixing reduced model

We conduct a sensitivity analysis in order to assess the influence of the epidemic parameters on the performance of the uniformly mixing reduced model of Chapter 2. More precisely, we focus on its capacity to accurately predict the epidemic peak and final size of the household-workplace model in the large population limit. Let us start by recalling the epidemic models, which allow us to introduce a general overview of the sensitivity analysis approach adopted here. Next, we present the experimental design and analyze the results.

### 4.2.1 General approach and quantities of interest

In this section, we focus on the uniformly mixing reduced model of Chapter 2. Recall that it corresponds to a deterministic uniformly mixing *SIR* model, calibrated using the exponential growth rate  $r$  and recovery rate  $\gamma$  of the household-workplace model. Let  $s(t)$  and  $i(t)$  be the proportion of susceptibles and infected in the population, at time  $t$ . If we assume that at time 0, the proportion of susceptible and infected is given by  $(1 - \varepsilon, \varepsilon)$  for  $\varepsilon > 0$  fixed, the reduced model is given by:

$$\begin{aligned} s'(t) &= -(r + \gamma)s(t)i(t), & i'(t) &= (r + \gamma)s(t)i(t) - \gamma i(t), \\ s(0) &= 1 - \varepsilon, & i(0) &= \varepsilon. \end{aligned} \tag{4.1}$$

Throughout the chapter, the epidemic growth rate  $r$  is computed using the R package `communityEpidemics` associated to Bansaye et al., 2023a.

The impact of the epidemic parameters on the performance of the uniformly mixing reduced model (4.1) is explored through global sensitivity analysis based on Sobol's variance decomposition. As the large population limit derived in Chapter 3 is asymptotically exact and its simulation numerically advantageous on a large part of the parameter space when compared to the household-workplace model, it will be used as reference for comparison. Here, its initial condition corresponds to a proportion  $\varepsilon$  of infected chosen uniformly within in otherwise susceptible population, and we refer to Section 3.3.1 for detail. Thus, performance is measured by the absolute value of the difference between either the epidemic peak size or the epidemic final size predicted by the uniformly mixing reduced model, and by the large population limit.

Before presenting our numerical experiments in detail, let us briefly recall the definition of main and total Sobol' indices. The basic idea is to consider that the contact rates  $\beta_G$ ,  $\lambda_H$  and  $\lambda_W$  are essentially random variables, whose laws are assumed to be known and thus

**Table 4.1:** Parameters of the contact rate distributions for sensitivity analysis.

Contact rate	$(a, b)$	$\alpha$	$\beta$
$\beta_G$	(0.2, 2.5)	0.8	2.27
$\lambda_H$	(0.3, 2.0)	0.55	4.0
$\lambda_W$	(0.013, 0.03)	1.6	5.5

are fixed in the experimental design. Let  $Y$  be a model outcome of interest, in our case the absolute difference between the peak (or final) size estimates using the uniformly mixing reduced model and given by the large population limit. Notice that  $Y$  itself corresponds to a random variable, which depends on the realizations of  $\beta_G$ ,  $\lambda_H$  and  $\lambda_W$ . Then the main effect  $S_G$  and total effect  $S_{T,G}$  of  $\beta_G$  are defined as follows:

$$S_G = \frac{\text{Var}[\mathbb{E}[Y|\beta_G]]}{\text{Var}(Y)} \quad \text{and} \quad S_{T,G} = \frac{\mathbb{E}[\text{Var}[Y|\lambda_H, \lambda_W]]}{\text{Var}(Y)}.$$

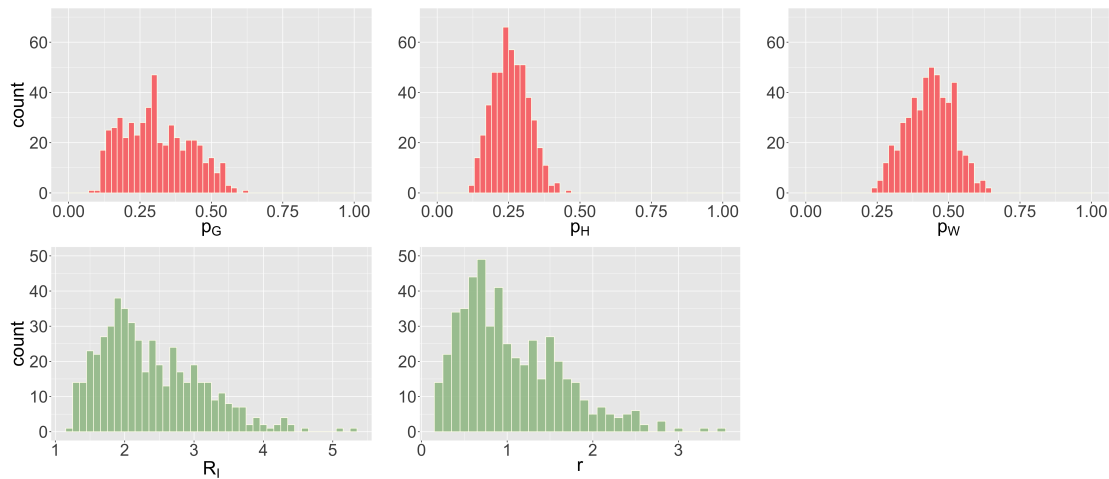
In summary, the main effect  $S_G$  yields the reduction of the variance of model output  $Y$  if the contact rate  $\beta_G$  were to be fixed: hence if  $S_G$  is large,  $\beta_G$  explains a large part of the variance of  $Y$ . Reciprocally, the total effect  $S_{T,G}$  corresponds to the expected reduction of the variance of  $Y$  if all factors but  $\beta_G$  are fixed, and thus additionally captures the impact of the interaction between  $\beta_G$  and all other parameters. The Sobol' indices of  $\lambda_H$  and  $\lambda_W$  are computed analogously, by exchanging the respective roles for each contact rate. We refer to Saltelli et al., 2002, Chapters 1 and 5 for more detail on sensitivity analysis based on variance decomposition.

#### 4.2.2 Experimental design

Let us now turn to the experimental design. Similarly to previous chapters, we consider the French household and workplace size distributions introduced in Section 2.2.3, based on INSEE data collected in 2018. Household sizes range from one to six, while the workplace size distribution has been limited to workplaces up to size fifty by splitting larger workplaces into several smaller ones. We refer to Chapter 2 and upcoming Figure 4.3 for further detail.

It remains to introduce the epidemic parameters considered. First, without loss of generality, the recovery rate is kept fixed at  $\gamma = 1$ . Hence, only the Sobol' indices of the contact rates  $\beta_G$ ,  $\lambda_H$  and  $\lambda_W$  will be computed. For this purpose, we assume that the contact rates follow shifted beta distributions, *i.e.* their laws are given by  $a + (b - a)\text{beta}(\alpha, \beta)$  where  $(a, b)$  yields the support of the contact rate distribution, and  $\text{beta}(\alpha, \beta)$  designates a random variable of density  $c_{\alpha, \beta} x^{\alpha-1} (1-x)^{\beta-1} \mathbf{1}_{\{0 \leq x \leq 1\}}$ , where  $c_{\alpha, \beta} = \Gamma(\alpha + \beta) / (\Gamma(\alpha)\Gamma(\beta))$  is the appropriate normalization constant. The considered parameters for each contact rate distribution are given in Table 4.1.

The parameters of these distributions have been chosen as to explore epidemic scenarios with reasonable characteristics similar to scenarios 1-11 of Chapter 2, in terms of  $R_I$  (Pellis et al., 2009, Supplementary Material), growth rate  $r$  (Pellis et al., 2011) and proportions of infections per layer (general population  $p_G$ , households  $p_H$  and workplaces  $p_W$ ). Namely, we focus on intermediate values of epidemic growth rate and reproduction number, while occasionally larger values occur. Similarly, proportions of infections per layer allow for the



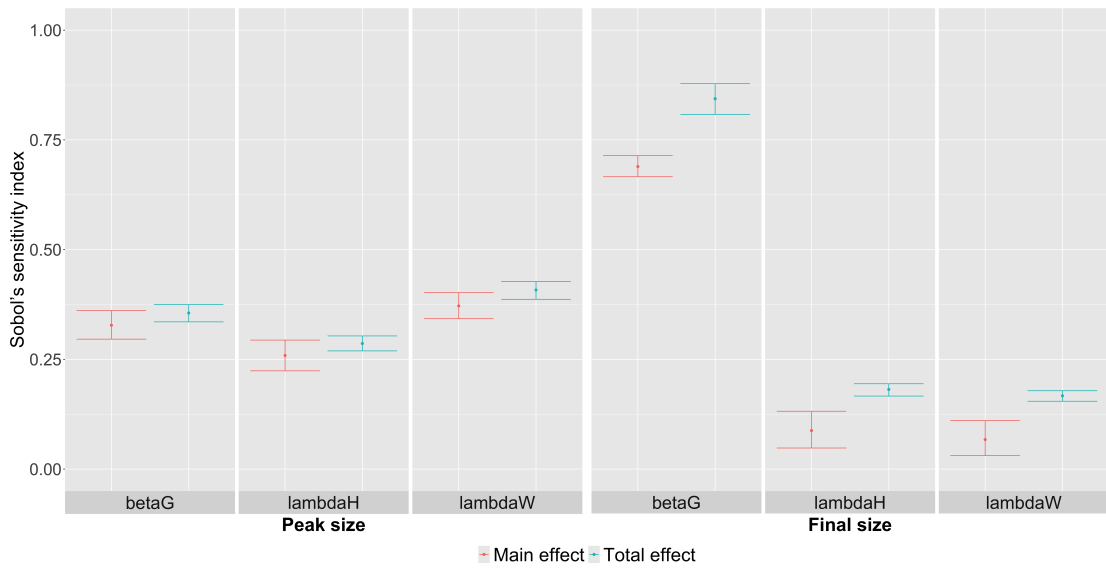
**Figure 4.1:** Empirical distributions of proportions ( $p_G, p_H, p_W$ ) of infections per layer (top), reproduction number  $R_I$  and epidemic growth rate  $r$  (bottom) for the household-workplace model with French structure size distributions, recovery rate  $\gamma = 1$  and contact rates ( $\beta_G, \lambda_H, \lambda_W$ ) sampled from the distributions given in Table 4.1. The histograms are based on 500 independent samples of the contact rates.

main source of infection to either be the general population, or the contact structures. Notably, the distribution of proportions of infections per household concentrates on lower values than the other sources of infection, which is coherent with French households being of relatively small size. We refer to Figure 4.1 for detail.

### 4.2.3 Results

Based on this experiment design, we obtain the main and total Sobol' indices shown in Figure 4.2. Let us start by focusing on the peak size error. In this case, all three contact rates appear to be influential, with  $\lambda_W$  and  $\beta_G$  having slightly predominant roles compared to  $\lambda_H$ . This may be correlated to households being relatively small, leading to generally smaller proportions of infections occurring within households as shown in Figure 4.1. In addition, there seem to be almost no interaction effects, as for each parameter, total and main effects are very close to one another. Given the importance of the epidemic growth rate for capturing the beginning of the epidemic and calibrating the reduced model, the shared importance of all three parameters may reflect the fact that the growth rate finely depends on all three contact rates.

For the final size error, however, the picture is quite different. Notably, contact rate  $\beta_G$  largely has the highest impact on the final size error, whereas  $\lambda_H$  and  $\lambda_W$  play minor roles, their effects being similar to one another. This essentially corroborates our findings in Chapter 2, which seemed to indicate that the precision of the final size prediction of the uniformly mixing reduced model mainly depends on the proportion of infections occurring in the general population. Notably, the difference between the total and main effect of  $\beta_G$  implies a significant part of interactions with the other contact rates. These interaction effects however remain minor compared to the main effect of  $\beta_G$ .



**Figure 4.2:** Main and total Sobol' indices associated to each contact rate, for the peak and final size error of the uniformly mixing reduced model compared to the asymptotic dynamics given by the large population limit. The experimental design is described in Section 4.2.2, and the sample size is equal to 5000. Computation of the Sobol' indices is achieved by the `soboljansen` implementation of the Jansen-Sobol's scheme (Jansen, 1999), in the R package `sensitivity`. 95% confidence intervals are estimated from 1000 bootstrap replicates.

### 4.3 The epidemic impact of network model perturbations

In Chapter 3, we have derived the large population limit of the household-workplace model. Notably, this was achieved under two major assumptions on the contact network, namely (1) that individuals choose their household and workplace independently from one another and from other individuals, and (2) that uniform mixing holds within all households and workplaces. As both of these assumptions do not hold in real-life setting, we aim to investigate their impact on epidemic outcomes. In other terms, if (1) and/or (2) are not satisfied, does the large population limit of the household-workplace model still yield a good approximation of the epidemic dynamics at hand?

In order to address this question, we start by introducing a generalized household-workplace model. Next, we describe the parameter sets which serve for simulations, and discuss the results of our numerical study.

#### 4.3.1 A generalization of the household-workplace model

We propose a generalized household-workplace model. As previously, it contains two levels of mixing: a global, uniformly mixing level, and a local level distinguishing two layers of household and workplace contacts, respectively. However, we modify the construction of the local contact network in order to relax Assumptions (1) and (2).

As will be described in detail below, the network model representing the local level of mixing depends on several parameters. We consider two distributions  $\pi^H$  and  $\pi^W$  on  $\llbracket 1, n_{\max} \rrbracket$ , which correspond respectively to the household and workplace size distribution

in the large population limit. In addition, we require an application  $\rho : \llbracket 1, n_{\max} \rrbracket \rightarrow [0, 1]$  such that within-workplace contact networks of workplaces of size  $n \in \llbracket 1, n_{\max} \rrbracket$  are given by Erdős-Rényi graphs  $G(n, \rho(n))$ . In particular, within workplaces, the probability that two nodes are connected by an edge thus depends on the workplace size. Finally, parameter  $q \in [0, 1]$  depicts how likely members of the same household are to belong to the same workplace.

Notably, if  $\rho$  is constant equal to one and  $q$  is null, we recover the household-workplace model. Conversely, if  $\rho$  is constant equal to zero, there are no within-workplace contacts and we recover a household model.

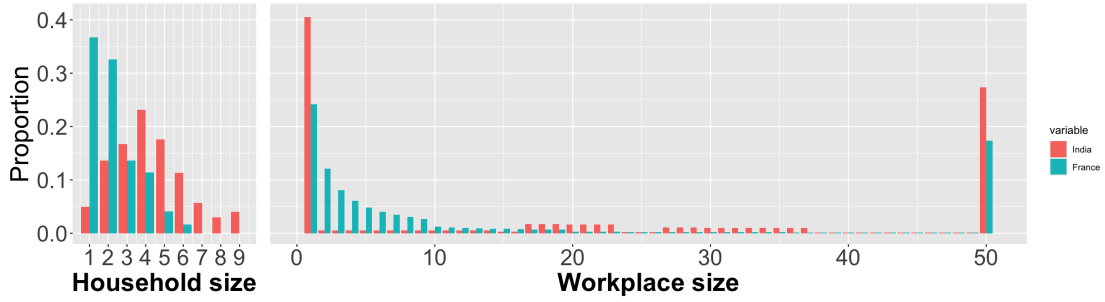
More precisely, for a population of size  $K$ , the local level of mixing of the generalized household-workplace model is constructed according to the following procedure.

1. **The household layer.** In order to assign individuals to households, we proceed as described in the previous chapters. Let  $k$  be the number of individuals who not yet belong to a household. While  $k > 0$ , pick a size  $n \sim \pi^H$  and gather a new household of size  $n^H = \min(k, n)$  by choosing uniformly at random  $n^H$  individuals among those  $k$ . Subsequently, update  $k = k - n^H$ . The procedure stops when  $k = 0$ , *i.e.* when all individuals are assigned to a household. We let  $K_H$  be the number of households, and  $\{n_h^H, h \in \llbracket 1, K_H \rrbracket\}$  the household sizes.
2. **The workplace layer.** The layer of workplaces is obtained in three steps.
  - (a) *Resolve the workplace sizes.* We start by determining the sizes of the workplaces which will constitute the workplace layer. This is achieved by proceeding as in Step 1, replacing  $\pi^H$  by  $\pi^W$ , but without actually assigning individuals to workplaces. Instead, we only keep in memory the number of required workplaces  $K_W$  and their sizes  $\{n_w^W, w \in \llbracket 1, K_W \rrbracket\}$ .
  - (b) *Assign individuals to workplaces.* Independently for each household  $h \in \llbracket 1, K_H \rrbracket$ , the workplaces of its members are chosen as follows. First, the number of its members who would like to share a workplace is given by  $\tilde{X}_h \sim 1 + \mathcal{B}(n_h^H - 1, q)$ . However, at most  $N_W = \max_{1 \leq w \leq K_W} (n_w^W)$  individuals can actually go to the same workplace. We hence uniformly pick  $X_h = \min(\tilde{X}_h, N_W)$  individuals among the  $n_h^H$  members of the household, and all of them are assigned to workplace  $w \in \llbracket 1, K_W \rrbracket$  with probability proportional to the lower integer part of  $n_w^W / X_h$ . We then let  $n_w^W = n_w^W - X_h$  in order to keep track of the remaining number of individuals who need to be assigned to this workplace. Finally, each remaining household member independently is assigned to workplace  $w \in \llbracket 1, K_W \rrbracket$  with probability proportional to  $n_w^W$ , which then becomes  $n_w^W = n_w^W - 1$ .
  - (c) *Determine the within-workplace contact networks.* For each workplace  $w \in \llbracket 1, K_W \rrbracket$ , the within-workplace contact network is given by an Erdős-Rényi graph  $G(n_w^W, \rho(n_w^W))$ .

Throughout the following, we refer to  $\rho$  as the within-workplace density. Similarly,  $q$  will occasionally be called structure overlap, as high values of  $q$  imply on average more individuals belonging to a given pair of household and workplace.

Subsequently, the epidemic spreads on the contact network. Susceptible individuals may be contaminated by contacts with infected individuals in each layer. At the global level,





**Figure 4.3:** Household and workplace size distribution distribution derived from data collected in France and India.

we consider a one-to-all contact rate  $\beta_G$ , meaning that when there are  $S$  susceptible and  $I$  infected individuals, the next infectious contact occurs at rate  $\beta_G SI/K$ . Similarly, we fix a one-to-one contact rate  $\lambda_H$  such that within a household containing  $s$  susceptible and  $i$  infected members, as uniform mixing is assumed, the next infection happens at rate  $\lambda_H si$ . Within workplaces, a contact occurs along each edge at rate  $\lambda_W$ , leading to the transmission of the disease if the edge connects a susceptible to an infected. In particular, if the workplace is totally connected and has  $s$  susceptible and  $i$  infected members, the next infectious contact occurs at rate  $\lambda_W si$ . Finally, infected individuals recover at rate  $\gamma$ .

### 4.3.2 Simulation study design

Let us now turn towards the set-up for numerical explorations. Taken together, the generalized household-workplace model depends on two groups of parameter, namely parameters  $(\pi^H, \pi^W, \rho, q)$  related to the contact network, and epidemic parameters  $(\beta_G, \lambda_H, \lambda_W, \gamma)$ . Here, we present and motivate the different parameter sets used throughout this chapter.

#### Network parameters

For the present study, two sets of structure size distributions are considered. In addition to the French structure size distributions of Section 4.2, we make use of a second couple of structure size distributions, inspired by data collected by the Indian government. Indeed, households of sizes between 1 and 9 or more are reported in IIPS, 2021, based on a survey conducted between 2019 and 2021. A proxy of a workplace size distribution has been obtained by modifying the factory size distribution reported in Government of India, 2022, which as previously is restricted to workplace sizes 1 to 50, and pondered by the Indian labor force. Inactive individuals are assigned to workplaces of size one. Children attending school contribute to workplaces of sizes chosen uniformly between 17-23 or 27-37, inspired by the average school class sizes for primary and secondary schools (Government of India, 2018). The labor force and number of children attending school are approximated using World Bank, 2024.

Both sets of structure size distributions are depicted in Figure 4.3, and will be referred to as French and Indian distributions throughout the chapter. Notably, Indian households tend to be larger than French households, implying that households may have a stronger impact on epidemic spread. Further, while the Indian workplace size distribution is a coarse

approximation of the real workplace size distribution, it shares some similarity with the French workplace size distribution due to the large share of workplaces of size 1 or 50, while being quite different otherwise (few small workplaces as data is based on factories, notable presence of school classes). As a consequence, it will not be regarded as ground truth, but occasionally serve to explore the impact of perturbing the workplace size distribution. Let us mention here that detailed data on workplace sizes is rarely available, making it difficult to construct truly realistic workplace size distributions.

Further, it remains to specify the within-workplace density  $\rho$ . Here, we will consider that within small workplaces, uniform mixing holds, whereas we want the within-workplace density to decrease for larger workplace sizes. As a first proxy, we will consider  $\rho$  to be a continuous, piecewise linear function, depending on a parameter  $p \in [0, 1]$ :

$$\rho : n \in \llbracket 1, n_{\max} \rrbracket \mapsto 1 - (1 - p) \frac{n - 20}{n_{\max} - 20} \mathbf{1}_{\{n \geq 20\}}.$$

In words, this means that workplaces of size less than 20 are uniformly mixing, whereas for  $n \geq 20$ , the within-workplace density decreases linearly with the workplace size from  $\rho(20) = 1$  to  $\rho(n_{\max}) = p$ . Here, fixing the threshold value at which workplaces cease to be uniformly mixing at 20 is an arbitrary choice. However, given that the considered workplace size distributions concentrate essentially either on very small or on very large workplaces, we do not expect this choice to have a strong qualitative impact, at least if workplaces of size less than ten are assumed uniformly mixing. Hence, this choice of  $\rho$  reduces the within-workplace density to a unique parameter  $p \in [0, 1]$ . In particular, if  $(p, q) = (1, 0)$ , we recover the household-workplace model.

### Epidemic scenarios for numerical explorations of the generalized household-workplace model

In order to assess the impact of parameters  $p$  and  $q$  on epidemic outcomes of the generalized household-workplace model when compared to the reference  $p = 1, q = 0$ , we consider several combinations of epidemic parameters and structure size distributions. More precisely, we consider the covid-like scenario ( $\beta_G = 1.00, \lambda_H = 11.852, \lambda_W = 0.009, \gamma = 1$ ) and influenza-like scenario ( $\beta_G = 0.313, \lambda_H = 0.730, \lambda_W = 0.013, \gamma = 1$ ) for the French size distributions, as proposed in Chapter 2. In addition, the same epidemic parameters are used in combination with either the Indian household size distribution and French workplace size distribution, to investigate the impact of larger household sizes, or with both Indian size distributions. Table 4.2 yields information on all 6 scenarios obtained.

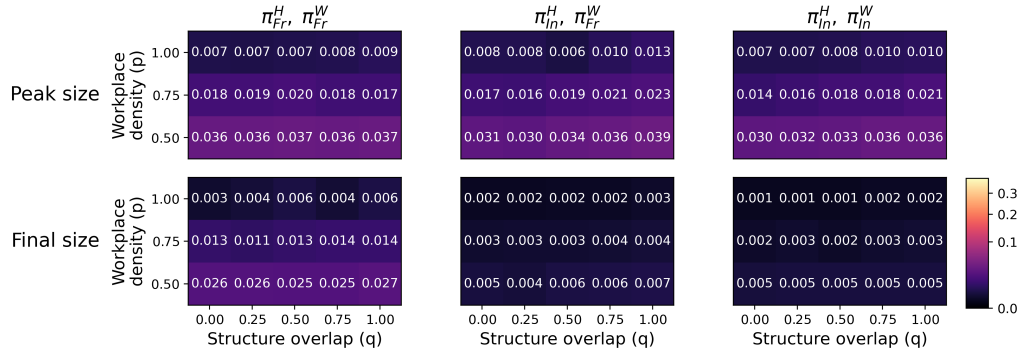
#### 4.3.3 Results

We have computed the epidemic peak and final size of all 6 scenarios in Table 4.2 using first the large population limit of the household-workplace model derived in Chapter 3. Second, for the same scenarios, peak and final size are computed using stochastic simulations of the generalized household-workplace model for  $(p, q) \in \{0.5, 0.75, 1\} \times \{0, 0.25, 0.5, 0.75, 1\}$ . The average absolute difference between the former and the latter is shown in Figures 4.4 and 4.5, for covid-like and influenza-like epidemic scenarios, respectively.

Notice that for a large share of scenarios, the approximation by the large population limit of Chapter 3 remains reasonably good, the committed error often being less than

**Table 4.2:** Values of growth rate  $r$ , reproduction number  $R_I$  and proportions of infection in the general population  $p_G$ , within households  $p_H$  and workplaces  $p_W$  for each epidemic scenario, considering the household-workplace model (recall that  $\gamma = 1$  in all scenarios).

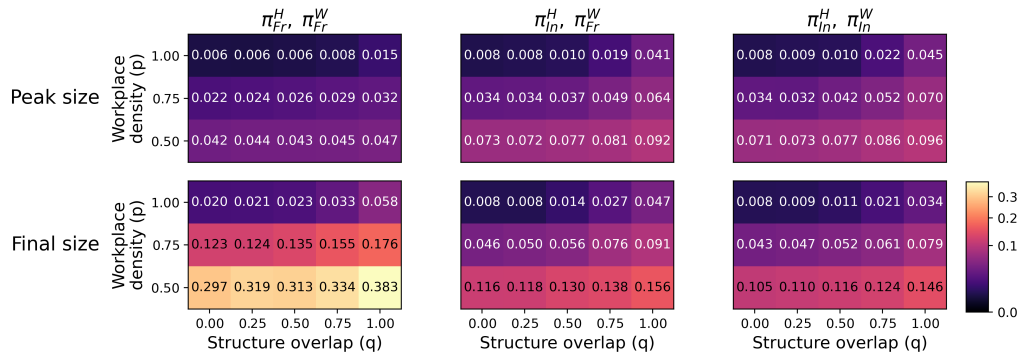
	$\pi^H$	$\pi^W$	$r$	$R_I$	$p_G$	$p_H$	$p_W$
covid-like	France	France	2.48	2.49	0.40	0.42	0.18
	India	France	4.86	3.23	0.30	0.55	0.15
	India	India	4.97	3.32	0.30	0.55	0.15
influenza-like	France	France	0.48	1.54	0.20	0.40	0.40
	India	France	1.15	2.26	0.14	0.55	0.31
	India	India	1.21	2.33	0.13	0.55	0.31



**Figure 4.4:** Heatmap of the absolute difference of the peak size (top) or final size (bottom) between the large population limit of Chapter 3 and stochastic simulations of the generalized household-workplace model, for either French structure size distributions (left), French household and Indian workplace size distribution (middle) or Indian structure size distributions (right) in the covid-like setting of Table 4.2. For the generalized reduced model, both within-workplace density parameter  $p$  and structure overlap parameter  $q$  vary for each scenario. For each set of parameters, 50 epidemic trajectories of the generalized household-workplace model have been simulated. In order to ease comparison, the colorbar is consistent between Figures 4.4 and 4.5.

5%. It performs particularly well in the covid-like setting of Figure 4.4, which may be explained by two factors. Indeed, compared to the influenza-like case, less infections take place within structures, and especially within workplaces. As  $p$  and  $q$  act on this part of the contact network, they are thus expected to have less influence in covid-like than influenza-like settings. Further, when compared to the influenza-like setting, as shown in Table 4.2, the epidemic is more intense (higher  $R_I$  and growth rate). As a consequence, the final epidemic size is expected to be very high, as corroborated by Figure 4.6, possibly also leading to smaller errors. This further explains why, in the covid-like setting, the error committed on the final size tends to be smaller than the one committed on the peak size, despite errors possibly accumulating over time.

A closer look at the results in Figure 4.5 yields further information. First, as values of growth rate and reproduction number are lower, the final size error exceeds the peak size error, as expected. Further, in the case of French structure size distributions, the strongest



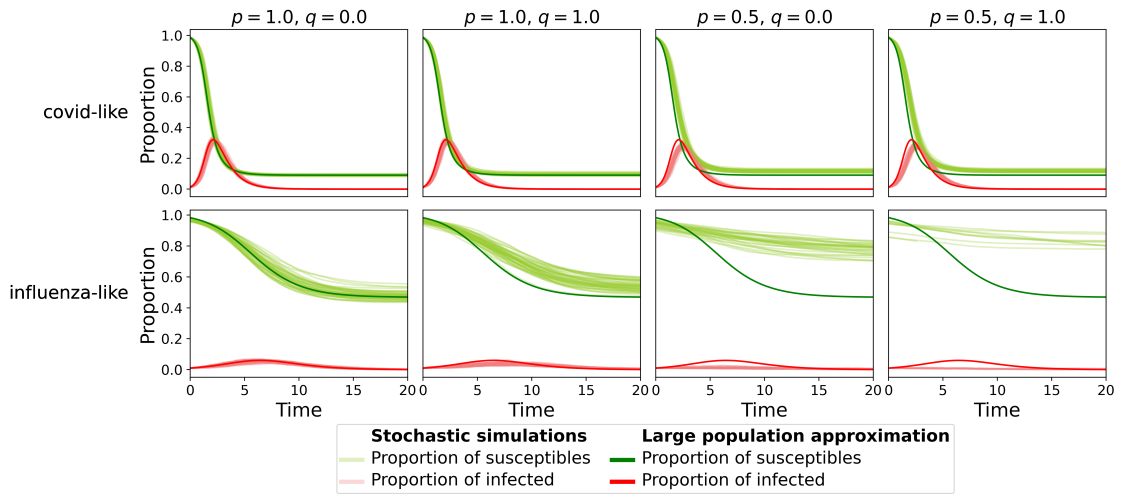
**Figure 4.5:** Heatmap of the absolute difference of the peak size (top) or final size (bottom) between the large population limit of Chapter 3 and stochastic simulations of the generalized household-workplace model, for either French structure size distributions (left), French household and Indian workplace size distribution (middle) or Indian structure size distributions (right) in the influenza-like setting of Table 4.2. For the generalized reduced model, both within-workplace density parameter  $p$  and structure overlap parameter  $q$  vary for each scenario. Numerical values are reported within each cell, in addition to the color code. For each set of parameters, 50 epidemic trajectories of the generalized household-workplace model have been simulated. In order to ease comparison, the colorbar is consistent between Figures 4.4 and 4.5.

perturbations of the within-workplace contact network density ( $p = 0.5$ ) actually appear to hinder epidemic outbreaks as illustrated in Figure 4.6, thus leading to the highest mistake in approximating the epidemic using the large population limit of Chapter 3.

Notice that when compared to the household-workplace model, the within-workplace density  $p$  generally affects a larger share of an individual's contact than the structure overlap  $q$ . This is due to households being overall smaller than workplaces. Let us illustrate this for the French structure size distribution. In the case of total structure overlap  $q = 1$ , the number of different contacts of a given individual in the local layer is reduced on average by 2.2. On the contrary, within workplaces of size  $n \geq 20$ , a fraction  $1 - \rho(n)$  of contacts is removed on average due to the reduction in within-workplace density. This represents on average 2.4 less contacts if  $p = 0.75$ , and 4.8 less contacts if  $p = 0.5$ . Thus,  $p$  is expected to be the more influential parameter. Indeed, overall, the impact of the within-workplace density  $p$  exceeds that of the structure overlap  $q$ .

When replacing the French household size distribution by its Indian counterpart (e.g. second column of Figure 4.5), households are of larger size and a higher proportion of infections occurs within that layer (Table 4.2). As a consequence, the impact of  $q$  is stronger in that context, but still remains overshadowed by  $p$ , emphasizing the importance of that parameter.

Let us also mention here that in all scenarios, the large population limit of Chapter 3 generally overestimates the epidemic peak and final size, as illustrated in Figure 4.6. This again seems consistent with the effect of parameters  $p$  and  $q$ . Indeed, the case  $(p, q) = (1, 0)$  corresponds to the setting where each individual has the highest number of distinct contacts. In other words, the large population limit of Chapter 3 corresponds to a worst-case scenario. As previously argued in the literature (Trapman et al., 2016), this can be considered to be a favorable setting as controlling the epidemic for the large population limit of Chapter 3 thus ensures to control it in more realistic cases as well.



**Figure 4.6:** Plot of the epidemic trajectories of the large population limit of Chapter 3 and stochastic simulations of the generalized household-workplace model, for French structure size distributions in the covid-like (top) or influenza-like (bottom) setting of Table 4.2. For the generalized reduced model, both within-workplace density parameter  $p$  and structure overlap parameter  $q$  vary as indicated per column. For each set of parameters, 50 epidemic trajectories of the generalized household-workplace model have been simulated; only those whose maximum proportion of infected exceeds 1% are depicted. Time is shifted so that the origin corresponds to the first time that threshold is reached, for each of the deterministic and stochastic trajectories.

Finally, it is worth mentioning that considering Indian household size distributions and French workplace size distributions yields very similar epidemic characteristics as in the case of both Indian size distributions (Table 4.2). Figures 4.4 and 4.5 show that the numerical results are very close for these settings, both in the covid-like and influenza-like scenarios. This tends to indicate a certain reproducibility of our results. In a future work, we plan to conduct a more thorough numerical exploration, regarding both the choice of epidemic parameters and structure size distributions.

## 4.4 Discussion

In this chapter, we first have conducted a global sensitivity analysis based on Sobol's variance decomposition to assess the impact of epidemic parameters on the accuracy of the uniformly mixing reduced model of Chapter 2. This analysis has been designed in order to quantify the influence of the contact rates  $(\beta_G, \lambda_H, \lambda_W)$  on the reduced model's capacity of predicting epidemic peak and final size. In particular, the structure size distributions have been fixed for this purpose. Our findings indicate that while all three parameters have significant and similar influence on the quality of the prediction of the epidemic peak size, the accuracy of the final size estimation mainly depends on  $\beta_G$ . This is coherent with previous findings in Chapter 2.

In order to consolidate our results, we plan to conduct further experiments. Notably, repeating this analysis for different workplace size distributions is crucial, in order to confirm that our conclusions are not restricted to our particular choice of French structure size

distribution. In addition, it is well known that the experimental design can have a strong impact on Sobol' indices; see Rimbaud et al., 2018 for an illustration on an epidemic model. It thus would be interesting to confront the results of several experimental designs, exploring for instance parts of the parameter space which lead to epidemic characteristics (reproduction number, growth rate and proportions of infection per layer) which differ from those explored here.

Importantly, a natural perspective is to include the effect of the structure size distributions  $\pi^H$  and  $\pi^W$  in the sensitivity analysis. This however is not straightforward. Indeed, one needs to find a pertinent parameterization of these distributions, which should combine the following two constraints. The involved parameters on which the size distributions depend should allow for meaningful interpretation of the sensitivity analysis results, and they should lead to an exploration of realistic structure size distributions. This is left for a future work.

Second, we have investigated the epidemic impact of two parameters of a generalized household-workplace model, which represent the within-workplace density and correlation of workplace choice among members of the same household. Numerical explorations show that the former has a strong impact on the epidemic peak and final size. Conversely, if the within-workplace density is high enough, or if the epidemic is intense, we find that the large population limit of Chapter 3 which ignores these phenomena remains a satisfying approximation of these epidemic dynamics. These results still need to be confirmed by a more exhaustive simulation study, covering a larger spectrum of epidemic scenarios and structure size distributions.

The importance of the within-workplace density echoes considerations in the literature about effective workplace sizes. Indeed, the number of contacts within a workplace not only depends on its size, but also on the tasks achieved by employees which may condition their contact with coworkers (Timpka et al., 2016). Additionally, it illustrates the importance of empirical studies aiming at reporting detailed contact networks within schools or workplaces (Colosi et al., 2022; Mossong et al., 2008), in order to inform modeling choices for within-workplace contacts. An interesting perspective would thus be to include more sophisticated within-workplace contact network models, possibly based on sensor contact data to mimic real-life settings (Colosi et al., 2022; Contreras et al., 2022).

Finally, a global sensitivity analysis would allow to corroborate and quantify the influence of both network and epidemic parameters on epidemic outcomes and their deviation with respect to the household-workplace model. Given the computation time necessary for numerical simulations of the epidemic trajectories of the generalized household-workplace model, we aim at constructing a metamodel on which the sensitivity analysis can be performed (Saltelli et al., 2002, Chapter 5.3 and Faivre et al., 2013, Chapter 5).



# On spinal constructions for interacting populations

We study a density-dependent Markov jump process describing a population where each individual is characterized by a type, and reproduces at rates depending both on its type and on the population state. First, using an appropriate change in probability, we show that a sampled lineage of the population process, together with the population state, can be captured by a time-inhomogeneous Markov process. This construction allows for more general sampling procedures than what was previously obtained in the literature, such as sampling restricted to subpopulations. Second, we consider the  $\psi$ -spine construction from Bansaye, 2024 and its large population approximation. Under the assumption that the deviations of the spinal population state with regard to its large population limit are controlled, we establish the large population approximation error for the spine's lineage.

## Contents

<b>5.1</b>	<b>Introduction</b>	<b>179</b>
<b>5.2</b>	<b>The population process and its <math>\psi</math>-spine construction</b>	<b>181</b>
5.2.1	The population process	181
5.2.2	Reminders on the $\psi$ -spine construction for interacting populations	182
<b>5.3</b>	<b>The <math>\psi</math>-sampled lineage</b>	<b>185</b>
5.3.1	A many-to-one formula	185
5.3.2	Proofs	187
<b>5.4</b>	<b>Quantification of the <math>\psi</math>-spine large population approximation</b>	<b>194</b>
5.4.1	Coupling the $\psi$ -spine with its large population limit	195
5.4.2	Quantifying the approximation error	196
<b>5.5</b>	<b>Discussion</b>	<b>198</b>

## 5.1 Introduction

For branching processes, the ancestral lineage of a uniformly sampled (*typical*) individual is well understood. Indeed, using an appropriate change in probability, it is possible to exhibit another genealogical tree, with one distinguished individual whose lineage behaves as the lineage of a typical individual in the original process. This distinguished individual is commonly called the spine. The key to the construction of the spinal process is that the reproduction rates of the spine are biased towards leaving more numerous descendants than other individuals, making emerge size-biased distributions. Generally speaking, the obtained spinal construction has several strengths. Notably, it allows to establish *many-to-one* formulas (e.g.



Harris et al., 2016; Harris and Roberts, 2017), which are closely related to Feynman-Kac path equations (Del Moral, 2004, Sections 1.3 and 1.4.4). Many-to-one formulas translate the average value of a functional evaluated over the lineages in the branching process, into the average of the functional evaluated along the spine, whose trajectories are exponentially weighted to capture the growth of the population. If the exponential weight is deterministic, this immediately implies a numerical advantage for computing such averages through Monte-Carlo simulations. This is due to simulations of the spine being numerically affordable, whereas simulations of the whole genealogical tree in the original branching process can be numerically challenging; see for instance Nagel et al., 2021 for comments on simulations of infection trees. Also, spinal constructions have proven an effective way of establishing classical key results on branching processes, such as the Kesten Stigum theorem (Lyons et al., 1995; Georgii and Baake, 2003). More recently, the semi-group associated to the spinal construction has proven a successful tool in the analysis of non-conservative semi-groups, extending its applications beyond branching processes (Bansaye et al., 2020; Bansaye et al., 2022).

While many models for population dynamics arising in biology and epidemiology do not satisfy *per se* the branching approximation, a classical approach is to consider regimes in which the population process can be well approached by a branching process, using coupling arguments. For instance, in epidemiology, it is well-known that at the beginning of an epidemic, the tree of infections can be captured by a branching process which neglects the depletion in susceptible individuals (Ball and Donnelly, 1995). Similarly, in order to analyze the lineage of a uniformly sampled individual in a population which is subject to evolution under a changing environment, Calvez et al., 2022 consider the stationary regime in which the population's type distribution is constant. However, such branching approximations are restricted to specific parts of the dynamics of interest only; see for instance Barbour and Reinert, 2013 and Bansaye et al., 2023b for details in the case of epidemic models and invasion processes.

In order to address this limitation, there have recently been developments towards capturing the ancestral lineage of a sampled individual, as well as the whole genealogical tree, in populations with interactions. In Bansaye, 2024 a spinal construction is developed for this setting, focusing on multi-type processes with discrete type space. The general idea consists in biasing the reproduction rates of the process, both along the spine and outside of it, according to a positive function  $\psi$  of the reproducing particle's trait  $x$  and the population's type composition  $\mathbf{z}$ . Intuitively,  $\psi(x, \mathbf{z})$  can be regarded as the individual's reproductive value or long-term fertility. Hence, when the spine reproduces, descendances with higher values of  $\psi$  given the population state are favored, while the descendances of individuals outside of the spine are biased towards rendering the population more favorable for the spine. The author derives a many-to-one formula, both for finite populations and in the large population approximation. In the latter, the population's type composition converges to the unique solution of an ordinary differential equation. Hence, a time-inhomogeneous branching process is recovered which describes the lineage of a uniformly sampled individual in the deterministic regime. Further, this spinal construction has been extended to include more general type spaces (Medous, 2023). Finally, while our focus lies on spinal constructions, let us mention Duchamps et al., 2023 where information on the infection tree in epidemic models are achieved outside of the branching regime, by tracing contamination chains backwards in time. Again, both finite populations and the large population limit are considered.

In the present chapter, we aim at complementing the above spinal constructions for populations with interactions, by addressing the following topics.

- (i) Assuming  $\psi > 0$  can be restrictive, as it prohibits for instance sampling restricted to subpopulations. Indeed, this is a pertinent setting arising in applications. For instance, in epidemiology, chains of infection leading to the contamination of vulnerable individuals, such as the elderly or immunocompromised, are naturally of particular interest. In this context, the motivation for a spinal construction designed for targeted sampling is two-fold. First, the dynamics of the spinal lineage may open the door to an intuitive understanding of such contamination chains. Second, when considering the  $\psi$ -spine under the condition that  $\psi > 0$ , sampling at time  $t$  restricted to a given subpopulation would amount to sampling within the whole population and subsequently discarding all the trajectories which do not reach the target subpopulation at time  $t$ . If the many-to-one formula is to be evaluated by Monte-Carlo methods, this can be computationally expensive if the subpopulation is rare or rarely visited by the spine.
- (ii) In addition, in the many-to-one formula of Bansaye, 2024, spinal trajectories are penalized by an exponential weight, which depends on a functional of the spine's type and population state (*cf.* Section 5.2.2). In particular, this exponential weight is not necessarily deterministic. In this case, the interpretation of the spinal process is less straight-forward, as this penalization needs to be taken into account. Further, Monte-Carlo estimations of the many-to-one formula become delicate as rare trajectories may have a tremendous impact.

As a consequence, we aim at proposing an alternative many-to-one formula which relaxes the positivity assumption (i) and which does not require exponential weighting of trajectories (ii). This will be achieved by making use of a different change in probability inspired by Marguet, 2019.

- (iii) Finally, an open question so far is to evaluate the approximation error of the spinal large population approximation. Here, we will work under the assumption that the approximation of the spinal population state by its deterministic limit is well quantified, which is reasonable as this is generally the case for density-dependent Markov jump processes, for instance through diffusion approximations (Ethier and Kurtz, 1986, Chapter 11), moderate or large deviations (Britton and Pardoux, 2019b; Pardoux, 2020)). This allows us to yield a first control of the gap between the trait sequence of the finite population spine and its large population limit by a coupling argument.

This chapter is structured as follows. The population process of interest is defined in Section 5.2, and we recall the associated  $\psi$ -spine construction introduced by Bansaye, 2024. Next, we address questions (i) and (ii) in Section 5.3, followed by (iii) in Section 5.4. Finally, Section 5.5 discusses our results and the perspectives for applications in epidemiology.

## 5.2 The population process and its $\psi$ -spine construction

### 5.2.1 The population process

We consider a structured population, where each individual has a type  $x \in \mathcal{X}$ , and we assume that the type space  $\mathcal{X}$  is finite. The number of individuals of type  $x$  in the population is

referred to as  $\mathbf{z}_x$ , and the corresponding vector  $\mathbf{z}$  describes the composition of the population. Here, we will assume that the population size cannot exceed  $K$  individuals (carrying capacity, absence of demographic births and deaths, etc.). Thus

$$\mathbf{z} \in \mathcal{Z}_K = \{\mathbf{z} \in (\mathbb{N} \cup \{0\})^{\mathcal{X}} : \|\mathbf{z}\|_1 \leq K\}.$$

Further, individuals will reproduce at rates depending on their type and the current population state. More precisely, an individual of type  $x$  may be replaced by an offspring  $\mathbf{k} = (\mathbf{k}_y \in \mathbb{N} \cup \{0\}, y \in \mathcal{X}) \in \mathcal{Z}_K$ , meaning that the individual dies and for any  $y \in \mathcal{X}$ ,  $\mathbf{k}_y$  individuals of type  $y$  are born. This occurs at rate  $\tau_{\mathbf{k}}(x, \mathbf{z})$ . We suppose  $\tau(x, \mathbf{z}) = \sum_{\mathbf{k}} \tau_{\mathbf{k}}(x, \mathbf{z}) < \infty$  for all  $x \in \mathcal{X}$ . Further, let  $(\mathbf{e}(x) : x \in \mathcal{X})$  be the canonical base of  $\mathcal{Z}_K$ , in the sense that for  $x \in \mathcal{X}$ , the only non-zero component of  $\mathbf{e}(x)$  is its  $x$  component. Then, as the population size is bounded by  $K$ ,

$$\tau_{\mathbf{k}}(x, \mathbf{z}) = 0 \quad \text{if} \quad \|\mathbf{z} + \mathbf{k} - \mathbf{e}(x)\|_1 > K.$$

In order to keep track of the genealogy, we will make use of the Ulam-Harris-Neveu notations. Let  $\mathcal{U} = \bigcup_{k \geq 1} \mathbb{N}^k$ , then  $u = (u_1, \dots, u_k) \in \mathcal{U}$  represents the  $u_k$ -th descendent of  $(u_1, \dots, u_{k-1})$  and for  $u, v \in \mathcal{U}$  we write  $u \geq v$  if  $v$  is an ancestor of  $u$ . The type of  $u \in \mathcal{U}$  will be called  $x_u$ . Hence when an individual  $u$  is replaced by its offspring  $\mathbf{k}$ , the new individuals are  $(u, 1), \dots, (u, \|\mathbf{k}\|_1)$  and we need to decide the type of each descendent. We thus consider a probability distribution  $\mathcal{Q}_{\mathbf{k}}$  on

$$\mathcal{X}_{\mathbf{k}} = \{\mathbf{x} \in \mathcal{X}^{\|\mathbf{k}\|_1} : \forall x \in \mathcal{X}, \#\{i : \mathbf{x}_i = x\} = \mathbf{k}_x\},$$

and  $(x_{(u,i)} : i \in \llbracket 1, \|\mathbf{k}\|_1 \rrbracket)$  is distributed as  $\mathcal{Q}_{\mathbf{k}}$ .

Let us now introduce the stochastic process of interest. Intuitively, it corresponds to describing the set of individuals alive and their types, at each time  $t \geq 0$ . In order to give its definition, we need the following notations. Consider the set  $E = \{(u, x_u) : u \in \mathcal{U}, x_u \in \mathcal{X}\}$ , then  $\mathcal{M}_{P,K}(E)$  is defined as the set of positive point measures  $\mu$  on  $E$  such that  $\mu(E) \leq K$ . The set of atoms of  $\mu$  is given by  $A(\mu) = \{(u, x_u) : \mu(\{(u, x_u)\}) > 0\}$  and we define  $\mathfrak{g}(\mu) \subset \mathcal{U}$  as the set of labels of its atoms, and  $\mathbf{z}(\mu) \in \mathcal{Z}_K$  as the corresponding type composition.

We start from an initial set of individuals  $\mathbb{G}(0) = \mathfrak{g} \subset \mathbb{N}$ , and the population will evolve as explained above. At each time  $t$ , let  $\mathbb{G}(t) \subset \mathcal{U}$  be the set of individuals alive. The process of interest  $(X(t), t \geq 0)$  is a Markov jump process with càdlàg trajectories, which can be defined informally by

$$\forall t \geq 0, \quad X(t) = \sum_{u \in \mathbb{G}(t)} \delta_{(u, x_u)} \in \mathcal{M}_{P,K}(E).$$

In particular, notice that there cannot be explosion, since there are at most  $K$  individuals reproducing at rate less than  $\max_{x \in \mathcal{X}, \mathbf{z} \in \mathcal{Z}_K} \tau(x, \mathbf{z})$ , which is a finite bound as  $\mathcal{X}$  and  $\mathcal{Z}_K$  are finite sets. Finally,  $Z(t) = \mathbf{z}(X(t))$  yields the composition of the population at time  $t$ , and for  $u \in \mathbb{G}(t)$  and  $s \leq t$ ,  $x_u(s)$  stands for the type of the unique ancestor of  $u$  alive at time  $s$ .

## 5.2.2 Reminders on the $\psi$ -spine construction for interacting populations

Now that the population process is properly introduced, let us briefly summarize some of the results obtained in Bansaye, 2024 which will be stated in our setting. In particular, we will

restrict ourselves to understanding the evolution of the types of the ancestors of a typical individual, and of the population composition over time, instead of dealing with the entire genealogical tree.

### The $\psi$ -spine for finite-size populations

Define the set

$$\mathcal{S}_K = \{(x, \mathbf{z}) \in \mathcal{X} \times \mathcal{Z}_K : \mathbf{z}_x \geq 1\}. \quad (5.1)$$

For any  $(x, \mathbf{z}) \in \mathcal{S}_K$ , we consider an (arbitrary) labeling  $\mathfrak{g}(\mathbf{z}) \subset \mathbb{N}$  of individuals and fix  $u_x \in \mathfrak{g}(\mathbf{z})$  such that  $u_x$  is of type  $x$ . We define the corresponding population state

$$\mathfrak{X}(x, \mathbf{z}) = \sum_{u \in \mathfrak{g}(\mathbf{z})} \delta_{(u, x_u)}.$$

We let  $\mathbb{E}_{\mathfrak{X}(x, \mathbf{z})}$  and  $\mathbb{P}_{\mathfrak{X}(x, \mathbf{z})}$  designate the expectation and probability conditionally on  $X(0) = \mathfrak{X}(x, \mathbf{z})$ . With these notations,

$$M_t f(x, \mathbf{z}) = \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{u \in \mathfrak{G}(t), u \geq u_x} f(x_u(t), Z(t)) \right]$$

is a semi-group whose generator  $\mathcal{G}$  is defined by its action on functions  $f : \mathcal{S}_K \rightarrow \mathbb{R}_+$  as follows. For any  $(x, \mathbf{z}) \in \mathcal{S}_K$ ,

$$\begin{aligned} \mathcal{G}f(x, \mathbf{z}) &= \sum_{\mathbf{k} \in \mathcal{Z}_K} \tau_{\mathbf{k}}(x, \mathbf{z}) \left( \sum_{y \in \mathcal{X}} \mathbf{k}_y f(y, \mathbf{z} + \mathbf{k} - \mathbf{e}(x)) - f(x, \mathbf{z}) \right) \\ &+ \sum_{\substack{y \in \mathcal{X} \\ \mathbf{k} \in \mathcal{Z}_K}} (\mathbf{z}_y - \mathbf{1}_{\{x=y\}}) \tau_{\mathbf{k}}(y, \mathbf{z}) (f(x, \mathbf{z} + \mathbf{k} - \mathbf{e}(y)) - f(x, \mathbf{z})). \end{aligned} \quad (5.2)$$

Consider a positive function  $\psi : \mathcal{S}_K \rightarrow (0, +\infty)$  and let  $\lambda = \mathcal{G}\psi/\psi$ . The  $\psi$ -spine construction corresponds to the stochastic process  $(Y(s), \zeta(s))_{s \geq 0}$  whose generator  $\mathcal{L}$  is defined as follows. For any  $f : \mathcal{S}_K \rightarrow \mathbb{R}_+$ ,

$$\mathcal{L}f = \frac{\mathcal{G}(\psi f)}{\psi} - \lambda f.$$

Generally speaking, the idea is that  $Y$  yields the evolution of the type along the lineage of a distinguished individual, the spine, whose reproduction rates are biased by  $\psi(\cdot, \zeta)$ , where  $\zeta$  describes the spinal population's composition. Similarly, reproduction rates outside of the spine are also biased, favoring descendances which create a population state in which the spine is more fertile. The interest in this  $\psi$ -spinal process is illustrated by the following many-to-one formula. Let  $(x, \mathbf{z}) \in \mathcal{S}_K$  and consider  $\mathbb{E}_{(x, \mathbf{z})}$  the expectation conditionally on the event  $(Y(0), \zeta(0)) = (x, \mathbf{z})$ . Also, for  $t \geq 0$ , define

$$W(t) = \exp \left( \int_0^t \lambda(Y(s), \zeta(s)) ds \right). \quad (5.3)$$

Then for any  $t \geq 0$  and any measurable function  $F : \mathbb{D}([0, t], \mathcal{S}_K) \rightarrow \mathbb{R}_+$ , for any  $(x, \mathbf{z}) \in \mathcal{S}_K$ ,

$$\mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{\substack{u \in \mathfrak{G}(t) \\ u \geq u_x}} \psi(x_u(t), Z(t)) F((x_u(s), Z(s))_{s \leq t}) \right] = \psi(x, \mathbf{z}) \mathbb{E}_{(x, \mathbf{z})} [W(t) F((Y(s), \zeta(s))_{s \leq t})]. \quad (5.4)$$

As mentioned in the introduction, the exponential weight  $W$  implies that rare trajectories of the  $\psi$ -spine may have a large impact on the right-hand side of Equation (5.4). This leads to a potential numerical challenge despite the  $\psi$ -spine being easier to simulate than a typical lineage of the original population process by itself. The issue vanishes if  $\psi$  is an eigenfunction of  $\mathcal{G}$ , as  $\lambda$  is constant in this case. However, the existence and uniqueness of such a positive eigenfunction is not always satisfied, even for classical models such as the epidemic *SIR* model. This motivates our interest in obtaining a many-to-one formula without exponential weighting the spinal trajectories. Additionally, one may be interested for instance in the typical lineage of individuals who at time  $t$  have a specific type  $x$ . The positivity assumption  $\psi > 0$  does not allow choices such as  $\psi = \mathbf{1}_{\{x\}}$ , hence this constraint needs to be incorporated in the function  $F$ . Again, this emphasizes some potential difficulties for numerically computing the right-hand side of Equation (5.4), as the event  $\{Y(t) = x\}$  may be rare itself, particularly if the type  $x$  is rare and/or poorly fertile. This will be addressed in Section 5.3.

### Large population approximation of the $\psi$ -spine

Further, it is possible to establish the large population limit of the  $\psi$ -spine construction. Let  $d = \text{Card}(\mathcal{X})$  and  $\mathcal{Z} = [0, 1]^d$ . Let  $(\tau_{\mathbf{k}}, \mathbf{k} \in (\mathbb{N} \cup \{0\})^d)$  be a family of continuous functions  $\tau_{\mathbf{k}} : \mathcal{X} \times \mathcal{Z} \rightarrow \mathbb{R}_+$  such that for each  $x \in \mathcal{X}$ , the set  $\mathcal{J}_x = \{\mathbf{k} \in (\mathbb{N} \cup \{0\})^d : \tau_{\mathbf{k}}(x, \cdot) \neq 0\}$  is finite.

For  $K \geq 1$ , consider the population process  $X^K$  where an individual of type  $x$  in a population of composition  $\mathbf{z}$  is replaced by descendance  $\mathbf{k}$  at rate  $\tau_{\mathbf{k}}(x, \mathbf{z}/K)$ . For the population size to be bounded by  $K$ , this imposes the following condition:

$$\forall K \geq 1, \forall \mathbf{z} \in \mathcal{Z}_K, \forall x \in \mathcal{X} : \mathbf{z}_x \geq 1, \forall \mathbf{k} \in \mathcal{J}_x, \quad \tau_{\mathbf{k}}(x, \mathbf{z}/K) = 0 \text{ if } \|\mathbf{z} + \mathbf{k} - \mathbf{e}(x)\|_1 \geq 1.$$

Notice that for instance, classical epidemic models satisfy this condition, as the effective population size is often kept constant. Letting  $Z^K = Z(X^K)/K$  then ensures that  $Z^K(t) \in \mathcal{Z}$  almost surely, for every  $t \geq 0$ .

Also, we consider some  $x \in \mathcal{X}$  and  $\mathbf{z}_0 \in (0, 1)^d$  and for every  $K \geq 1$ , the initial condition is set to  $X(0) = \mathfrak{X}(x, \lfloor K\mathbf{z}_0 \rfloor)$ . Define  $A(\mathbf{z}) = (A_{x,y}(\mathbf{z}))_{x,y \in \mathcal{X}}$  for  $\mathbf{z} \in \mathcal{Z}$  by  $A_{x,y}(\mathbf{z}) = \sum_{\mathbf{k} \in \mathcal{J}_x} (\mathbf{k}_y - 1) \tau_{\mathbf{k}}(x, \mathbf{z})$ . Under the hypothesis that  $\mathbf{z} \mapsto A(\mathbf{z})$  is Lipschitz continuous on  $\mathcal{Z}$ , there exists a unique solution  $z$  to the differential equation

$$z'(t) = z(t)A(z(t)), \quad z(0) = \mathbf{z}_0, \tag{5.5}$$

and we assume that for all  $t \geq 0$ ,  $\inf_{x \in \mathcal{X}, s \in [0, t]} z_x(s) > 0$ . In particular, we know that  $Z^K$  converges uniformly in probability to  $z$  on finite time intervals.

Consider a continuous function  $\psi : \mathcal{X} \times [0, 1]^d \rightarrow (0, +\infty)$  which is continuously differentiable on  $(0, 1)^d$ . Let  $(Y^K, \zeta^K)$  be the associated  $\psi$ -spine construction with initial condition  $(x, \lfloor K\mathbf{z}_0 \rfloor / K)$ , the spinal population state  $\zeta^K$  being normalized by  $K$  analogously to  $Z^K$ . Notice that for any  $t \geq 0$ ,  $(Y^K(t), \zeta^K(t)) \in \mathcal{S}$  where

$$\mathcal{S} = \{(x, \mathbf{z}) \in \mathcal{X} \times [0, 1]^d : \mathbf{z}_x > 0\}.$$

Then under the assumptions above, the  $\psi$ -spine converges in law, on finite time intervals, to  $(\Upsilon, z)$  where  $z$  is the unique solution to (5.5), and  $\Upsilon$  is a time-inhomogeneous  $\mathcal{X}$ -valued

Markov jump process which, at time  $t$ , transitions from  $x$  to  $y$  at rate

$$\sum_{\mathbf{k} \in \mathcal{J}_x} \tau_{\mathbf{k}}(x, z(t)) \mathbf{k}_y \frac{\psi(y, z(t))}{\psi(x, z(t))}.$$

In Section 5.4, we exhibit and quantify the speed of this convergence.

### 5.3 The $\psi$ -sampled lineage

In this section, our aim is to propose a many-to-one formula in the line of Equation (5.4), which relaxes the positivity assumption on  $\psi$ . In order to do so, we will introduce a change in probability, which yields a time inhomogeneous spinal construction. As we will see, this many-to-one formula further does not require to exponentially penalize the spinal trajectories. This approach is inspired by Marguet, 2019, in continuity with which we refer to the obtained time-inhomogeneous process as the  $\psi$ -auxiliary process.

#### 5.3.1 A many-to-one formula

Consider a non-negative function  $\psi$  on  $\mathcal{X} \times \mathcal{Z}_K$  such that the following application  $m_\psi$  is positive on  $\mathcal{S}_K \times [0, t]$ :

$$m_\psi(x, \mathbf{z}, s) = \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{u \in \mathbb{G}(s), u \geq u_x} \psi(x_u(s), Z(s)) \right]. \quad (5.6)$$

Notice that, by the Markov property, for  $s \in [0, t]$ ,

$$m_\psi(x, \mathbf{z}, t - s) = \mathbb{E} \left[ \sum_{u \in \mathbb{G}(t), u \geq u_x} \psi(x_u(t), Z(t)) \middle| X(s) = \mathfrak{X}(x, \mathbf{z}) \right].$$

In words,  $m_\psi(x, \mathbf{z}, t - s)$  corresponds to the  $\psi$ -weighted average of the types of individuals alive at time  $t$  who descend from a given individual of type  $x$  at time  $s$ , given that at time  $s$ , the population was in state  $\mathbf{z}$ . For instance, if  $\psi = 1$ , this yields the average number of individuals alive at time  $t$ , who descend from an individual of type  $x$  at time  $s$  when the population was in state  $\mathbf{z}$ . Similarly, if  $\psi = \mathbf{1}_{\{y\}}$  for a given  $y \in \mathcal{X}$ ,  $m_\psi(x, \mathbf{z}, t - s)$  yields the average number of descendants of type  $y$  at time  $t$ , starting at time  $s$  from an individual of type  $x$  in a population of composition  $\mathbf{z}$ .

Let us now introduce the  $\psi$ -auxiliary process, which will allow to capture the behavior of the ancestral lineage of a  $\psi$ -weighted sample of the population process. For  $t \geq 0$  fixed, we will consider the time-inhomogeneous Markov process  $(Y^{(t)}(s), \zeta^{(t)}(s))_{s \leq t}$  defined as follows. The main idea is to follow the type  $Y^{(t)}$  of a distinguished individual, which will be referred to as *spine* in analogy to classical spinal constructions. At time  $s \leq t$ , when of type  $x$  in a population of state  $\mathbf{z}$ , the spine divides to leave descendance  $\mathbf{k}$  and switch to type  $y$  with rate

$$\rho_{y, \mathbf{k}}^{(t)}(s, x, \mathbf{z}) = \tau_{\mathbf{k}}(x, \mathbf{z}) \mathbf{k}_y \frac{m_\psi(y, \mathbf{z} + \mathbf{k} - \mathbf{e}(x), t - s)}{m_\psi(x, \mathbf{z}, t - s)}.$$

In other words, compared to the original process, at any time  $s \leq t$ , transitions along the distinguished lineage are biased in favor of those which lead to a larger  $\psi$ -average descendance

at the final time  $t$ . However, due to the density-dependence of division rates, it is necessary to keep track of the population state  $\zeta^{(t)}$ . Again, transitions need to be biased, in order to account for the modified behavior of the distinguished individual when compared to the original process. As a consequence, when the population is in state  $\mathbf{z}$  and the spine of type  $x$  at time  $s \leq t$ , individuals of type  $y$  other than the spine divide to leave descendance  $\mathbf{k}$  at rate

$$\hat{\rho}_{\mathbf{k}}^{(t)}(s, y, x, \mathbf{z}) = \tau_{\mathbf{k}}(y, \mathbf{z}) \frac{m_{\psi}(x, \mathbf{z} + \mathbf{k} - \mathbf{e}(y), t - s)}{m_{\psi}(x, \mathbf{z}, t - s)}.$$

Here, the bias favors those transitions which lead to a more favorable environment for the spine, *i.e.* a population composition in which the  $\psi$ -average of the spine's descendance is high.

We will now characterize  $(Y^{(t)}(s), \zeta^{(t)}(s))_{s \leq t}$  as the unique solution of a stochastic differential equation. In order to do so, we let  $Y^{(t)}(s) \in \{\mathbf{e}(x) : x \in \mathcal{X}\}$  for any  $s \in [0, t]$ , where  $Y^{(t)}(s) = \mathbf{e}(x)$  means that the spine is of type  $x$ . Define  $E = \mathbb{R}_+ \times \mathcal{S}_K$ , and consider two independent Poisson point processes  $Q$  and  $\hat{Q}$  on  $\mathbb{R}_+ \times E$ , of density  $dr \otimes d\theta \otimes n(dy, d\mathbf{k})$  where  $dr, d\theta$  designate the Lebesgue measure and  $n$  the counting measure on  $\mathcal{S}_K$ . Here, we assume that  $Q$  and  $\hat{Q}$  are defined on the same probability space as and independently from  $(Y^{(t)}(0), \zeta^{(t)}(0))$ , whose law is supposed to be given. Then, for any  $s \in [0, t]$ ,

$$\begin{aligned} Y^{(t)}(s) &= Y^{(t)}(0) + \int_0^s \int_E \mathbf{1}_{\{\theta \leq \rho_{y, \mathbf{k}}^{(t)}(r, Y^{(t)}(r-), \zeta^{(t)}(r-))\}} (\mathbf{e}(y) - Y^{(t)}(r-)) Q(dr, d\theta, n(dy, d\mathbf{k})), \\ \zeta^{(t)}(s) &= \zeta^{(t)}(0) + \int_0^s \int_E \mathbf{1}_{\{\theta \leq \rho_{y, \mathbf{k}}^{(t)}(r, Y^{(t)}(r-), \zeta^{(t)}(r-))\}} (\mathbf{k} - Y^{(t)}(r-)) Q(dr, d\theta, n(dy, d\mathbf{k})) \\ &\quad + \int_0^s \int_E \mathbf{1}_{\{\theta \leq (\zeta_y^{(t)}(r-) - \mathbf{1}_{\{Y^{(t)}(r-)=y\}}) \hat{\rho}_{\mathbf{k}}^{(t)}(r, y, Y^{(t)}(r-), \zeta^{(t)}(r-))\}} (\mathbf{k} - \mathbf{e}(y)) \hat{Q}(dr, d\theta, n(dy, d\mathbf{k})). \end{aligned} \tag{5.7}$$

**Remark 5.3.1.** *Throughout the following, in order to simplify notations, we will make no distinction between the sets  $\mathcal{X}$  and  $\{\mathbf{e}(x) : x \in \mathcal{X}\}$ , based on the natural bijection between the two sets. For example,  $Y^{(t)}(s) = x$  is equivalent to  $Y^{(t)}(s) = \mathbf{e}(x)$ . Similarly, to every real-valued function  $f$  on  $\mathcal{S}_K$ , we assign a function  $\hat{f}$  on  $\{\mathbf{e}(x) : x \in \mathcal{X}\}$  by  $\hat{f}(\mathbf{e}(x)) = f(x)$ , the application  $f \mapsto \hat{f}$  being a bijection between the sets of real-valued functions on  $\mathcal{S}_K$  and on  $\{\mathbf{e}(x) : x \in \mathcal{X}\}$ . Thus, we will always consider  $Y^{(t)}$  to take values in the Skorokhod space  $\mathbb{D}([0, t], \mathcal{X})$ , unless mentioned otherwise.*

Our first result shows that the process  $(Y^{(t)}(s), \zeta^{(t)}(s))_{s \leq t}$  is now well defined, and additionally provides its semi-group  $\mathcal{R}^{(t)} = (\mathcal{R}_{r,s}^{(t)}, r \leq s \leq t)$ . We recall that the latter is characterized by its action on non-negative functions  $f$  on  $\mathcal{S}_K$ : for  $r \leq s \leq t$  and  $(x, \mathbf{z}) \in \mathcal{S}_K$ ,

$$\mathcal{R}_{r,s}^{(t)}(x, \mathbf{z}) = \mathbb{E}[f(Y^{(t)}(s), \zeta^{(t)}(s)) | (Y^{(t)}(r), \zeta^{(t)}(r)) = (x, \mathbf{z})].$$

**Proposition 5.3.2.** *Equation (5.7) admits a unique strong solution  $(Y^{(t)}, \zeta^{(t)})$  in the Skorokhod space  $\mathbb{D}([0, t], \mathcal{S}_K)$ . Its semi-group  $\mathcal{R}^{(t)}$  is defined by:*

$$\forall 0 \leq r \leq s \leq t, \quad \mathcal{R}_{r,s}^{(t)} = e^{\int_r^s A_{\tau}^{(t)} d\tau}, \tag{5.8}$$

where the operator  $A^{(t)}$  is characterized by its action on non-negative functions  $f$  on  $\mathcal{S}_K$ . For any  $s \in [0, t)$  and  $(x, \mathbf{z}) \in \mathcal{S}_K$ ,

$$A_s^{(t)} f(x, \mathbf{z}) = m_\psi(x, \mathbf{z}, t - s)^{-1} (\mathcal{G}(m_\psi(\cdot, t - s) f(\cdot))(x, \mathbf{z}) - \mathcal{G}(m_\psi(\cdot, t - s))(x, \mathbf{z})). \quad (5.9)$$

Notice that the operator  $A^{(t)}$  corresponds to the generator of the semi-group  $\mathcal{R}^{(t)}$ . As our state space  $\mathcal{S}_K$  is finite,  $A_\tau^{(t)}$  can be represented as a matrix, whose elements correspond to the instantaneous transition rates at time  $\tau$  which can be recovered by taking  $f = \mathbf{1}_{\{(y, \mathbf{v})\}}$  for  $(y, \mathbf{v}) \in \mathcal{S}_K$ . In particular, this ensures that the generator  $A^{(t)}$  uniquely characterizes the semi-group  $\mathcal{R}^{(t)}$ , and thus the Markov process  $(Y^{(t)}, \zeta^{(t)})$ . The proof of the proposition is postponed to Section 5.3.2.

We are now ready to state our main result. With slight abuse of notation,  $\mathbb{E}_{(x, \mathbf{z})}$  will designate the expectation conditionally on the event  $(Y^{(t)}(0), \zeta^{(t)}(0)) = (x, \mathbf{z})$ .

**Theorem 5.3.3.** *For any  $t \geq 0$  and any measurable function  $F : \mathbb{D}([0, t], \mathcal{S}_K) \rightarrow \mathbb{R}_+$ , for any  $(x, \mathbf{z}) \in \mathcal{S}_K$ ,*

$$\begin{aligned} \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{u \in \mathbb{G}(t), u \geq u_x} \psi(x_u(t), Z(t)) F((x_u(s), Z(s))_{s \leq t}) \right] \\ = m_\psi(x, \mathbf{z}, t) \mathbb{E}_{(x, \mathbf{z})} [F((Y^{(t)}(s), \zeta^{(t)}(s))_{s \leq t})]. \end{aligned} \quad (5.10)$$

Before proceeding to the proof of Theorem 5.3.3, let us compare the obtained  $\psi$ -auxiliary process with the  $\psi$ -spine from Bansaye, 2024. First, we may notice that both spinal constructions remain similar in spirit, as ours may be regarded as a *time-inhomogeneous  $m_\psi$ -transform*, instead of the classical  $\psi$ -transform. Second, in the special case where  $\psi$  is an eigenfunction of the generator  $\mathcal{G}$  introduced above, a brief computation shows that Equation (5.10) amounts to the Feynman-Kac formula of Equation (5.4), as one would expect.

### 5.3.2 Proofs

The general idea is to proceed as follows. We start by showing that the  $\psi$ -auxiliary process is well defined, by Proposition 5.3.2, and compute its generator  $A^{(t)}$ . Next, we introduce a time-inhomogeneous semi-group corresponding essentially to the left-hand side of the many-to-one formula given by Equation (5.10) normalized by  $m_\psi$ , and show that its generator is equal to  $A^{(t)}$ . As mentioned previously, the considered state space being finite, the generator uniquely characterizes the time-inhomogeneous semi-group. This finally allows to establish Theorem 5.3.3.

#### Existence and uniqueness of the $\psi$ -auxiliary process

We first establish Proposition 5.3.2, ensuring that the  $\psi$ -auxiliary process is well defined. We start with a technical lemma.

**Lemma 5.3.4.** *For any  $t \geq 0$ , for any  $(x, \mathbf{z}) \in \mathcal{S}_K$ , the function  $s \mapsto m_\psi(x, \mathbf{z}, t - s)$  is differentiable on  $(0, t)$ , and we have:*

$$\partial_s m_\psi(x, \mathbf{z}, t - s) = -\mathcal{G}(m_\psi(\cdot, t - s))(x, \mathbf{z}).$$



*Proof.* Let  $(x, \mathbf{z}) \in \mathcal{S}_K$ . Showing that  $t \mapsto m_\psi(x, \mathbf{z}, t)$  is differentiable on  $\mathbb{R}_+$  and computing its derivative is sufficient, as the desired result follows by composition. Let  $t \geq 0$  and  $h > 0$ . The Markov property ensures that

$$m_\psi(x, \mathbf{z}, t + h) = \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{u \in \mathbb{G}(h), u \geq u_x(0)} m_\psi(x_u(h), Z(h), t) \right].$$

For  $i \geq 1$ , Let  $T_i$  be the time of the  $i$ -th jump of the population process. Then on the one hand, if  $T_1 > h$ , then the population at time  $h$  is identical to the population at time 0, and thus:

$$\begin{aligned} m_\psi(x, \mathbf{z}, t + h) &= \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{u \in \mathbb{G}(h), u \geq u_x(0)} m_\psi(x_u(h), Z(h), t) \mathbf{1}_{\{T_1 < h\}} \right] \\ &\quad + m_\psi(x, \mathbf{z}, t) \mathbb{P}_{\mathbf{z}}(T_1 < h). \end{aligned}$$

Similarly, on the event  $\{T_1 < h < T_2\}$ ,  $Z(h) = Z(T_1)$  whence

$$\mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{u \in \mathbb{G}(h), u \geq u_x(0)} m_\psi(x_u(h), Z(h), t) \mathbf{1}_{\{T_1 < h\}} \right] = a(h) + b(h),$$

where

$$\begin{aligned} a(h) &= \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{u \in \mathbb{G}(h), u \geq u_x(0)} m_\psi(x_u(T_1), Z(T_1), t) \mathbf{1}_{\{T_1 < h < T_2\}} \right], \\ b(h) &= \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{u \in \mathbb{G}(h), u \geq u_x(0)} m_\psi(x_u(h), Z(h), t) \mathbf{1}_{\{T_2 < h\}} \right]. \end{aligned}$$

As a consequence, we obtain that

$$m_\psi(x, \mathbf{z}, t + h) - m_\psi(x, \mathbf{z}, t) = A(h) + B(h), \quad (5.11)$$

with

$$A(h) = a(h) - m_\psi(x, \mathbf{z}, t) \mathbb{P}_{\mathbf{z}}(T_1 < h < T_2)$$

and

$$B(h) = b(h) - m_\psi(x, \mathbf{z}, t) \mathbb{P}_{\mathbf{z}}(T_2 < h).$$

Let us first focus on  $B(h)$ . For any  $t \geq 0$  and  $(y, \mathbf{v}) \in \mathcal{S}_K$ , it holds that  $m_\psi(y, \mathbf{v}, t) \leq K \|\psi\|_\infty$ . As  $\mathcal{S}_K$  is a finite set, it follows that there exists a constant  $c > 0$  such that

$$|B(h)| \leq c \mathbb{P}_{\mathbf{z}}(T_2 < h).$$

For  $\mathbf{v} \in \mathcal{Z}_K$ , let us write  $\Lambda(\mathbf{v}) = \sum_{y \in \mathcal{X}} \sum_{\mathbf{k} \in \mathcal{Z}_K} \mathbf{v}_y \tau_{\mathbf{k}}(y, \mathbf{v})$  for the total jump rate in a population whose type distribution is given by  $\mathbf{v}$ . In particular,  $\Lambda$  is bounded on  $\mathcal{Z}_K$ . Using the law of  $T_1$  given  $Z(0) = \mathbf{z}$  and the law of  $T_2 - T_1$  given  $T_1$  and  $Z(T_1)$ , we then obtain:

$$\mathbb{P}_{\mathbf{z}}(T_2 < h) = \int_0^h e^{-\Lambda(\mathbf{z})t_1} \sum_{\substack{y \in \mathcal{X} \\ \mathbf{k} \in \mathcal{Z}_K}} \mathbf{z}_y \tau_{\mathbf{k}}(y, \mathbf{z}) \int_0^{h-t_1} \Lambda(\mathbf{z} + \mathbf{k} - \mathbf{e}(y)) e^{-\Lambda(\mathbf{z} + \mathbf{k} - \mathbf{e}(y))t_2} dt_2 dt_1 \leq \frac{\|\Lambda\|_\infty^2}{2} h^2.$$

We deduce that

$$\frac{B(h)}{h} \xrightarrow{h \rightarrow 0^+} 0. \quad (5.12)$$

Let us now focus on  $A(h)$ . Proceeding in the same way, we have

$$\begin{aligned} A(h) &= \sum_{\substack{\mathbf{k} \in \mathcal{Z}_K \\ y \in \mathcal{X}}} \int_0^h e^{-\Lambda(\mathbf{z})t_1} e^{-\Lambda(\mathbf{z}+\mathbf{k}-\mathbf{e}(x))(h-t_1)} dt_1 \tau_{\mathbf{k}}(x, \mathbf{z}) \mathbf{k}_y (m_{\psi}(y, \mathbf{z} + \mathbf{k} - \mathbf{e}(x), t) - m_{\psi}(x, \mathbf{z}, t)) \\ &+ \sum_{\substack{\mathbf{k} \in \mathcal{Z}_K \\ y \in \mathcal{X}}} \int_0^h e^{-\Lambda(\mathbf{z})t_1} e^{-\Lambda(\mathbf{z}+\mathbf{k}-\mathbf{e}(y))(h-t_1)} dt_1 \tau_{\mathbf{k}}(y, \mathbf{z}) (m_{\psi}(x, \mathbf{z} + \mathbf{k} - \mathbf{e}(y), t) - m_{\psi}(x, \mathbf{z}, t)), \end{aligned}$$

from which it follows that

$$\begin{aligned} \frac{A(h)}{h} \xrightarrow{h \rightarrow 0^+} & \sum_{\mathbf{k} \in \mathcal{Z}_K} \tau_{\mathbf{k}}(x, \mathbf{z}) \sum_{y \in \mathcal{X}} \mathbf{k}_y (m_{\psi}(y, \mathbf{z} + \mathbf{k} - \mathbf{e}(x), t) - m_{\psi}(x, \mathbf{z}, t)) \\ &+ \sum_{\substack{y \in \mathcal{X} \\ \mathbf{k} \in \mathcal{Z}_K}} \tau_{\mathbf{k}}(y, \mathbf{z}) (m_{\psi}(x, \mathbf{z} + \mathbf{k} - \mathbf{e}(y), t) - m_{\psi}(x, \mathbf{z}, t)). \end{aligned} \quad (5.13)$$

As a consequence, right differentiability of  $t \mapsto m_{\psi}(x, \mathbf{z}, t)$  is established by Equations (5.11), (5.12) and (5.13), and its right derivative is given by the right-hand side of Equation (5.13). As this corresponds to a continuous function on  $\mathbb{R}_+$ , we deduce that  $t \mapsto m_{\psi}(x, \mathbf{z}, t)$  is differentiable on  $\mathbb{R}_+$  (see e.g. Corollary 1.2 of Chapter 2 in Pazy, 2012) and

$$\begin{aligned} \frac{d}{dt} m_{\psi}(x, \mathbf{z}, t) &= \sum_{\mathbf{k} \in \mathcal{Z}_K} \tau_{\mathbf{k}}(x, \mathbf{z}) \sum_{y \in \mathcal{X}} \mathbf{k}_y (m_{\psi}(y, \mathbf{z} + \mathbf{k} - \mathbf{e}(x), t) - m_{\psi}(x, \mathbf{z}, t)) \\ &+ \sum_{\substack{y \in \mathcal{X} \\ \mathbf{k} \in \mathcal{Z}_K}} \tau_{\mathbf{k}}(y, \mathbf{z}) (m_{\psi}(x, \mathbf{z} + \mathbf{k} - \mathbf{e}(y), t) - m_{\psi}(x, \mathbf{z}, t)). \end{aligned}$$

This concludes the proof.  $\square$

We are now ready to establish the desired result.

*Proof of Proposition 5.3.2.* The proof is decomposed in three steps, establishing (i) existence and (ii) uniqueness of the solution to Equation (5.7) by classical arguments, before (iii) characterizing the associated semi-group  $\mathcal{R}^{(t)}$ .

For ease of notation, throughout the proof, for  $0 \leq s \leq t$  we let

$$\mathbb{Y}^{(t)}(s) = (Y^{(t)}(s), \zeta^{(t)}(s)).$$

(i) *Existence.* First, notice that by assumption on  $\tau(x, \mathbf{z})$  for  $(x, \mathbf{z}) \in \mathcal{S}_K$  and continuity of  $m_{\psi}$ , both applications  $\rho_{y, \mathbf{k}}^{(t)}$  and  $\hat{\rho}_{\mathbf{k}}^{(t)}$  are bounded for any  $y \in \mathcal{X}$  and  $\mathbf{k} \in \mathcal{Z}_K$ . As a consequence, existence of at least one solution to Equation (5.7) is ensured, as the associated sequence of jump times  $(T_k)_{k \geq 0}$  cannot admit an accumulation point on  $\mathbb{R}_+$ .

(ii) *Uniqueness.* Subsequently, in order to establish uniqueness, let us show by induction that for any  $k \geq 0$  such that  $T_k \leq t$ ,  $(T_k, \mathbb{Y}^{(t)}(T_k))$  is entirely determined by  $(\mathbb{Y}^{(t)}(0), Q, \hat{Q})$ . As  $T_0 = 0$ , initialization of the induction argument is immediate. If the property holds for  $k \geq 1$ , then by construction,  $T_{k+1}$  only depends on  $(T_k, \mathbb{Y}^{(t)}(T_k), Q, \hat{Q})$ . Similarly, given  $T_{k+1}$  and the corresponding atoms  $A_{k+1}$  and  $\hat{A}_{k+1}$  of  $Q$  and  $\hat{Q}$ , it is clear that  $\mathbb{Y}^{(t)}(T_{k+1})$  is

fixed by  $(T_{k+1}, A_{k+1}, \widehat{A}_{k+1}, \mathbb{Y}^{(t)}(T_k))$ . The desired conclusion thus is a consequence of the induction hypothesis.

(ii) *Characterization of  $\mathcal{R}^{(t)}$* . In order to establish Equation (5.8), it is sufficient to show that for any non-negative function  $f$  on  $\mathcal{S}_K$  and  $(x, \mathbf{z}) \in \mathcal{S}_K$ , the function

$$\tau \mapsto \mathcal{R}_{s,\tau}^{(t)} f(x, \mathbf{z}) = \mathbb{E}[f(\mathbb{Y}^{(t)}(\tau)) | \mathbb{Y}^{(t)}(s) = (x, \mathbf{z})]$$

is right differentiable at  $\tau = s$ . Indeed, it then follows that Equation (5.8) holds with the operator  $\mathcal{A}_s^{(t)}$  defined for  $s \in [0, t)$  by

$$\forall f : \mathcal{S}_K \rightarrow \mathbb{R}_+ \quad \forall (x, \mathbf{z}) \in \mathcal{S}_K, \quad \mathcal{A}_s^{(t)} f(x, \mathbf{z}) = \lim_{h \rightarrow 0^+} \frac{1}{h} \left( \mathcal{R}_{s,s+h}^{(t)} f(x, \mathbf{z}) - f(x, \mathbf{z}) \right). \quad (5.14)$$

As we will see, computing the right-hand side of Equation (5.14) leads to Equation (5.9).

Let  $f : \mathcal{S}_K \rightarrow \mathbb{R}_+$  and  $(x, \mathbf{z}) \in \mathcal{S}_K$ . We introduce the following notations. For any  $(y, \mathbf{k}) \in \mathcal{S}_K$  such that  $\tau_{\mathbf{k}}(x, \mathbf{z}) > 0$ ,

$$\mathfrak{d}_{y,\mathbf{k}} f(x, \mathbf{z}) = f(y, \mathbf{z} + \mathbf{k} - \mathbf{e}(x)) - f(x, \mathbf{z}).$$

Further, for any  $y \in \mathcal{X}$  such that  $\mathbf{z}_y > 0$  and  $\mathbf{k} \in \mathcal{Z}_K$  such that  $\tau_{\mathbf{k}}(y, \mathbf{z}) > 0$ , let

$$\widehat{\mathfrak{d}}_{y,\mathbf{k}} f(x, \mathbf{z}) = f(x, \mathbf{z} + \mathbf{k} - \mathbf{e}(y)) - f(x, \mathbf{z}).$$

Equation (5.7) then ensures that, on the event  $\mathbb{Y}^{(t)}(s) = (x, \mathbf{z})$ , we have for any  $h \in [0, t - s]$ :

$$\begin{aligned} f(\mathbb{Y}^{(t)}(s+h)) - f(x, \mathbf{z}) &= \int_s^{s+h} \int_E \mathbf{1}_{\{\theta \leq \rho_{y,\mathbf{k}}^{(t)}(r, \mathbb{Y}^{(t)}(r-))\}} \mathfrak{d}_{y,\mathbf{k}} f(\mathbb{Y}^{(t)}(r-)) Q(dr, d\theta, n(dy, d\mathbf{k})) \\ &+ \int_s^{s+h} \int_E \mathbf{1}_{\{\theta \leq (\zeta_y^{(t)}(r-) - \mathbf{1}_{\{Y^{(t)}(r-)=y\}}) \widehat{\rho}_{\mathbf{k}}^{(t)}(r, y, \mathbb{Y}^{(t)}(r-))\}} \widehat{\mathfrak{d}}_{y,\mathbf{k}} f(\mathbb{Y}^{(t)}(r-)) \widehat{Q}(dr, d\theta, n(dy, d\mathbf{k})). \end{aligned}$$

Notice that, for instance,

$$\begin{aligned} \mathbb{E} \left[ \int_s^{s+h} \int_E \mathbf{1}_{\{\theta \leq \rho_{y,\mathbf{k}}^{(t)}(r, \mathbb{Y}^{(t)}(r-))\}} \mathfrak{d}_{y,\mathbf{k}} f(\mathbb{Y}^{(t)}(r-)) Q(dr, d\theta, n(dy, d\mathbf{k})) \middle| \mathbb{Y}^{(t)}(s) = (x, \mathbf{z}) \right] \\ = \mathbb{E} \left[ \int_s^{s+h} \sum_{(y,\mathbf{k}) \in \mathcal{S}_K} \rho_{y,\mathbf{k}}^{(t)}(r, \mathbb{Y}^{(t)}(r)) \mathfrak{d}_{y,\mathbf{k}} f(\mathbb{Y}^{(t)}(r)) dr \middle| \mathbb{Y}^{(t)}(s) = (x, \mathbf{z}) \right]. \end{aligned}$$

On the one hand, almost surely,

$$\lim_{h \rightarrow 0^+} \frac{1}{h} \int_s^{s+h} \sum_{(y,\mathbf{k}) \in \mathcal{S}_K} \rho_{y,\mathbf{k}}^{(t)}(r, \mathbb{Y}^{(t)}(r)) \mathfrak{d}_{y,\mathbf{k}} f(\mathbb{Y}^{(t)}(r)) dr = \sum_{(y,\mathbf{k}) \in \mathcal{S}_K} \rho_{y,\mathbf{k}}^{(t)}(s, \mathbb{Y}^{(t)}(s)) \mathfrak{d}_{y,\mathbf{k}} f(\mathbb{Y}^{(t)}(s)).$$

On the other hand, as mentioned at the beginning of the proof,

$$\|\rho^{(t)}\|_\infty = \max_{s \in [0,t], (y,\mathbf{k}) \in \mathcal{S}, (x,\mathbf{z}) \in \mathcal{S}_K} \rho_{y,\mathbf{k}}^{(t)}(s, x, \mathbf{z}) < \infty.$$

Further, as  $\mathcal{S}_K$  is a finite set,  $\|f\|_\infty = \max_{(x,\mathbf{z}) \in \mathcal{S}_K} f(x, \mathbf{z}) < \infty$ . Thus, for any  $h \in [0, t - s]$ ,

$$\frac{1}{h} \int_s^{s+h} \left| \sum_{(y,\mathbf{k}) \in \mathcal{S}_K} \rho_{y,\mathbf{k}}^{(t)}(r, \mathbb{Y}^{(t)}(r)) \mathfrak{d}_{y,\mathbf{k}} f(\mathbb{Y}^{(t)}(r)) \right| dr \leq 2 \text{Card}(\mathcal{S}_K) \|\rho^{(t)}\|_\infty \|f\|_\infty < \infty.$$

Taken together, we obtain by dominated convergence:

$$\begin{aligned} \frac{1}{h} \mathbb{E} \left[ \int_s^{s+h} \int_E \mathbf{1}_{\{\theta \leq \rho_{y,\mathbf{k}}^{(t)}(r, \mathbb{Y}^{(t)}(r-))\}} \mathfrak{d}_{y,\mathbf{k}} f(\mathbb{Y}^{(t)}(r-)) Q(dr, d\theta, n(dy, d\mathbf{k})) \mid \mathbb{Y}^{(t)}(s) = (x, \mathbf{z}) \right] \\ \xrightarrow{h \rightarrow 0+} \sum_{(y,\mathbf{k}) \in \mathcal{S}_K} \rho_{y,\mathbf{k}}^{(t)}(s, x, \mathbf{z}) \mathfrak{d}_{y,\mathbf{k}} f(x, \mathbf{z}). \end{aligned}$$

The other terms arising on the right-hand side of Equation (5.14) can be treated analogously. This leads to the desired result.  $\square$

### Proof of the many-to-one formula

We are now ready to turn to the proof of Theorem 5.3.3, which comprises several steps. Let  $t \geq 0$ , and start by introducing the time-inhomogeneous semi-group of interest  $\mathcal{P}^{(t)} = (\mathcal{P}_{r,s}^{(t)}, r \leq s \leq t)$  through its action on applications  $f : \mathcal{S}_K \rightarrow \mathbb{R}_+$ . For  $s \leq t$ ,  $u_x(s)$  will designate a chosen individual of type  $x$  in  $\mathbb{G}(s)$ , if it exists. For any  $(x, \mathbf{z}) \in \mathcal{S}_K$  and  $0 \leq r \leq s \leq t$ ,

$$\mathcal{P}_{r,s}^{(t)} f(x, \mathbf{z}) = m_\psi(x, \mathbf{z}, t - r)^{-1} \mathbb{E} \left[ \sum_{\substack{u \in \mathbb{G}(t) \\ u \geq u_x(r)}} \psi(x_u(t), Z(t)) f(x_u(s), Z(s)) \mid X(r) = \mathfrak{X}(x, \mathbf{z}) \right]. \quad (5.15)$$

**Lemma 5.3.5.**  $(\mathcal{P}_{r,s}^{(t)}, r \leq s \leq t)$  defines a conservative, time-inhomogeneous semi-group acting on the set of functions  $\{f : \mathcal{S}_K \rightarrow \mathbb{R}_+\}$ .

*Proof.* The conservativity of  $\mathcal{P}^{(t)}$  follows directly from Equation (5.15) applied to  $f \equiv 1$ , which shows that  $\mathcal{P}^{(t)} 1 \equiv 1$ .

Let us now turn to the inhomogeneous semi-group property. Let  $r \leq \tau \leq t$ , and consider  $f : \mathcal{S}_K \rightarrow \mathbb{R}_+$  and  $(x, \mathbf{z}) \in \mathcal{S}_K$ . Throughout the proof, we let  $\mathfrak{X}_0 = \mathfrak{X}(x, \mathbf{z})$ . By definition of the semi-group,

$$\begin{aligned} \mathcal{P}_{r,s}^{(t)} f(x, \mathbf{z}) &= m_\psi(x, \mathbf{z}, t - r)^{-1} \mathbb{E} \left[ \sum_{\substack{u \in \mathbb{G}(t) \\ u \geq u_x(r)}} \psi(x_u(t), Z(t)) f(x_u(s), Z(s)) \mid X(r) = \mathfrak{X}_0 \right] \\ &= m_\psi(x, \mathbf{z}, t - r)^{-1} \mathbb{E} \left[ \sum_{\substack{v \in \mathbb{G}(\tau) \\ v \geq u_x(r)}} \sum_{\substack{u \in \mathbb{G}(t) \\ u \geq v}} \psi(x_u(t), Z(t)) f(x_u(s), Z(s)) \mid X(r) = \mathfrak{X}_0 \right] \\ &= m_\psi(x, \mathbf{z}, t - r)^{-1} \mathbb{E} \left[ \sum_{\substack{v \in \mathbb{G}(\tau) \\ v \geq u_x(r)}} g(x_v(\tau), Z(\tau)) \mid X(r) = \mathfrak{X}_0 \right], \end{aligned} \quad (5.16)$$

where we define the function  $g : \mathcal{S}_K \rightarrow \mathbb{R}_+$  by

$$g(x, \mathbf{z}) = \mathbb{E} \left[ \sum_{\substack{u \in \mathbb{G}(t) \\ u \geq u_x(\tau)}} \psi(x_u(t), Z(t)) f(x_u(s), Z(s)) \mid X(\tau) = \mathfrak{X}(x, \mathbf{z}) \right].$$

Notice that, for any measurable function  $G : \mathbb{D}([0, \tau], \mathcal{S}_K) \rightarrow \mathbb{R}_+$ ,

$$\begin{aligned}
& \mathbb{E} \left[ \sum_{\substack{u \in \mathbb{G}(t) \\ u \geq u_x(r)}} \psi(x_u(t), Z(t)) m_\psi(x_u(\tau), Z(\tau), t - \tau)^{-1} G((x_u(s), Z(s))_{s \leq \tau}) \middle| X(r) = \mathfrak{x}_0 \right] \\
&= \mathbb{E} \left[ \sum_{\substack{v \in \mathbb{G}(\tau) \\ v \geq u_x(r)}} \mathbb{E} \left[ \sum_{\substack{u \in \mathbb{G}(t) \\ u \geq v}} \psi(x_u(t), Z(t)) \middle| X(\tau) \right] m_\psi(x_v(\tau), Z(\tau), t - \tau)^{-1} G((x_v(s), Z(s))_{s \leq \tau}) \middle| X(r) = \mathfrak{x}_0 \right] \\
&= \mathbb{E} \left[ \sum_{\substack{v \in \mathbb{G}(\tau) \\ v \geq u_x(r)}} G((x_v(s), Z(s))_{s \leq \tau}) \middle| X(r) = \mathfrak{x}_0 \right].
\end{aligned} \tag{5.17}$$

Applying this equality to  $G((x_v(s), Z(s))_{s \leq \tau}) = g(x(\tau), Z(\tau))$  finally yields the desired semi-group property:

$$\mathcal{P}_{r,s}^{(t)} f(x, \mathbf{z}) = \mathcal{P}_{r,\tau}^{(t)} \mathcal{P}_{\tau,s}^{(t)} f(x, \mathbf{z}).$$

This concludes the proof.  $\square$

Let us now compute the generator of  $(\mathcal{P}_{r,s}^{(t)}, r \leq s \leq t)$ .

**Lemma 5.3.6.** *Let  $t \geq 0$ . The generator of the semi-group  $(\mathcal{P}_{r,s}^{(t)}, r \leq s \leq t)$  is  $(\mathcal{A}_s^{(t)}, s \leq t)$ .*

*Proof.* Consider  $f : \mathcal{S}_K \rightarrow \mathbb{R}_+$ . Let  $(x, \mathbf{z}) \in \mathcal{S}_K$  and  $t \geq 0$ . For any  $0 \leq s \leq t$  and  $h > 0$  such that  $s + h \leq t$ , it follows from Equation (5.17) and the Markov property that

$$\mathcal{P}_{s,s+h}^{(t)} f(x, \mathbf{z}) = m_\psi(x, \mathbf{z}, t - s)^{-1} \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{\substack{u \in \mathbb{G}(h) \\ u \geq u_x(0)}} m_\psi(x_u(h), Z(h), t - (s + h)) f(x_u(h), Z(h)) \right].$$

Using Lemma 5.3.4 as well as the fact that  $\mathcal{S}_K$  is a finite set, we obtain the following Taylor expansion:

$$\begin{aligned}
m_\psi((x, \mathbf{z}), t - s) \mathcal{P}_{s,s+h}^{(t)} f(x, \mathbf{z}) &= \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{\substack{u \in \mathbb{G}(h) \\ u \geq u_x(0)}} m_\psi(x_u(h), Z(h), t - s) f(x_u(h), Z(h)) \right] \\
&\quad + h \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{\substack{u \in \mathbb{G}(h) \\ u \geq u_x(0)}} \partial_s m_\psi(x_u(h), Z(h), t - s) f(x_u(h), Z(h)) \right] + o(h).
\end{aligned}$$

As a consequence,

$$\begin{aligned}
m_\psi(x, \mathbf{z}, t - s) \frac{\mathcal{P}_{s,s+h}^{(t)} f(x, \mathbf{z}) - f(x, \mathbf{z})}{h} &= \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{\substack{u \in \mathbb{G}(h) \\ u \geq u_x(0)}} \partial_s m_\psi(x_u(h), Z(h), t - s) f(x_u(h), Z(h)) \right] \\
&\quad + h^{-1} \left( \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{\substack{u \in \mathbb{G}(h) \\ u \geq u_x(0)}} m_\psi(x_u(h), Z(h), t - s) f(x_u(h), Z(h)) \right] - m(x, \mathbf{z}, t - s) f(x, \mathbf{z}) \right) + \epsilon(h),
\end{aligned}$$

where  $\epsilon(h)$  is such that  $\lim_{h \rightarrow 0^+} \epsilon(h) = 0$ . We thus obtain that

$$\lim_{h \rightarrow 0^+} \frac{\mathcal{P}_{s,s+h}^{(t)} f(x, \mathbf{z}) - f(x, \mathbf{z})}{h} = m_\psi(x, \mathbf{z}, t - s)^{-1} (\mathcal{G}(m(\cdot, t - s) f(\cdot))(x, \mathbf{z}) + \partial_s m_\psi(x, \mathbf{z}, t - s)),$$

where we recall that  $\mathcal{G}$  is defined by (5.2). Lemma 5.3.4 yields the desired result.  $\square$

We finally are ready to establish Theorem 5.3.3. The proof follows the lines of Marguet, 2019, and is detailed here for the sake of completeness.

*Proof of Theorem 5.3.3.* Throughout this proof, for readability, we will make use of the following notations. On the one hand, for  $t \geq 0$  and  $u$  such that there exists  $v \in \mathbb{G}(t)$  satisfying  $u \geq v$ , let

$$\mathbb{X}_u(t) = (x_u(t), Z(t)).$$

Similarly, for  $0 \leq s \leq t$ , we let

$$\mathbb{Y}^{(t)}(s) = (Y^{(t)}(s), \zeta^{(t)}(s))$$

and for  $\mathbb{X} = (x, \mathbf{z}) \in \mathcal{S}_K$ , we let  $m_\psi(\mathbb{X}, t) = m_\psi(x, \mathbf{z}, t)$ .

Let us start by showing that Equation (5.10) holds for

$$F((x(s), \mathbf{z}(s))_{s \leq t}) = \prod_{j=1}^k f_j(x(s_j), \mathbf{z}(s_j)),$$

where  $k \geq 1$ ,  $0 \leq s_1 \leq \dots \leq s_k \leq t$  and  $f_1, \dots, f_k : \mathcal{S}_K \rightarrow \mathbb{R}_+$ .

This part of the proof proceeds by induction. For  $k \geq 1$ , let  $H_k$  be the property that for any  $0 \leq s_1 \leq \dots \leq s_k \leq t$  and  $f_1, \dots, f_k : \mathcal{S}_K \rightarrow \mathbb{R}_+$ ,

$$\mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{u \in \mathbb{G}(t), u \geq u_x(0)} \psi(\mathbb{X}_u(t)) \prod_{j=1}^k f_j(\mathbb{X}_u(s_j)) \right] = m_\psi(x, \mathbf{z}, t) \mathbb{E}_{(x, \mathbf{z})} \left[ \prod_{j=1}^k f_j(\mathbb{Y}^{(t)}(s_j)) \right].$$

Let us turn our attention to the initialization step. As  $\mathcal{S}_K$  is a finite set, a semi-group acting on non-negative functions on  $\mathcal{S}_K$  is uniquely characterized by its generator. Thus Lemma 5.3.6 implies that the semi-groups  $\mathcal{P}^{(t)}$  and  $\mathcal{R}^{(t)}$  are identical. Hence for any  $s \in [0, t]$  and  $f : \mathcal{S}_K \rightarrow \mathbb{R}_+$ , Equation (5.15) becomes

$$\mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{u \in \mathbb{G}(t), u \geq u_x(0)} \psi(\mathbb{X}_u(t)) f(\mathbb{X}_u(s)) \right] = m_\psi(x, \mathbf{z}, t) \mathcal{R}_{0,s}^{(t)} f(x, \mathbf{z}).$$

This exactly corresponds to  $H_1$  by definition of  $\mathcal{R}^{(t)}$ .

Suppose now that  $H_{k-1}$  is true for  $k > 1$ , and let us show that  $H_k$  follows. Consider functions  $f_1, \dots, f_k : \mathcal{S}_K \rightarrow \mathbb{R}_+$  and  $0 \leq s_1 \leq \dots \leq s_k \leq t$ . Notice that

$$\begin{aligned} \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{\substack{u \in \mathbb{G}(t) \\ u \geq u_x(0)}} \psi(\mathbb{X}_u(t)) \prod_{j=1}^k f_j(\mathbb{X}_u(s_j)) \right] = \\ \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{\substack{u \in \mathbb{G}(s_{k-1}) \\ u \geq u_x(0)}} \prod_{j=1}^{k-1} f_j(\mathbb{X}_u(s_j)) \mathbb{E} \left[ \sum_{\substack{v \in \mathbb{G}(t) \\ v \geq u_{\mathfrak{X}(u(s_{k-1}))}}} \psi(\mathbb{X}_v(t)) f_k(\mathbb{X}_v(s_k)) \middle| X(s_{k-1}) = \mathfrak{X}(u(s_{k-1})) \right] \right]. \end{aligned}$$

Using the Markov property and  $H_1$  leads to:

$$\begin{aligned} \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{\substack{u \in \mathbb{G}(t) \\ u \geq u_x(0)}} \psi(\mathbb{X}_u(t)) \prod_{j=1}^k f_j(\mathbb{X}_u(s_j)) \right] = \\ \mathbb{E}_{\mathfrak{X}(x, \mathbf{z})} \left[ \sum_{\substack{u \in \mathbb{G}(s_{k-1}) \\ u \geq u_x(0)}} m_\psi(\mathbb{X}_u(s_{k-1}), t - s_{k-1}) \prod_{j=1}^{k-1} f_j(\mathbb{X}_u(s_j)) \mathbb{E}[f_k(\mathbb{Y}^{(t)}(s_k)) | \mathbb{Y}^{(t)}(s_{k-1}) = \mathbb{X}_u(s_{k-1})] \right]. \end{aligned}$$

Equation (5.17) allows to rewrite this as:

$$\begin{aligned} \mathbb{E}_{\mathcal{X}(x, \mathbf{z})} \left[ \sum_{\substack{u \in \mathbb{G}(t) \\ u \geq u_x(0)}} \psi(\mathbb{X}_u(t)) \prod_{j=1}^k f_j(\mathbb{X}_u(s_j)) \right] = \\ \mathbb{E}_{\mathcal{X}(x, \mathbf{z})} \left[ \sum_{\substack{u \in \mathbb{G}(t) \\ u \geq u_x(0)}} \psi(\mathbb{X}_u(t)) \prod_{j=1}^{k-1} f_j(\mathbb{X}_u(s_j)) \mathbb{E}[f_k(\mathbb{Y}^{(t)}(s_k)) | \mathbb{Y}^{(t)}(s_{k-1}) = \mathbb{X}_u(s_{k-1})] \right]. \end{aligned}$$

Finally,  $H_{k-1}$  yields:

$$\begin{aligned} \mathbb{E}_{\mathcal{X}(x, \mathbf{z})} \left[ \sum_{\substack{u \in \mathbb{G}(t) \\ u \geq u_x(0)}} \psi(\mathbb{X}_u(t)) \prod_{j=1}^k f_j(\mathbb{X}_u(s_j)) \right] \\ = m_\psi(x, \mathbf{z}, t) \mathbb{E}_{(x, \mathbf{z})} \left[ \prod_{j=1}^{k-1} f_j(\mathbb{Y}^{(t)}(s_j)) \mathbb{E}[f_k(\mathbb{Y}^{(t)}(s_k)) | \mathbb{Y}^{(t)}(s_{k-1})] \right] \\ = m_\psi(x, \mathbf{z}, t) \mathbb{E}_{(x, \mathbf{z})} \left[ \prod_{j=1}^k f_j(\mathbb{Y}^{(t)}(s_j)) \right]. \end{aligned}$$

This concludes the induction argument.

In order to obtain the desired result, we will reason using the monotone class theorem. Let us introduce the set

$$I = \left\{ \bigcap_{j=1}^k \{x \in \mathbb{D}([0, t], \mathcal{S}_K) : x(s_j) \in B_j\}, k \in \mathbb{N}, s_j \in [0, t], B_j \in \mathcal{P}(\mathcal{S}_K) \right\}$$

where  $\mathcal{P}(\mathcal{S}_K)$  is the set of subsets of  $\mathcal{S}_K$ . The set  $I$  is a  $\pi$ -system, which induces the Borel  $\sigma$ -algebra  $\mathcal{B}(\mathbb{D}([0, t], \mathcal{S}_K))$  on the Skorokhod space  $\mathbb{D}([0, t], \mathcal{S}_K)$  (Billingsley, 1999, Theorem 12.5). Further, define

$$M = \{\mathbf{B} \in \mathcal{B}(\mathbb{D}([0, t], \mathcal{S}_K)) : \text{Equation (5.10) is satisfied for } F = \mathbf{1}_{\mathbf{B}}\}.$$

$M$  is a monotone class which contains  $I$  according to our induction argument. It thus follows from the monotone class theorem that  $M = \mathcal{B}(\mathbb{D}([0, t], \mathcal{S}_K))$ . In other words, for any  $\mathbf{B} \in \mathcal{B}(\mathbb{D}([0, t], \mathcal{S}_K))$ , Equation (5.10) is satisfied for  $F = \mathbf{1}_{\mathbf{B}}$ . As a consequence, Equation (5.10) holds for any positive measurable function  $F : \mathbb{D}([0, t], \mathcal{S}_K) \rightarrow \mathbb{R}_+$  as there exists an increasing sequence of simple functions converging pointwise to  $F$ , from which the result follows by monotone convergence.  $\square$

## 5.4 Quantification of the $\psi$ -spine large population approximation

Throughout this section, we will consider the large population  $\psi$ -spine construction introduced by Bansaye, 2024, as summarized in Section 5.2.2 whose assumptions and notations carry on to the following results. In particular,  $\psi$  is a positive, continuous function on  $\mathcal{X} \times [0, 1]^d$  which is continuously differentiable on  $\mathcal{X} \times (0, 1)^d$ .

In a population of size  $K$ , let  $(Y^K, \zeta^K)$  be the corresponding  $\psi$ -spine construction where the population composition is normalized by  $K$ . We suppose that there exists  $(y_0, \mathbf{z}_0) \in \mathcal{X} \times (0, 1)^d$  such that  $Y^K(0) = y_0$  and  $\zeta^K(0) = \lfloor K\mathbf{z}_0 \rfloor / K$ , for all  $K \geq 1$ . Recall that  $\zeta^K$  converges to the unique solution  $z$  of the differential equation (5.5), while  $Y^K$  tends towards an inhomogeneous continuous-time Markov chain  $\Upsilon$ . In order to quantify the convergence to this large population limit, we will make use of a specific construction which couples the sequence of finite population  $\psi$ -spines with their large population limit.

### 5.4.1 Coupling the $\psi$ -spine with its large population limit

For the purpose of this section, we let  $Y^K(t) \in \{\mathbf{e}(x) : x \in \mathcal{X}\}$  for any  $t \geq 0$ , where  $Y^K(t) = \mathbf{e}(x)$  means that the spine is of type  $x$ . As previously, we will identify  $\{\mathbf{e}(x) : x \in \mathcal{X}\}$  and  $\mathcal{X}$  through their natural bijection. Also, recall that  $\mathcal{S} = \{(x, \mathbf{z}) \in \mathcal{X} \times [0, 1]^d : \mathbf{z}_x > 0\}$ .

Let  $\mathcal{J} = \cup_{x \in \mathcal{X}} \{(x, \mathbf{k}) : \mathbf{k} \in \mathcal{J}_x\}$  be the set of possible descendance of any individual, and  $\overline{\mathcal{J}} = \{(y, \mathbf{k}) \in \mathcal{X} \times \mathbb{N}^d : \mathbf{k} \in \cup_{x \in \mathcal{X}} \mathcal{J}_x, \mathbf{k}_y > 0\}$  the set of possible descendance for the spine, *i.e.* the type of the child who will become the spine along with the types of all the spine's children. For  $(x, \mathbf{z}) \in \mathcal{S}$  and  $(y, \mathbf{k}) \in \overline{\mathcal{J}}$ , we introduce the rate  $\rho_{y, \mathbf{k}}^K(x, \mathbf{z})$  at which the spine  $Y^K$  leaves descendance  $\mathbf{k}$  and becomes of type  $y$ , given that it is in state  $x$  in a population of type  $\mathbf{z} \in [0, 1]^d$ :

$$\rho_{y, \mathbf{k}}^K(x, \mathbf{z}) = \tau_{\mathbf{k}}(x, \mathbf{z}) \mathbf{k}_y \frac{\psi(y, \mathbf{z} + (\mathbf{k} - \mathbf{e}(x))/K)}{\psi(x, \mathbf{z})}. \quad (5.18)$$

Similarly, let  $\hat{\rho}_{\mathbf{k}}^K(x, y, \mathbf{z})$  be the rate at which an individual of type  $x$  other than the spine is replaced by  $\mathbf{k}$ , given that the population state is  $\mathbf{z}$  and the spine is of type  $y$ :

$$\hat{\rho}_{\mathbf{k}}^K(x, y, \mathbf{z}) = \tau_{\mathbf{k}}(x, \mathbf{z}) \frac{\psi(y, \mathbf{z} + (\mathbf{k} - \mathbf{e}(x))/K)}{\psi(y, \mathbf{z})}.$$

Define  $E = \mathbb{R}_+ \times \overline{\mathcal{J}}$  and  $Q$  a Poisson point process taking values in  $\mathbb{R}_+ \times E$ , of density  $ds \otimes d\theta \otimes n(dy, d\mathbf{k})$ , where  $ds, d\theta$  represent the Lebesgue measure on  $\mathbb{R}_+$  and  $n(dy, d\mathbf{k})$  the counting measure on  $\overline{\mathcal{J}}$ . Further, let  $\tilde{E} = \mathbb{R}_+ \times \mathcal{J}$  and consider a Poisson point process  $\tilde{Q}$  on  $\mathbb{R}_+ \times \tilde{E}$ , with density  $ds \otimes d\theta \otimes n(dx, d\mathbf{k})$  where with slight abuse of notation,  $n$  represents the counting measure on  $\mathcal{J}$ .

Then for  $K \geq 1$ , the process  $(Y^K, \zeta^K)$  can be defined as the unique strong solution in the Skorokhod space  $\mathbb{D}(\mathbb{R}_+, \mathcal{S})$  of the following system of stochastic differential equations. For any  $t \geq 0$ ,

$$\begin{aligned} Y^K(t) &= y_0 + \int_0^t \int_E \mathbf{1}_{\{\theta \leq \rho_{y, \mathbf{k}}^K(Y^K(s-), \zeta^K(s-))\}} (\mathbf{e}(y) - Y^K(s-)) Q(ds, d\theta, n(dy, d\mathbf{k})), \\ \zeta^K(t) &= \frac{\lfloor K\mathbf{z}_0 \rfloor}{K} + \frac{1}{K} \int_0^t \int_E \mathbf{1}_{\{\theta \leq \rho_{y, \mathbf{k}}^K(Y^K(s-), \zeta^K(s-))\}} (\mathbf{k} - Y^K(s-)) Q(ds, d\theta, n(dy, d\mathbf{k})) \\ &\quad + \frac{1}{K} \int_0^t \int_{\tilde{E}} \mathbf{1}_{\{\theta \leq (\zeta_x^K(s-))^{-1} \mathbf{1}_{\{Y^K(s-)=x\}} \hat{\rho}_{\mathbf{k}}^K(x, Y^K(s-), \zeta^K(s-))\}} (\mathbf{k} - \mathbf{e}(x)) \tilde{Q}(ds, d\theta, n(dx, d\mathbf{k})). \end{aligned} \quad (5.19)$$

In order to couple the  $\psi$ -spine with its large population limit, we will construct  $\Upsilon$  using the same Poisson point process  $Q$  as for  $Y^K$ . Define  $\rho_{y, \mathbf{k}}(x, \mathbf{z})$  as the rate at which the



large population spine leaves descendance  $\mathbf{k}$  and becomes of type  $y$ , given that it is of type  $x$  in a population of composition  $\mathbf{z}$ :

$$\rho_{y,\mathbf{k}}(x, \mathbf{z}) = \tau_{\mathbf{k}}(x, \mathbf{z}) \mathbf{k}_y \frac{\psi(y, \mathbf{z})}{\psi(x, \mathbf{z})}. \quad (5.20)$$

Then  $\Upsilon$  is characterized as the unique strong solution in  $\mathbb{D}(\mathbb{R}_+, \mathcal{X})$  of the following stochastic differential equation. For any  $t \geq 0$ ,

$$\Upsilon(t) = y_0 + \int_0^t \int_E \mathbf{1}_{\{\theta \leq \rho_{y,\mathbf{k}}(\Upsilon(s-), z(s))\}} (\mathbf{e}(y) - \Upsilon(s-)) Q(ds, d\theta, n(dy, d\mathbf{k})). \quad (5.21)$$

**Remark 5.4.1.** *Existence and uniqueness of the solutions to Equations (5.19) and (5.21) can be established following the lines of the Proof of Proposition 5.3.2. Indeed, the continuity and positivity Assumption on  $\psi$ , the continuity of the applications  $\tau_{\mathbf{k}}$  and the finiteness of  $\mathcal{J}$  ensure that the jump rates are bounded. This allows to derive existence, and uniqueness is again obtained by an induction argument.*

#### 5.4.2 Quantifying the approximation error

Let us now turn towards quantifying the large population approximation error. We will work under the following regularity assumption.

**Assumption 5.4.2.** *For every  $x \in \mathcal{X}$  and  $\mathbf{k} \in \mathcal{J}_x$ , there exists a positive constant  $L_x$  such that the application  $\tau_{\mathbf{k}}(x, \cdot)$  is  $L_x$ -Lipschitz continuous on  $[0, 1]^d$ .*

Recall that  $\psi$  is continuously differentiable on  $\mathcal{X} \times (0, 1)^d$  and continuous on  $\mathcal{X} \times [0, 1]^d$ . Hence for every  $x \in \mathcal{X}$ ,  $\psi(x, \cdot)$  and  $\psi(x, \cdot)^{-1}$  are Lipschitz continuous with respective Lipschitz constant  $M_x$  and  $N_x$ . Let  $M = \max_{x \in \mathcal{X}} M_x$ ,  $N = \max_{x \in \mathcal{X}} N_x$  and  $L = \max_{x \in \mathcal{X}} L_x$ . As the set  $\mathcal{J}$  is finite, it holds that:

$$\begin{aligned} V &= \max_{\mathbf{k} \in \mathcal{J}} (\|\mathbf{k}\|_1 \|\tau_{\mathbf{k}}\|_{\infty}) (M \|\psi^{-1}\|_{\infty} + N \|\psi\|_{\infty}) + \|\psi^{-1}\|_{\infty} \max_{\mathbf{k} \in \mathcal{J}} \|\mathbf{k}\|_1 L < \infty, \\ c &= M \|\psi^{-1}\|_{\infty} \max_{\mathbf{k} \in \mathcal{J}} (\|\mathbf{k}\|_1 (\|\mathbf{k}\|_1 + 1) \|\tau_{\mathbf{k}}\|_{\infty}) < \infty. \end{aligned}$$

Hence, Equations (5.18) and (5.20) as well as Assumption 5.4.2 ensure that for every  $K \geq 1$ , for every  $x \in \mathcal{X}$  and  $\mathbf{z}_1, \mathbf{z}_2 \in [0, 1]^d$ ,

$$|\rho_{y,\mathbf{k}}^K(x, \mathbf{z}_1) - \rho_{y,\mathbf{k}}(x, \mathbf{z}_2)| \leq V \|\mathbf{z}_1 - \mathbf{z}_2\|_1 + \frac{c}{K}. \quad (5.22)$$

With this in mind, we make an assumption on the convergence of  $\zeta^K$  to  $z$ .

**Assumption 5.4.3.** *There exist non-negative sequences  $(t_K)_{K \geq 1}$ ,  $(\varepsilon_K)_{K \geq 1}$  and  $(\alpha_K)_{K \geq 1}$  such that, for every  $K \geq 1$ ,*

$$\mathbb{P} \left( \sup_{s \in [0, t_K]} \|\zeta^K(s) - z(s)\|_1 \geq \varepsilon_K \right) \leq \alpha_K.$$

In particular, Assumption 5.4.3 implies that we already know the precision of the large population approximation at the population scale, which we consider to be reasonable. Indeed, for many density-dependent Markov jump processes describing population dynamics, the divergence of the finite-population process from its large population limit is well studied. Classical regimes include diffusion approximations (Ethier and Kurtz, 1986, Chapter 11), moderate and large deviations (Britton and Pardoux, 2019b; Pardoux, 2020, for applications to epidemic models). This will be discussed in more detail in Section 5.5.

As a consequence, it remains to focus on understanding how well  $\Upsilon$  approximates  $\zeta^K$ . The idea is to control the probability of the coupled spines to differ before time  $t_K$  thanks to the control of the fluctuations of  $\zeta^K$  around  $z$  on the same time interval. This will allow to establish the following result.

**Theorem 5.4.4.** *Under Assumptions 5.4.2 and 5.4.3, there exists  $C > 0$  such that for every  $K \geq 1$ ,*

$$\mathbb{P}(\exists t \leq t_K : Y^K(t) \neq \Upsilon(t)) \leq \alpha_K + C(\varepsilon_K + K^{-1})t_K.$$

A particularly nice setting would be to consider

$$\lim_{K \rightarrow \infty} t_K/K = \lim_{K \rightarrow \infty} t_K \varepsilon_K = \lim_{K \rightarrow \infty} \alpha_K = 0, \quad (5.23)$$

as the probability of the coupling being exact on  $[0, t_K]$  then converges to one as  $K$  grows large. Again, this is expected to be achievable, given the aforementioned results on the fluctuations of density-dependent population processes around their deterministic large population limits. We refer to Section 5.5 for detail.

*Proof.* We are interested in the first instant  $T_K$  at which  $\Upsilon$  differs from  $Y^K$ :

$$T_K = \inf\{t \geq 0 : Y^K(t) \neq \Upsilon(t)\}.$$

Let  $K \geq 1$ . For two sets  $A$  and  $B$ , we let  $A \Delta B$  designate their symmetric difference. Notice that, by the coupling of  $\Upsilon$  and  $Y^K$  through Equations (5.19) and (5.21),

$$\begin{aligned} \{T_K \geq t_K\} &\supseteq \left\{ T_K \geq t_K, \int_0^{t_K} \int_E \mathbf{1}_{\{\theta \leq \rho_{y,\mathbf{k}}^K(Y^K(s-), \zeta^K(s-)) \Delta \theta \leq \rho_{y,\mathbf{k}}(\Upsilon(s-), z(s))\}} Q(ds, d\theta, n(dy, d\mathbf{k})) = 0 \right\} \\ &\supseteq \left\{ T_K \geq t_K, \int_0^{t_K} \int_E \mathbf{1}_{\{\theta \leq \rho_{y,\mathbf{k}}^K(\Upsilon(s-), \zeta^K(s-)) \Delta \theta \leq \rho_{y,\mathbf{k}}(\Upsilon(s-), z(s))\}} Q(ds, d\theta, n(dy, d\mathbf{k})) = 0 \right\} \\ \{T_K \geq t_K\} &\supseteq \left\{ \int_0^{t_K} \int_E \mathbf{1}_{\{\theta \leq \rho_{y,\mathbf{k}}^K(\Upsilon(s-), \zeta^K(s-)) \Delta \theta \leq \rho_{y,\mathbf{k}}(\Upsilon(s-), z(s))\}} Q(ds, d\theta, n(dy, d\mathbf{k})) = 0 \right\}. \end{aligned}$$

Define the constant  $R = V \text{Card}(\mathcal{J})$  and for  $K \geq 1$ ,  $c_K = R\varepsilon_K + \text{Card}(\mathcal{J})c/K$ . We introduce the event

$$A_K = \left\{ \sup_{s \in [0, t_K]} \sum_{(y, \mathbf{k}) \in \mathcal{S}} |\rho_{y,\mathbf{k}}^K(\Upsilon(s), \zeta^K(s)) - \rho_{y,\mathbf{k}}(\Upsilon(s), z(s))| < c_K \right\}.$$

It follows that

$$\begin{aligned} \mathbb{P}(T_K \leq t_K) &\leq \mathbb{P}(A_K, \int_0^{t_K} \int_E \mathbf{1}_{\{\theta \leq \rho_{y,\mathbf{k}}^K(\Upsilon(s-), \zeta^K(s-)) \Delta \theta \leq \rho_{y,\mathbf{k}}(\Upsilon(s-), z(s))\}} Q(ds, d\theta, n(dy, d\mathbf{k})) \geq 1) \\ &\quad + \mathbb{P}(A_K^C). \end{aligned} \quad (5.24)$$

First, we may notice that Assumption 5.4.2 ensures through Equation (5.22) that

$$A_K^C \subseteq \{R \sup_{s \in [0, t_K]} \|\zeta^K(s) - z(s)\|_1 + \text{Card}(\mathcal{J}) \frac{c}{K} \geq c_K\},$$

from which we deduce:

$$\mathbb{P}(A_K^C) \leq \mathbb{P}(\sup_{s \in [0, t_K]} \|\zeta^K(s) - z(s)\|_1 \geq \varepsilon_K) \leq \alpha_K. \quad (5.25)$$

Second, on the event  $A_K$ , it holds that for any  $(\theta, y, \mathbf{k}) \in E$ ,

$$\begin{aligned} \{\theta \leq \rho_{y, \mathbf{k}}^K(\Upsilon(s-), \zeta^K(s-)) \Delta \{\theta \leq \rho_{y, \mathbf{k}}(\Upsilon(s-), z(s))\}\} \\ \subseteq \{\theta \in [\rho_{y, \mathbf{k}}(\Upsilon(s-), z(s)) - c_K, \rho_{y, \mathbf{k}}(\Upsilon(s-), z(s)) + c_K]\}. \end{aligned}$$

As a consequence,

$$\begin{aligned} \left\{ A_K, \int_0^{t_K} \int_E \mathbf{1}_{\{\theta \leq \rho_{y, \mathbf{k}}^K(\Upsilon(s-), \zeta^K(s-)) \Delta \{\theta \leq \rho_{y, \mathbf{k}}(\Upsilon(s-), z(s))\}\}} Q(ds, d\theta, n(dy, d\mathbf{k})) \geq 1 \right\} \\ \subseteq \left\{ \int_0^{t_K} \int_E \mathbf{1}_{\{\theta \in [\rho_{y, \mathbf{k}}(\Upsilon(s-), z(s)) - c_K, \rho_{y, \mathbf{k}}(\Upsilon(s-), z(s)) + c_K]\}} Q(ds, d\theta, n(dy, d\mathbf{k})) \geq 1 \right\}. \end{aligned}$$

Hence, Markov's inequality leads to

$$\begin{aligned} \mathbb{P}(A_K, \int_0^{t_K} \int_E \mathbf{1}_{\{\theta \leq \rho_{y, \mathbf{k}}^K(\Upsilon(s-), \zeta^K(s-)) \Delta \{\theta \leq \rho_{y, \mathbf{k}}(\Upsilon(s-), z(s))\}\}} Q(ds, d\theta, n(dy, d\mathbf{k})) \geq 1) \\ \leq \mathbb{E} \left[ \int_0^{t_K} \int_E \mathbf{1}_{\{\theta \in [\rho_{y, \mathbf{k}}(\Upsilon(s-), z(s)) - c_K, \rho_{y, \mathbf{k}}(\Upsilon(s-), z(s)) + c_K]\}} Q(ds, d\theta, n(dy, d\mathbf{k})) \right] \\ \leq \mathbb{E} \left[ \int_0^{t_K} \sum_{(y, \mathbf{k}) \in \mathcal{J}} \int_0^{+\infty} \mathbf{1}_{\{\theta \in [\rho_{y, \mathbf{k}}(\Upsilon(s-), z(s)) - c_K, \rho_{y, \mathbf{k}}(\Upsilon(s-), z(s)) + c_K]\}} Q(ds, d\theta, n(dy, d\mathbf{k})) \right] \\ \leq C(\varepsilon_K + K^{-1})t_K, \end{aligned} \quad (5.26)$$

with  $C = 2 \max(c, V) \text{Card}(\mathcal{J})^2$ . Thus, the conclusion follows by injecting Inequalities (5.25) and (5.26) into Equation (5.24).  $\square$

## 5.5 Discussion

Using a change of probability, we first exhibit a time-inhomogeneous Markov process which allows to establish a many-to-one formula, *i.e.* which represents trajectories of a typical lineage of the population process. When compared to the  $\psi$ -spine of Bansaye, 2024, the general idea remains familiar. Indeed, reproduction rates are biased both along the spine and outside of it, transitions being most likely if they increase the average fertility of the spine, where fertility is measured by evaluating  $\psi$  over the spine's descendants at time  $t$ .

However, the  $\psi$ -auxiliary process has the advantage of relaxing the initial positivity assumption on  $\psi$ , which for instance opens the door to more efficient sampling in subpopulations. Further, it does not require exponential weighting of trajectories, potentially facilitating

the numerical evaluation of the many-to-one formula. This comes at the cost of introducing a time-inhomogeneity, a possible inconvenient being the necessity to evaluate the application  $m_\psi$  defined in Equation (5.6). More precisely, it can be delicate to compute explicitly, even for classical models such as the *SIR* model. As a consequence, it may need to be computed numerically, in which case simulating trajectories of the  $\psi$ -auxiliary process through standard algorithms such as Thanh and Priami, 2015 may require excessive computation time.

In this context, the large population approximation may be a pertinent regime. Heuristically, one would expect a similar result as in the case of the  $\psi$ -spine: under appropriate assumptions, the effect of the spine on the population state should vanish, as well as the bias of reproduction rates outside of the spine. Hence, the dynamics of the population's type composition are given by the unique solution  $z$  to differential equation (5.5). Similarly, at time  $s \leq t$ , a spine of type  $x$  is expected to be replaced by descendance  $\mathbf{k}$  and switch to type  $y$  at rate

$$\tau_{\mathbf{k}}(x, z(s)) \mathbf{k}_y \frac{m_\psi(y, z(s), t-s)}{m_\psi(x, z(s), t-s)}.$$

In particular, the process describing the descendance of the spine then corresponds to a time-inhomogeneous multi-type branching process whose reproduction rates depend on a changing environment, given by  $z$ . If similarly to Calvez et al., 2022, we assume that  $z$  admits a stable equilibrium, then starting from this equilibrium, the descendance is described by a classical multi-type branching process.

In the case of the  $\psi$ -spine, the large population approximation has already been established in Bansaye, 2024. Here, we have quantified the error of this approximation when compared to the  $\psi$ -spine in finite population, through the means of a coupling argument. This argument makes use of Assumption 5.4.3 on the convergence of  $\zeta^K$  to  $z$ , which we consider to be reasonable. Indeed, it corresponds to understanding and controlling the fluctuations of the finite-population stochastic process around its deterministic limit, which is a well studied question with several classical regimes: Gaussian fluctuations ( $\varepsilon_K = O(K^{1/2})$ ), moderate deviations ( $\varepsilon_K = O(K^p)$  for  $p \in (0, 1/2)$ ) and large deviations ( $\varepsilon_K = O(1)$ ).

In the case where the deterministic limit has a stable equilibrium at which the process is initialized, moderate deviations appear to be a pertinent regime. Indeed, if we were to replace  $\zeta^K$  by  $Z^K$ , moderate deviations are then known to occur on time scales growing exponentially with  $K$  (Pardoux, 2020; Prodhomme, 2023). This would yield a suitable choice of the sequences  $t_K, \varepsilon_K, \alpha_K$  satisfying Equation (5.23), and further allowing for  $t_K$  to grow to infinity sublinearly with  $K$ . In addition, conditions for the existence of stable equilibria are well understood for classical endemic disease models (Brauer et al., 2019). However, the dynamics of  $\zeta^K$  and  $Z^K$  differ due to the presence of a distinguished individual and the induced bias in the reproduction rates. While intuitively the impact of the spine vanishes in the large population limit, the fluctuations around the deterministic limit still need to rigorously be established in this case.

Among the many possible applications, let us conclude with a perspective connected to understanding the impact of measures such as teleworking on epidemic spread. More precisely, it would be interesting to investigate the modifications of typical contamination chains induced by these interventions. Hence, we would like to consider an endemic model with age classes, where contacts are represented using the contact matrixes from the POLYMOD survey (Mossong et al., 2008). Indeed, these contact matrixes have the advantage of being split into several components, such as household, school and work contacts. This makes it

possible to rescale each component separately, yielding a proxy for teleworking and/or school closure (Di Lauro et al., 2021). As a consequence, spinal constructions for populations with interactions may allow to investigate how the epidemic typically spreads once different control measures are applied.

# Bibliography

- Aguilar, Javier, Aleix Bassolas, Gourab Ghoshal, Surendra Hazarie, Alec Kirkley, Mattia Mazzoli, Sandro Meloni, Sayat Mimar, Vincenzo Nicosia, José J. Ramasco, and Adam Sadilek (2022). "Impact of Urban Structure on Infectious Disease Spreading". In: *Scientific Reports* 12.1, p. 3816. DOI: [10.1038/s41598-022-06720-8](https://doi.org/10.1038/s41598-022-06720-8).
- Ajelli, Marco, Piero Poletti, Alessia Melegaro, and Stefano Merler (2014). "The Role of Different Social Contexts in Shaping Influenza Transmission during the 2009 Pandemic". In: *Scientific Reports* 4, p. 7218. DOI: [10.1038/srep07218](https://doi.org/10.1038/srep07218).
- Aldous, David (1978). "Stopping Times and Tightness". In: *The Annals of Probability* 6.2, pp. 335–340. DOI: [10.1214/aop/1176995579](https://doi.org/10.1214/aop/1176995579).
- Andersson, Håkan and Tom Britton (2000a). "Density Dependent Jump Markov Processes". In: *Stochastic Epidemic Models and Their Statistical Analysis*. Lecture Notes in Statistics. New York, NY: Springer. ISBN: 978-1-4612-1158-7. DOI: [10.1007/978-1-4612-1158-7\\_2](https://doi.org/10.1007/978-1-4612-1158-7_2).
- Andersson, Håkan and Tom Britton (2000b). *Stochastic Epidemic Models and Their Statistical Analysis*. Lecture Notes in Statistics. New York, NY: Springer. ISBN: 978-1-4612-1158-7. DOI: [10.1007/978-1-4612-1158-7\\_2](https://doi.org/10.1007/978-1-4612-1158-7_2).
- Angel, Omer, Remco van der Hofstad, and Cecilia Holmgren (2019). "Limit Laws for Self-Loops and Multiple Edges in the Configuration Model". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 55.3, pp. 1509–1530. DOI: [10.1214/18-AIHP926](https://doi.org/10.1214/18-AIHP926).
- Aparicio, Juan Pablo and Mercedes Pascual (2006). "Building Epidemiological Models from R0: An Implicit Treatment of Transmission in Networks". In: *Proceedings of the Royal Society B: Biological Sciences* 274.1609, pp. 505–512. DOI: [10.1098/rspb.2006.0057](https://doi.org/10.1098/rspb.2006.0057).
- Athreya, Krishna B. and Peter E. Ney (1972). *Branching Processes*. Berlin, Heidelberg: Springer. ISBN: 978-3-642-65373-5 978-3-642-65371-1. DOI: [10.1007/978-3-642-65371-1](https://doi.org/10.1007/978-3-642-65371-1).
- Backhausz, Ágnes, István Z. Kiss, and Péter L. Simon (2022). "The Impact of Spatial and Social Structure on an SIR Epidemic on a Weighted Multilayer Network". In: *Periodica Mathematica Hungarica*. DOI: [10.1007/s10998-021-00440-8](https://doi.org/10.1007/s10998-021-00440-8).
- Bailey, Norman T. J. (1975). *The Mathematical Theory of Infectious Diseases and Its Applications*. Second Edition. Charles Griffin. ISBN: 0-85264-231-8.
- Ball, Frank G., Tom Britton, Ka Yin Leung, and David Sirl (2019). "A Stochastic SIR Network Epidemic Model with Preventive Dropping of Edges". In: *Journal of Mathematical Biology* 78.6, pp. 1875–1951. DOI: [10.1007/s00285-019-01329-4](https://doi.org/10.1007/s00285-019-01329-4).
- Ball, Frank G. and Peter Donnelly (1995). "Strong Approximations for Epidemic Models". In: *Stochastic Processes and their Applications* 55.1, pp. 1–21. DOI: [10.1016/0304-4149\(94\)00034-Q](https://doi.org/10.1016/0304-4149(94)00034-Q).
- Ball, Frank G., Denis Mollison, and Gianpaolo Scalia-Tomba (1997). "Epidemics with Two Levels of Mixing". In: *The Annals of Applied Probability* 7.1, pp. 46–89. DOI: [10.1214/aoap/1034625252](https://doi.org/10.1214/aoap/1034625252).
- Ball, Frank G. and Peter Neal (2002). "A General Model for Stochastic SIR Epidemics with Two Levels of Mixing". In: *Mathematical Biosciences* 180.1, pp. 73–102. DOI: [10.1016/S0025-5564\(02\)00125-6](https://doi.org/10.1016/S0025-5564(02)00125-6).

- Ball, Frank G. and Peter Neal (2008). “Network Epidemic Models with Two Levels of Mixing”. In: *Mathematical Biosciences* 212.1, pp. 69–87. DOI: [10.1016/j.mbs.2008.01.001](https://doi.org/10.1016/j.mbs.2008.01.001).
- Ball, Frank G., Lorenzo Pellis, and Pieter Trapman (2016). “Reproduction Numbers for Epidemic Models with Households and Other Social Structures II: Comparisons and Implications for Vaccination”. In: *Mathematical Biosciences* 274, pp. 108–139. DOI: [10.1016/j.mbs.2016.01.006](https://doi.org/10.1016/j.mbs.2016.01.006).
- Ball, Frank G., David J. Sirl, and Pieter Trapman (2014). “Epidemics on Random Intersection Graphs”. In: *The Annals of Applied Probability* 24.3, pp. 1081–1128. DOI: [10.1214/13-AAP942](https://doi.org/10.1214/13-AAP942).
- Bansal, Shweta, Jonathan Read, Babak Pourbohloul, and Lauren Ancel Meyers (2010). “The Dynamic Nature of Contact Networks in Infectious Disease Epidemiology”. In: *Journal of Biological Dynamics* 4.5, pp. 478–489. DOI: [10.1080/17513758.2010.503376](https://doi.org/10.1080/17513758.2010.503376).
- Bansaye, Vincent (2024). “Spine for Interacting Populations and Sampling”. In: *Bernoulli* 30.2, pp. 1555–1585. DOI: [10.3150/23-BEJ1645](https://doi.org/10.3150/23-BEJ1645).
- Bansaye, Vincent, Bertrand Cloez, and Pierre Gabriel (2020). “Ergodic Behavior of Non-Conservative Semigroups via Generalized Doeblin’s Conditions”. In: *Acta Applicandae Mathematicae* 166.1, pp. 29–72. DOI: [10.1007/s10440-019-00253-5](https://doi.org/10.1007/s10440-019-00253-5).
- Bansaye, Vincent, Bertrand Cloez, Pierre Gabriel, and Aline Marguet (2022). “A Non-Conservative Harris Ergodic Theorem”. In: *Journal of the London Mathematical Society* 106.3. DOI: [10.1112/jlms.12639](https://doi.org/10.1112/jlms.12639).
- Bansaye, Vincent, François Deslandes, Madeleine Kubasch, and Elisabeta Vergu (2023a). *The Epidemiological Footprint of Contact Structures in Models with Two Levels of Mixing*. DOI: [10.48550/arXiv.2303.05287](https://doi.org/10.48550/arXiv.2303.05287). arXiv: 2303.05287.
- Bansaye, Vincent, Xavier Erny, and Sylvie Méléard (2023b). *Sharp Approximation and Hitting Times for Stochastic Invasion Processes*. DOI: [10.48550/arXiv.2212.14320](https://doi.org/10.48550/arXiv.2212.14320). arXiv: 2212.14320.
- Barbour, Andrew D. (1978). “A Stochastic Model for the Transmission of Bilharzia”. In: *Mathematical Biosciences* 38.3, pp. 303–312. DOI: [10.1016/0025-5564\(78\)90051-2](https://doi.org/10.1016/0025-5564(78)90051-2).
- Barbour, Andrew D. and Gesine Reinert (2013). “Approximating the Epidemic Curve”. In: *Electronic Journal of Probability* 18, pp. 1–30. DOI: [10.1214/EJP.v18-2557](https://doi.org/10.1214/EJP.v18-2557).
- Barbour, Andrew D. and Sergey Utev (2004). “Approximating the Reed–Frost Epidemic Process”. In: *Stochastic Processes and their Applications* 113.2, pp. 173–197. DOI: [10.1016/j.spa.2004.03.013](https://doi.org/10.1016/j.spa.2004.03.013).
- Barnard, Rosanna C., Istvan Z. Kiss, Luc Berthouze, and Joel C. Miller (2018). “Edge-Based Compartmental Modelling of an SIR Epidemic on a Dual-Layer Static–Dynamic Multiplex Network with Tunable Clustering”. In: *Bulletin of Mathematical Biology* 80.10, pp. 2698–2733. DOI: [10.1007/s11538-018-0484-5](https://doi.org/10.1007/s11538-018-0484-5).
- Barrat, Alain and Martin Weigt (2000). “On the Properties of Small-World Network Models”. In: *The European Physical Journal B - Condensed Matter and Complex Systems* 13.3, pp. 547–560. DOI: [10.1007/s100510050067](https://doi.org/10.1007/s100510050067).
- Barthélemy, Marc (2011). “Spatial Networks”. In: *Physics Reports* 499.1, pp. 1–101. DOI: [10.1016/j.physrep.2010.11.002](https://doi.org/10.1016/j.physrep.2010.11.002).
- Becker, Niels G. and Klaus Dietz (1995). “The Effect of Household Distribution on Transmission and Control of Highly Infectious Diseases”. In: *Mathematical Biosciences* 127.2, pp. 207–219. DOI: [10.1016/0025-5564\(94\)00055-5](https://doi.org/10.1016/0025-5564(94)00055-5).



- Billingsley, Patrick (1999). *Convergence of Probability Measures*. 2nd ed. Wiley Series in Probability and Statistics. Probability and Statistics Section. New York: Wiley. ISBN: 978-0-471-19745-4. DOI: [10.1002/9780470316962](https://doi.org/10.1002/9780470316962).
- Bin Nafisah, Sharafaldeen, Aliyah H. Alamery, Aminah Al Nafesa, Bakhitah Aleid, and Nora A. Brazanji (2018). "School Closure during Novel Influenza: A Systematic Review". In: *Journal of Infection and Public Health* 11.5, pp. 657–661. DOI: [10.1016/j.jiph.2018.01.003](https://doi.org/10.1016/j.jiph.2018.01.003).
- Bollobás, Béla (1980). "A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs". In: *European Journal of Combinatorics* 1.4, pp. 311–316. DOI: [10.1016/S0195-6698\(80\)80030-8](https://doi.org/10.1016/S0195-6698(80)80030-8).
- Bozon, Michel and Francois Heran (1989). "Finding a Spouse: A Survey of How French Couples Meet". In: *Population: An English Selection* 44.1, pp. 91–121. JSTOR: [2949076](https://www.jstor.org/stable/2949076).
- Brauer, Fred, Carlos Castillo-Chavez, and Zhilan Feng (2019). "Endemic Disease Models". In: *Mathematical Models in Epidemiology* 69, pp. 63–116. DOI: [10.1007/978-1-4939-9828-9\\_3](https://doi.org/10.1007/978-1-4939-9828-9_3).
- Brauer, Fred, Pauline van den Driessche, and Jianhong Wu, eds. (2008). *Mathematical Epidemiology*. 2008th edition. Berlin Heidelberg: Springer. ISBN: 978-3-540-78910-9.
- Britton, Tom, Frank Ball, and Pieter Trapman (2020). "A Mathematical Model Reveals the Influence of Population Heterogeneity on Herd Immunity to SARS-CoV-2". In: *Science* 369.6505, pp. 846–849. DOI: [10.1126/science.abc6810](https://doi.org/10.1126/science.abc6810).
- Britton, Tom, Svante Janson, and Anders Martin-Löf (2007). "Graphs with Specified Degree Distributions, Simple Epidemics, and Local Vaccination Strategies". In: *Advances in Applied Probability* 39, pp. 922–948. DOI: [10.1239/aap/1198177233](https://doi.org/10.1239/aap/1198177233).
- Britton, Tom and Lasse Leskelä (2023). "Optimal Intervention Strategies for Minimizing Total Incidence during an Epidemic". In: *SIAM Journal on Applied Mathematics* 83.2, pp. 354–373. DOI: [10.1137/22M1504433](https://doi.org/10.1137/22M1504433).
- Britton, Tom and Etienne Pardoux (2019a). "Chapter 3 A General Two-Level Mixing Model". In: *Stochastic Epidemic Models with Inference*. Ed. by Tom Britton and Etienne Pardoux. Lecture Notes in Mathematics. Cham: Springer International Publishing, pp. 159–213. ISBN: 978-3-030-30900-8. DOI: [10.1007/978-3-030-30900-8\\_7](https://doi.org/10.1007/978-3-030-30900-8_7).
- Britton, Tom and Etienne Pardoux (2019b). *Stochastic Epidemics in a Homogeneous Community*. Vol. 2255. DOI: [10.1007/978-3-030-30900-8](https://doi.org/10.1007/978-3-030-30900-8). arXiv: [1808.05350](https://arxiv.org/abs/1808.05350).
- Britton, Tom and Gianpaolo Scalia Tomba (2019). "Estimation in Emerging Epidemics: Biases and Remedies". In: *Journal of The Royal Society Interface* 16.150, p. 20180670. DOI: [10.1098/rsif.2018.0670](https://doi.org/10.1098/rsif.2018.0670).
- Brooks-Pollock, Ellen and James L. N. Wood (2015). "Eliminating Bovine Tuberculosis in Cattle and Badgers: Insight from a Dynamic Model". In: *Proceedings of the Royal Society B: Biological Sciences* 282.1808, p. 20150374. DOI: [10.1098/rspb.2015.0374](https://doi.org/10.1098/rspb.2015.0374).
- Calvez, Vincent, Benoît Henry, Sylvie Méléard, and Viet Chi Tran (2022). "Dynamics of Lineages in Adaptation to a Gradual Environmental Change". In: *Annales Henri Lebesgue* 5, pp. 729–777. DOI: [10.5802/ah1.135](https://doi.org/10.5802/ah1.135).
- Caulkins, Jonathan P., Dieter Grass, Gustav Feichtinger, Richard F. Hartl, Peter M. Kort, Alexia Prskawetz, Andrea Seidl, and Stefan Wrzaczek (2021). "The Optimal Lockdown Intensity for COVID-19". In: *Journal of Mathematical Economics* 93, p. 102489. DOI: [10.1016/j.jmateco.2021.102489](https://doi.org/10.1016/j.jmateco.2021.102489).



- Chatterjee, Gurudas (1974). "Negative Integral Powers of a Bidiagonal Matrix". In: *Mathematics of Computation* 28, pp. 713–714. DOI: [10.1090/S0025-5718-1974-0371049-5](https://doi.org/10.1090/S0025-5718-1974-0371049-5).
- Cléménçon, Stéphan, Hector De Arazoza, Fabrice Rossi, and Viet Chi Tran (2015). "A Statistical Network Analysis of the HIV/AIDS Epidemics in Cuba". In: *Social Network Analysis and Mining* 5.1, p. 58. DOI: [10.1007/s13278-015-0291-z](https://doi.org/10.1007/s13278-015-0291-z).
- Colosi, Elisabetta, Giulia Bassignana, Alain Barrat, and Vittoria Colizza (2022). "Modelling COVID-19 in School Settings to Evaluate Prevention and Control Protocols". In: *Anaesthesia Critical Care & Pain Medicine* 41.2, p. 101047. DOI: [10.1016/j.accpm.2022.101047](https://doi.org/10.1016/j.accpm.2022.101047).
- Contreras, Diego Andrés, Elisabetta Colosi, Giulia Bassignana, Vittoria Colizza, and Alain Barrat (2022). "Impact of Contact Data Resolution on the Evaluation of Interventions in Mathematical Models of Infectious Diseases". In: *Journal of the Royal Society Interface* 19.191, p. 20220164. DOI: [10.1098/rsif.2022.0164](https://doi.org/10.1098/rsif.2022.0164).
- Danon, Leon, Ashley P. Ford, Thomas House, Chris P. Jewell, Matt J. Keeling, Gareth O. Roberts, Joshua V. Ross, and Matthew C. Vernon (2011). "Networks and the Epidemiology of Infectious Disease". In: *Interdisciplinary Perspectives on Infectious Diseases* 2011, e284909. DOI: [10.1155/2011/284909](https://doi.org/10.1155/2011/284909).
- Davies, Nicholas G., Petra Klepac, Yang Liu, Kiesha Prem, Mark Jit, CMMID COVID-19 working group, and Rosalind M. Eggo (2020). "Age-Dependent Effects in the Transmission and Control of COVID-19 Epidemics". In: *Nature Medicine* 26.8, pp. 1205–1211. DOI: [10.1038/s41591-020-0962-9](https://doi.org/10.1038/s41591-020-0962-9).
- Decreusefond, Laurent, Jean-Stéphane Dherain, Pascal Moyal, and Viet Chi Tran (2012). "Large Graph Limit for an SIR Process in Random Network with Heterogeneous Connectivity". In: *The Annals of Applied Probability* 22.2. DOI: [10.1214/11-AAP773](https://doi.org/10.1214/11-AAP773).
- Deijfen, Maria and Willemien Kets (2009). "Random Intersection Graphs with Tunable Degree Distribution and Clustering". In: *Probability in the Engineering and Informational Sciences* 23.4, pp. 661–674. DOI: [10.1017/S0269964809990064](https://doi.org/10.1017/S0269964809990064).
- Del Moral, Pierre (2004). *Feynman-Kac Formulae*. Ed. by J. Gani, C. C. Heyde, and T. G. Kurtz. Probability and Its Applications. New York, NY: Springer. ISBN: 978-1-4419-1902-1 978-1-4684-9393-1. DOI: [10.1007/978-1-4684-9393-1](https://doi.org/10.1007/978-1-4684-9393-1).
- del Valle Rafo, María, Juan Pablo Di Mauro, and Juan Pablo Aparicio (2021). "Disease Dynamics and Mean Field Models for Clustered Networks". In: *Journal of Theoretical Biology* 526, p. 110554. DOI: [10.1016/j.jtbi.2020.110554](https://doi.org/10.1016/j.jtbi.2020.110554).
- Delmas, Jean-François, Paolo Frasca, Federica Garin, Viet Chi Tran, Aurélien Velleret, and Pierre-André Zitt (2023). *Individual Based SIS Models on (Not so) Dense Large Random Networks*. DOI: [10.48550/arXiv.2302.13385](https://doi.org/10.48550/arXiv.2302.13385). arXiv: [2302.13385](https://arxiv.org/abs/2302.13385).
- Dhanjal, Charanpal, Stephan Clemençon, Hector De Arazoza, Fabrice Rossi, and Viet Chi Tran (2011). *The Evolution of the Cuban HIV/AIDS Network*. DOI: [10.48550/arXiv.1109.2499](https://doi.org/10.48550/arXiv.1109.2499). arXiv: [1109.2499](https://arxiv.org/abs/1109.2499).
- Di Lauro, Francesco, Luc Berthouze, Matthew D. Dorey, Joel C. Miller, and István Z. Kiss (2021). "The Impact of Contact Structure and Mixing on Control Measures and Disease-Induced Herd Immunity in Epidemic Models: A Mean-Field Model Perspective". In: *Bulletin of Mathematical Biology* 83.11, p. 117. DOI: [10.1007/s11538-021-00947-8](https://doi.org/10.1007/s11538-021-00947-8).
- Diekmann, Odo and Heesterbeek (2000). *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Wiley Series in Mathematical and

- Computational Biology. Chichester: John Wiley & Sons. ISBN: 978-0-471-49241-2. DOI: [10.1093/ije/30.1.186](https://doi.org/10.1093/ije/30.1.186).
- Doenges, Philipp, Thomas Götz, Tyll Krueger, Karol Niedziewski, Viola Priesemann, and Moritz Schaefer (2023). *SIR-Model for Households*. DOI: [10.48550/arXiv.2301.04355](https://doi.org/10.48550/arXiv.2301.04355). arXiv: [2301.04355](https://arxiv.org/abs/2301.04355).
- Duchamps, Jean-Jil, Félix Foutel-Rodier, and Emmanuel Schertzer (2023). "General Epidemiological Models: Law of Large Numbers and Contact Tracing". In: *Electronic Journal of Probability* 28.none, pp. 1–37. DOI: [10.1214/23-EJP992](https://doi.org/10.1214/23-EJP992).
- Durrett, Rick (2006). *Random Graph Dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press. ISBN: 978-0-521-15016-3. DOI: [10.1017/CB09780511546594](https://doi.org/10.1017/CB09780511546594).
- Eames, Ken T. D., Shweta Bansal, Simon Frost, and Steven Riley (2015). "Six Challenges in Measuring Contact Networks for Use in Modelling". In: *Epidemics. Challenges in Modelling Infectious Disease Dynamics* 10, pp. 72–77. DOI: [10.1016/j.epidem.2014.08.006](https://doi.org/10.1016/j.epidem.2014.08.006).
- Eames, Ken T. D. and Matt J. Keeling (2002). "Modeling Dynamic and Network Heterogeneities in the Spread of Sexually Transmitted Diseases". In: *Proceedings of the National Academy of Sciences* 99.20, pp. 13330–13335. DOI: [10.1073/pnas.202244299](https://doi.org/10.1073/pnas.202244299).
- Erdős, Paul and Alfréd Rényi (1959). "On Random Graphs. I." In: *Publicaciones Matheamticae* 6, pp. 290–297. DOI: [10.5486](https://doi.org/10.5486).
- Ethier, Stewart N. and Thomas G. Kurtz (1986). *Markov Processes: Characterization and Convergence*. New York, NY: Wiley. ISBN: 978-0-471-08186-9. DOI: [10.1002/9780470316658](https://doi.org/10.1002/9780470316658).
- Faivre, Robert, Bertrand Iooss, Stéphanie Mahévas, David Makowski, and Herve Monod (2013). *Analyse de Sensibilité et Exploration de Modèles*. Collection Savoir-Faire. Editions Quae. ISBN: 978-2-7592-1906-3.
- Favero, Martina, Gianpaolo Scalia Tomba, and Tom Britton (2022). "Modelling Preventive Measures and Their Effect on Generation Times in Emerging Epidemics". In: *Journal of The Royal Society Interface* 19.191, p. 20220128. DOI: [10.1098/rsif.2022.0128](https://doi.org/10.1098/rsif.2022.0128).
- Feng, Zhilan, Dashun Xu, and Haiyun Zhao (2007). "Epidemiological Models with Non-Exponentially Distributed Disease Stages and Applications to Disease Control". In: *Bulletin of Mathematical Biology* 69.5, pp. 1511–1536. DOI: [10.1007/s11538-006-9174-9](https://doi.org/10.1007/s11538-006-9174-9).
- Ferguson, Neil M., Derek A. T. Cummings, Christophe Fraser, James C. Cajka, Philip C. Cooley, and Donald S. Burke (2006). "Strategies for Mitigating an Influenza Pandemic". In: *Nature* 442.7101, pp. 448–452. DOI: [10.1038/nature04795](https://doi.org/10.1038/nature04795).
- Fontanet, Arnaud and Simon Cauchemez (2020). "COVID-19 Herd Immunity: Where Are We?" In: *Nature Reviews Immunology* 20.10, pp. 583–584. DOI: [10.1038/s41577-020-00451-5](https://doi.org/10.1038/s41577-020-00451-5).
- Forien, Raphael, Guodong Pang, and Etienne Pardoux (2022). "Recent Advances in Epidemic Modeling: Non-Markov Stochastic Models and Their Scaling Limits". In: *The Graduate Journal of Mathematics* 7.2, pp. 19–75. DOI: [10.48550/arXiv.2106.08466](https://doi.org/10.48550/arXiv.2106.08466).
- Fournier, Nicolas and Sylvie Méléard (2004). "A Microscopic Probabilistic Description of a Locally Regulated Population and Macroscopic Approximations". In: *The Annals of Applied Probability* 14.4, pp. 1880–1919. DOI: [10.1214/105051604000000882](https://doi.org/10.1214/105051604000000882).
- Fransson, Carolina (2022). *The Real-Time Growth Rate of Stochastic Epidemics on Random Intersection Graphs*. DOI: [10.48550/arXiv.2211.11034](https://doi.org/10.48550/arXiv.2211.11034). arXiv: [2211.11034](https://arxiv.org/abs/2211.11034).

- Galmiche, Simon, Tiffany Charmet, Laura Schaeffer, Juliette Paireau, Rebecca Grant, Olivia Chény, Cassandre Von Platen, Alexandra Maurizot, Carole Blanc, Annika Dinis, Sophie Martin, Faïza Omar, Christophe David, Alexandra Septfons, Simon Cauchemez, Fabrice Carrat, Alexandra Mailles, Daniel Levy-Bruhl, and Arnaud Fontanet (2021). “Exposures Associated with SARS-CoV-2 Infection in France: A Nationwide Online Case-Control Study”. In: *The Lancet Regional Health – Europe* 7. DOI: [10.1016/j.lanepe.2021.100148](https://doi.org/10.1016/j.lanepe.2021.100148).
- Georgii, Hans-Otto and Ellen Baake (2003). “Supercritical Multitype Branching Processes: The Ancestral Types of Typical Individuals”. In: *Advances in Applied Probability* 35.4, pp. 1090–1110. DOI: [10.1239/aap/1067436336](https://doi.org/10.1239/aap/1067436336).
- Giles, John R., Elisabeth Zu Erbach-Schoenberg, Andrew J. Tatem, Lauren Gardner, Ottar N. Bjørnstad, C. J. E. Metcalf, and Amy Wesolowski (2020). “The Duration of Travel Impacts the Spatial Dynamics of Infectious Diseases”. In: *Proceedings of the National Academy of Sciences of the United States of America* 117.36, pp. 22572–22579. DOI: [10.1073/pnas.1922663117](https://doi.org/10.1073/pnas.1922663117).
- Goldstein, Edward, K. Paur, Christophe Fraser, Eben Kenah, Jacco Wallinga, and Marc Lipsitch (2009). “Reproductive Numbers, Epidemic Spread and Control in a Community of Households”. In: *Mathematical Biosciences* 221.1, pp. 11–25. DOI: [10.1016/j.mbs.2009.06.002](https://doi.org/10.1016/j.mbs.2009.06.002).
- Government of India (2018). *Educational Statistics at a Glance*. Tech. rep. New Delhi: Statistics division, Department of School Education & Literacy, Ministry of Human Resource Development, Government of India.
- Government of India (2022). *Annual Survey of Industries 2019-2020: Summary Results for Factory Sector*. Tech. rep. Kolkata: Data Quality Assurance Division (Industrial Statistics Wing), National Statistical Office, Ministry of Statistics and Programme Implementation, Government of India.
- Harris, Simon C., Marion Hesse, and Andreas E. Kyprianou (2016). “Branching Brownian Motion in a Strip: Survival near Criticality”. In: *The Annals of Probability* 44.1, pp. 235–275. DOI: [10.1214/14-AOP972](https://doi.org/10.1214/14-AOP972).
- Harris, Simon C. and Matthew I. Roberts (2017). “The Many-to-Few Lemma and Multiple Spines”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 53.1, pp. 226–242. DOI: [10.1214/15-AIHP714](https://doi.org/10.1214/15-AIHP714).
- Hébert-Dufresne, Laurent, Pierre-André Noël, Vincent Marceau, Antoine Allard, and Louis J. Dubé (2010). “Propagation Dynamics on Networks Featuring Complex Topologies”. In: *Physical Review E* 82.3, p. 036115. DOI: [10.1103/PhysRevE.82.036115](https://doi.org/10.1103/PhysRevE.82.036115).
- Heesterbeek, J. A. P. and Klaus Dietz (1996). “The Concept of  $R_0$  in Epidemic Theory”. In: *Statistica Neerlandica* 50.1, pp. 89–110. DOI: [10.1111/j.1467-9574.1996.tb01482.x](https://doi.org/10.1111/j.1467-9574.1996.tb01482.x).
- Hilton, Joe, Heather Riley, Lorenzo Pellis, Rabia Aziza, Samuel P. C. Brand, Ivy K. Kombe, John Ojal, Andrea Parisi, Matt J. Keeling, D. James Nokes, Robert Manson-Sawko, and Thomas House (2022). “A Computational Framework for Modelling Infectious Disease Policy Based on Age and Household Structure with Applications to the COVID-19 Pandemic”. In: *PLOS Computational Biology* 18.9, e1010390. DOI: [10.1371/journal.pcbi.1010390](https://doi.org/10.1371/journal.pcbi.1010390).

- Hofstad, Remco van der (2016). *Random Graphs and Complex Networks: Volume 1*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press. ISBN: 978-1-107-17287-6. DOI: [10.1017/9781316779422](https://doi.org/10.1017/9781316779422).
- Hofstad, Remco van der (2024). *Random Graphs and Complex Networks: Volume 2*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press. ISBN: 978-1-107-17400-9. DOI: [10.1017/9781316795552](https://doi.org/10.1017/9781316795552).
- House, Thomas and Matt J. Keeling (2008). "Deterministic Epidemic Models with Explicit Household Structure". In: *Mathematical Biosciences* 213.1, pp. 29–39. DOI: [10.1016/j.mbs.2008.01.011](https://doi.org/10.1016/j.mbs.2008.01.011).
- House, Thomas and Matt J. Keeling (2011). "Insights from Unifying Modern Approximations to Infections on Networks". In: *Journal of the Royal Society Interface* 8.54, pp. 67–73. DOI: [10.1098/rsif.2010.0179](https://doi.org/10.1098/rsif.2010.0179).
- Huber, Greg, Mason Kamb, Kyle Kawagoe, Lucy M. Li, Boris Veytsman, David Yllanes, and Dan Zigmund (2020). "A Minimal Model for Household Effects in Epidemics". In: *Physical Biology* 17.6, p. 065010. DOI: [10.1088/1478-3975/abb209](https://doi.org/10.1088/1478-3975/abb209).
- Hyman, James M. and Jia Li (2005). "Differential Susceptibility Epidemic Models". In: *Journal of Mathematical Biology* 50.6, pp. 626–644. DOI: [10.1007/s00285-004-0301-7](https://doi.org/10.1007/s00285-004-0301-7).
- IIPS (2021). *National Family Health Survey (NFHS-5) 2019-21: India: Volume I*. Tech. rep. Mumbai: International Institute for Population Sciences (IIPS) and ICF; Ministry of Health and Family Welfare, Government of India.
- Isella, Lorenzo, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-François Pinton, and Wouter Van den Broeck (2011). "What's in a Crowd? Analysis of Face-to-Face Behavioral Networks". In: *Journal of Theoretical Biology* 271.1, pp. 166–180. DOI: [10.1016/j.jtbi.2010.11.033](https://doi.org/10.1016/j.jtbi.2010.11.033).
- Jacobsen, Karly A., Mark G. Burch, Joseph H. Tien, and Grzegorz A. Rempała (2018). "The Large Graph Limit of a Stochastic Epidemic Model on a Dynamic Multilayer Network". In: *Journal of Biological Dynamics* 12.1, pp. 746–788. DOI: [10.1080/17513758.2018.1515993](https://doi.org/10.1080/17513758.2018.1515993).
- Jacquez, John A., Carl P. Simon, James Koopman, Lisa Sattenspiel, and Timothy Perry (1988). "Modeling and Analyzing HIV Transmission: The Effect of Contact Patterns". In: *Mathematical Biosciences* 92.2, pp. 119–199. DOI: [10.1016/0025-5564\(88\)90031-4](https://doi.org/10.1016/0025-5564(88)90031-4).
- Jansen, Michiel J. W. (1999). "Analysis of Variance Designs for Model Output". In: *Computer Physics Communications* 117.1, pp. 35–43. DOI: [10.1016/S0010-4655\(98\)00154-4](https://doi.org/10.1016/S0010-4655(98)00154-4).
- Janson, Svante, Malwina Luczak, and Peter Windridge (2014). "Law of Large Numbers for the SIR Epidemic on a Random Graph with given Degrees". In: *Random Structures & Algorithms* 45.4, pp. 726–763. DOI: [10.1002/rsa.20575](https://doi.org/10.1002/rsa.20575).
- Joffe, Anatole and Michel Metivier (1986). "Weak Convergence of Sequences of Semimartingales with Applications to Multitype Branching Processes". In: *Advances in Applied Probability* 18.1, pp. 20–65. DOI: [10.2307/1427238](https://doi.org/10.2307/1427238).
- Jourdain, Benjamin, Sylvie Méléard, and Wojbor A. Woyczynski (2012). "Lévy Flights in Evolutionary Ecology". In: *Journal of Mathematical Biology* 65.4, pp. 677–707. DOI: [10.1007/s00285-011-0478-5](https://doi.org/10.1007/s00285-011-0478-5).
- Kalmijn, Matthijs and Henk Flap (2001). "Assortative Meeting and Mating: Unintended Consequences of Organized Settings for Partner Choices\*". In: *Social Forces* 79.4, pp. 1289–1312. DOI: [10.1353/sof.2001.0044](https://doi.org/10.1353/sof.2001.0044).

- Karoński, Michał, Edward R. Scheinerman, and Karen B. Singer-Cohen (1999). "On Random Intersection Graphs: The Subgraph Problem". In: *Combinatorics, Probability and Computing* 8.1-2, pp. 131–159. DOI: [10.1017/S0963548398003459](https://doi.org/10.1017/S0963548398003459).
- Karrer, Brian and Mark E. J. Newman (2010a). "Message Passing Approach for General Epidemic Models". In: *Physical Review E* 82.1, p. 016101. DOI: [10.1103/PhysRevE.82.016101](https://doi.org/10.1103/PhysRevE.82.016101).
- Karrer, Brian and Mark E. J. Newman (2010b). "Random Graphs Containing Arbitrary Distributions of Subgraphs". In: *Physical Review E* 82.6, p. 066118. DOI: [10.1103/PhysRevE.82.066118](https://doi.org/10.1103/PhysRevE.82.066118).
- Keeling, Matt J. and Ken T. D. Eames (2005). "Networks and Epidemic Models". In: *Journal of the Royal Society, Interface* 2.4, pp. 295–307. DOI: [10.1098/rsif.2005.0051](https://doi.org/10.1098/rsif.2005.0051).
- Keeling, Matt J., David A. Rand, and A. J. Morris (1997). "Correlation Models for Childhood Epidemics". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 264.1385, pp. 1149–1156. DOI: [10.1098/rspb.1997.0159](https://doi.org/10.1098/rspb.1997.0159).
- Kermack, William O. and Anderson G. McKendrick (1997). "A Contribution to the Mathematical Theory of Epidemics". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772, pp. 700–721. DOI: [10.1098/rspa.1927.0118](https://doi.org/10.1098/rspa.1927.0118).
- Kiss, Istvan Z., Luc Berthouze, and Wasiur R. KhudaBukhsh (2024). "Towards Inferring Network Properties from Epidemic Data". In: *Bulletin of Mathematical Biology* 86.1, p. 6. DOI: [10.1007/s11538-023-01235-3](https://doi.org/10.1007/s11538-023-01235-3).
- Kiss, István Z., Eben Kenah, and Grzegorz A. Rempala (2023). "Necessary and Sufficient Conditions for Exact Closures of Epidemic Equations on Configuration Model Networks". In: *Journal of Mathematical Biology* 87.2, p. 36. DOI: [10.1007/s00285-023-01967-9](https://doi.org/10.1007/s00285-023-01967-9).
- Kiss, István Z., Joel C. Miller, and Péter L. Simon (2017). *Mathematics of Epidemics on Networks: From Exact to Approximate Models*. Vol. 46. Interdisciplinary Applied Mathematics. Cham: Springer International Publishing. ISBN: 978-3-319-50804-7 978-3-319-50806-1. DOI: [10.1007/978-3-319-50806-1](https://doi.org/10.1007/978-3-319-50806-1).
- Kubasch, Madeleine (2023). *Large Population Limit for a Multilayer SIR Model Including Households and Workplaces*. DOI: [10.48550/arXiv.2305.17064](https://doi.org/10.48550/arXiv.2305.17064). arXiv: [2305.17064](https://arxiv.org/abs/2305.17064).
- Kurauskas, Valentas (2022). "On Local Weak Limit and Subgraph Counts for Sparse Random Graphs". In: *Journal of Applied Probability* 59.3, pp. 755–776. DOI: [10.1017/jpr.2021.84](https://doi.org/10.1017/jpr.2021.84).
- Kurtz, Thomas G. (1981). "Epidemic Models". In: *Approximation of Population Processes*. CBMS-NSF Regional Conference Series in Applied Mathematics 36. Philadelphia, Pa: Society for Industrial and Applied Mathematics. ISBN: 978-0-89871-169-1. DOI: [10.1137/1.9781611970333](https://doi.org/10.1137/1.9781611970333).
- Lajmanovich, Ana and James A. Yorke (1976). "A Deterministic Model for Gonorrhoea in a Nonhomogeneous Population". In: *Mathematical Biosciences* 28.3, pp. 221–236. DOI: [10.1016/0025-5564\(76\)90125-5](https://doi.org/10.1016/0025-5564(76)90125-5).
- Lindquist, Jennifer, Junling Ma, Pauline van den Driessche, and Frederick H. Willeboordse (2011). "Effective Degree Network Disease Models". In: *Journal of Mathematical Biology* 62.2, pp. 143–164. DOI: [10.1007/s00285-010-0331-2](https://doi.org/10.1007/s00285-010-0331-2).
- Lloyd, Alun L. (2001). "Realistic Distributions of Infectious Periods in Epidemic Models: Changing Patterns of Persistence and Dynamics". In: *Theoretical Population Biology* 60.1, pp. 59–71. DOI: [10.1006/tpbi.2001.1525](https://doi.org/10.1006/tpbi.2001.1525).



- Locatelli, Isabella, Bastien Trächsel, and Valentin Rousson (2021). “Estimating the Basic Reproduction Number for COVID-19 in Western Europe”. In: *PLOS ONE* 16.3, e0248731. DOI: [10.1371/journal.pone.0248731](https://doi.org/10.1371/journal.pone.0248731).
- Luca, Giancarlo De, Kim Van Kerckhove, Pietro Coletti, Chiara Poletto, Nathalie Bossuyt, Niel Hens, and Vittoria Colizza (2018). “The Impact of Regular School Closure on Seasonal Influenza Epidemics: A Data-Driven Spatial Transmission Model for Belgium”. In: *BMC Infectious Diseases* 18.1, p. 29. DOI: [10.1186/s12879-017-2934-3](https://doi.org/10.1186/s12879-017-2934-3).
- Lyons, Russell, Robin Pemantle, and Yuval Peres (1995). “Conceptual Proofs of LLogL Criteria for Mean Behavior of Branching Processes”. In: *The Annals of Probability* 23.3, pp. 1125–1138. DOI: [10.1214/aop/1176988176](https://doi.org/10.1214/aop/1176988176).
- Ma, Junling, Pauline van den Driessche, and Frederick H. Willeboordse (2013). “Effective Degree Household Network Disease Model”. In: *Journal of Mathematical Biology* 66.1, pp. 75–94. DOI: [10.1007/s00285-011-0502-9](https://doi.org/10.1007/s00285-011-0502-9).
- Marguet, Aline (2019). “Uniform Sampling in a Structured Branching Population”. In: *Bernoulli* 25.4A, pp. 2649–2695. DOI: [10.3150/18-BEJ1066](https://doi.org/10.3150/18-BEJ1066).
- Medous, Charles (2023). *Spinal Constructions for Continuous Type-Space Branching Processes with Interactions*. DOI: [10.48550/arXiv.2309.15449](https://doi.org/10.48550/arXiv.2309.15449). arXiv: [2309.15449](https://arxiv.org/abs/2309.15449).
- Meleard, Sylvie and Vincent Bansaye (2015). *Stochastic Models for Structured Populations: Scaling Limits and Long Time Behavior*. Cham: Springer International Publishing. ISBN: 978-3-319-21710-9 978-3-319-21711-6. DOI: [10.1007/978-3-319-21711-6](https://doi.org/10.1007/978-3-319-21711-6).
- Mendez-Brito, Alba, Charbel El Bcheraoui, and Francisco Pozo-Martin (2021). “Systematic Review of Empirical Studies Comparing the Effectiveness of Non-Pharmaceutical Interventions against COVID-19”. In: *Journal of Infection* 83.3, pp. 281–293. DOI: [10.1016/j.jinf.2021.06.018](https://doi.org/10.1016/j.jinf.2021.06.018).
- Miller, Joel C. (2009). “Percolation and Epidemics in Random Clustered Networks”. In: *Physical Review E* 80.2, p. 020901. DOI: [10.1103/PhysRevE.80.020901](https://doi.org/10.1103/PhysRevE.80.020901).
- Miller, Joel C (2011). “A Note on a Paper by Erik Volz: SIR Dynamics in Random Networks”. In: *Journal of Mathematical Biology* 62, pp. 349–358. DOI: [10.1007/s00285-010-0337-9](https://doi.org/10.1007/s00285-010-0337-9).
- Mode, Charles J. and Candace K. Sleeman (2000). “Life Cycle Models and Mean Functions”. In: *Stochastic Processes in Epidemiology: HIV/AIDS, Other Infectious Diseases and Computers*. Singapore: World Scientific, pp. 175–180. ISBN: 978-981-02-4097-4.
- Molloy, Michael and Bruce Reed (1995). “A Critical Point for Random Graphs with a given Degree Sequence”. In: *Random Structures & Algorithms* 6.2-3, pp. 161–180. DOI: [10.1002/rsa.3240060204](https://doi.org/10.1002/rsa.3240060204).
- Molloy, Michael and Bruce Reed (1998). “The Size of the Giant Component of a Random Graph with a Given Degree Sequence”. In: *Combinatorics, Probability and Computing* 7.3, pp. 295–305. DOI: [10.1017/S0963548398003526](https://doi.org/10.1017/S0963548398003526).
- Mossong, Joël, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinska, and W. John Edmunds (2008). “Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases”. In: *PLoS Medicine* 5.3. Ed. by Steven Riley, e74. DOI: [10.1371/journal.pmed.0050074](https://doi.org/10.1371/journal.pmed.0050074).
- Nagel, Kai, Christian Rakow, and Sebastian A. Müller (2021). “Realistic Agent-Based Simulation of Infection Dynamics and Percolation”. In: *Physica A: Statistical Mechanics and its Applications* 584, p. 126322. DOI: [10.1016/j.physa.2021.126322](https://doi.org/10.1016/j.physa.2021.126322).

- Newman, Mark E. J. (2003). "The Structure and Function of Complex Networks". In: *SIAM Review* 45.2, pp. 167–256. DOI: [10.1137/S003614450342480](https://doi.org/10.1137/S003614450342480).
- Newman, Mark E. J. (2009). "Random Graphs with Clustering". In: *Physical Review Letters* 103.5, p. 058701. DOI: [10.1103/PhysRevLett.103.058701](https://doi.org/10.1103/PhysRevLett.103.058701).
- Newman, Mark E. J., Steven H. Strogatz, and Duncan J. Watts (2001). "Random Graphs with Arbitrary Degree Distributions and Their Applications". In: *Physical Review E* 64.2, p. 026118. DOI: [10.1103/PhysRevE.64.026118](https://doi.org/10.1103/PhysRevE.64.026118).
- Newman, Mark E. J., Duncan J. Watts, and Steven H. Strogatz (2002). "Random Graph Models of Social Networks". In: *Proceedings of the National Academy of Sciences* 99.suppl.1, pp. 2566–2572. DOI: [10.1073/pnas.012582999](https://doi.org/10.1073/pnas.012582999).
- Pardoux, Etienne (2020). "Moderate Deviations and Extinction of an Epidemic". In: *Electronic Journal of Probability* 25.none, pp. 1–27. DOI: [10.1214/20-EJP428](https://doi.org/10.1214/20-EJP428).
- Pazy, Amnon (2012). *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Place of publication not identified: Springer. ISBN: 978-1-4612-5563-5.
- Pellis, Lorenzo, Neil M. Ferguson, and Christophe Fraser (2009). "Threshold Parameters for a Model of Epidemic Spread among Households and Workplaces". In: *Journal of The Royal Society Interface* 6.40, pp. 979–987. DOI: [10.1098/rsif.2008.0493](https://doi.org/10.1098/rsif.2008.0493).
- Pellis, Lorenzo, Neil M. Ferguson, and Christophe Fraser (2011). "Epidemic Growth Rate and Household Reproduction Number in Communities of Households, Schools and Workplaces". In: *Journal of Mathematical Biology* 63.4, pp. 691–734. DOI: [10.1007/s00285-010-0386-0](https://doi.org/10.1007/s00285-010-0386-0).
- Peter, Olumuyiwa James, Sumit Kumar, Nitu Kumari, Festus Abiodun Oguntolu, Kayode Oshinubi, and Rabi Musa (2022). "Transmission Dynamics of Monkeypox Virus: A Mathematical Modelling Approach". In: *Modeling Earth Systems and Environment* 8.3, pp. 3423–3434. DOI: [10.1007/s40808-021-01313-2](https://doi.org/10.1007/s40808-021-01313-2).
- Prodhomme, Adrien (2023). "Strong Gaussian Approximation of Metastable Density-Dependent Markov Chains on Large Time Scales". In: *Stochastic Processes and their Applications* 160, pp. 218–264. DOI: [10.1016/j.spa.2023.01.018](https://doi.org/10.1016/j.spa.2023.01.018).
- Ram, Vishaal and Laura P. Schaposnik (2021). "A Modified Age-Structured SIR Model for COVID-19 Type Viruses". In: *Scientific Reports* 11, p. 15194. DOI: [10.1038/s41598-021-94609-3](https://doi.org/10.1038/s41598-021-94609-3).
- Richard, Quentin, Samuel Alizon, Marc Choisy, Mircea T. Sofonea, and Ramsès Djidjou-Demasse (2021). "Age-Structured Non-Pharmaceutical Interventions for Optimal Control of COVID-19 Epidemic". In: *PLoS computational biology* 17.3, e1008776. DOI: [10.1371/journal.pcbi.1008776](https://doi.org/10.1371/journal.pcbi.1008776).
- Rimbaud, Loup, Claude Bruchou, Sylvie Dallot, David R. J. Pleydell, Emmanuel Jacquot, Samuel Soubeyrand, and Gaël Thébaud (2018). "Using Sensitivity Analysis to Identify Key Factors for the Propagation of a Plant Epidemic". In: *Royal Society Open Science* 5.1, p. 171435. DOI: [10.1098/rsos.171435](https://doi.org/10.1098/rsos.171435).
- Riquelme, Fabián, Ana Aguilera, and Alonso Inostrosa-Psijas (2021). "Contagion Modeling and Simulation in Transport and Air Travel Networks During the COVID-19 Pandemic: A Survey". In: *IEEE Access* 9, pp. 149529–149541. DOI: [10.1109/ACCESS.2021.3123892](https://doi.org/10.1109/ACCESS.2021.3123892).
- Roelly-Coppoletta, Sylvie (1986). "A Criterion of Convergence of Measure-valued Processes: Application to Measure Branching Processes". In: *Stochastics: An International Journal of Probability and Stochastic Processes* 17.1-2, pp. 43–65. DOI: [10.1080/17442508608833382](https://doi.org/10.1080/17442508608833382).

- Romano, Silvio, Annalisa Fierro, and Antonella Liccardo (2020). "Beyond the Peak: A Deterministic Compartment Model for Exploring the Covid-19 Evolution in Italy". In: *PLOS ONE* 15.11, e0241951. DOI: [10.1371/journal.pone.0241951](https://doi.org/10.1371/journal.pone.0241951).
- Roux, Jonathan, Clément R. Massonnaud, Vittoria Colizza, Simon Cauchemez, and Pascal Crépey (2023). "Modeling the Impact of National and Regional Lockdowns on the 2020 Spring Wave of COVID-19 in France". In: *Scientific Reports* 13.1, p. 1834. DOI: [10.1038/s41598-023-28687-w](https://doi.org/10.1038/s41598-023-28687-w).
- Saltelli, Andrea, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola (2007). "Variance-Based Methods". In: *Global Sensitivity Analysis. The Primer*. John Wiley & Sons, Ltd. Chap. 4, pp. 155–182. ISBN: 978-0-470-72518-4. DOI: [10.1002/9780470725184.ch4](https://doi.org/10.1002/9780470725184.ch4).
- Saltelli, Andrea, Stefano Tarantola, Francesca Campolongo, and Marco Ratto (2002). *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley & Sons, Ltd. ISBN: 978-0-470-87095-2.
- Sherborne, Neil, Joel C. Miller, Konstantin B. Blyuss, and Istvan Z. Kiss (2018). "Mean-Field Models for Non-Markovian Epidemics on Networks". In: *Journal of Mathematical Biology* 76.3, pp. 755–778. DOI: [10.1007/s00285-017-1155-0](https://doi.org/10.1007/s00285-017-1155-0).
- Silhol, Romain and Pierre-Yves Boëlle (2011). "Modelling the Effects of Population Structure on Childhood Disease: The Case of Varicella". In: *PLoS Computational Biology* 7.7, e1002105. DOI: [10.1371/journal.pcbi.1002105](https://doi.org/10.1371/journal.pcbi.1002105).
- Simoy, Mario Ignacio and Juan Pablo Aparicio (2021). "Socially Structured Model for COVID-19 Pandemic: Design and Evaluation of Control Measures". In: *Computational and Applied Mathematics* 41.1, p. 14. DOI: [10.1007/s40314-021-01705-1](https://doi.org/10.1007/s40314-021-01705-1).
- Solomonoff, Ray and Anatol Rapoport (1951). "Connectivity of Random Nets". In: *Bulletin of Mathematical Biophysics* 13, pp. 107–117. DOI: [10.1007/BF02478357](https://doi.org/10.1007/BF02478357).
- St-Onge, Guillaume, Laurent Hébert-Dufresne, and Antoine Allard (2023). *Heterogeneous Transmission in Groups Induces a Superlinear Force of Infection*. DOI: [10.48550/arXiv.2302.13358](https://doi.org/10.48550/arXiv.2302.13358). arXiv: [2302.13358](https://arxiv.org/abs/2302.13358).
- Stark, Dudley (2004). "The Vertex Degree Distribution of Random Intersection Graphs". In: *Random Structures & Algorithms* 24.3, pp. 249–258. DOI: [10.1002/rsa.20005](https://doi.org/10.1002/rsa.20005).
- Stehlé, Juliette, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, and Philippe Vanhems (2011). "High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School". In: *PLOS ONE* 6.8, e23176. DOI: [10.1371/journal.pone.0023176](https://doi.org/10.1371/journal.pone.0023176).
- Strömgren, Magnus, Einar Holm, Örjan Dahlström, Joakim Ekberg, Henrik Eriksson, Armin Spreco, and Toomas Timpka (2017). "Place-Based Social Contact and Mixing: A Typology of Generic Meeting Places of Relevance for Infectious Disease Transmission". In: *Epidemiology and Infection* 145.12, pp. 2582–2593. DOI: [10.1017/S0950268817001169](https://doi.org/10.1017/S0950268817001169).
- Thanh, Vo Hong and Corrado Priami (2015). "Simulation of Biochemical Reactions with Time-Dependent Rates by the Rejection-Based Algorithm". In: *The Journal of Chemical Physics* 143.5, p. 054104. DOI: [10.1063/1.4927916](https://doi.org/10.1063/1.4927916).
- Timpka, Toomas, Henrik Eriksson, Einar Holm, Magnus Strömgren, Joakim Ekberg, Armin Spreco, and Örjan Dahlström (2016). "Relevance of Workplace Social Mixing during Influenza Pandemics: An Experimental Modelling Study of Workplace Cultures". In: *Epidemiology and Infection* 144.10, pp. 2031–2042. DOI: [10.1017/S0950268816000169](https://doi.org/10.1017/S0950268816000169).



- Tomori, Damilola Victoria, Johannes Horn, Nicole Rübsamen, Sven Kleine Bardenhorst, Christoph Kröger, Veronika K. Jaeger, André Karch, and Rafael Mikolajczyk (2022). "Sexual Contact Patterns in High-Income Countries—A Comparative Analysis Using Data From Germany, the United Kingdom, and the United States". In: *Frontiers in Epidemiology* 2. DOI: [10.3389/fepid.2022.858789](https://doi.org/10.3389/fepid.2022.858789).
- Tran, Viet Chi (2006). "Modèles particuliers stochastiques pour des problèmes d'évolution adaptative et pour l'approximation de solutions statistiques". PhD thesis. Université de Nanterre - Paris X. Accessible on HAL: tel-00125100.
- Tran, Viet Chi (2014). "Une ballade en forêts aléatoires". Habilitation a Diriger les Recherches. Université Lille 1. Accessible on HAL: tel-01087229.
- Trapman, Pieter, Frank Ball, Jean-Stephane Dhersin, Viet Chi Tran, Jacco Wallinga, and Tom Britton (2016). "Inferring  $R_0$  in Emerging Epidemics—the Effect of Common Population Structure Is Small". In: *Journal of The Royal Society Interface* 13.20160288. DOI: [10.1098/rsif.2016.0288](https://doi.org/10.1098/rsif.2016.0288).
- Van Kerckhove, Kim, Niel Hens, W. John Edmunds, and Ken T. D. Eames (2013). "The Impact of Illness on Social Networks: Implications for Transmission and Control of Influenza". In: *American Journal of Epidemiology* 178.11, pp. 1655–1662. DOI: [10.1093/aje/kwt196](https://doi.org/10.1093/aje/kwt196).
- Volz, Erik M. (2008). "SIR Dynamics in Random Networks with Heterogeneous Connectivity". In: *Journal of Mathematical Biology* 56.3, pp. 293–310. DOI: [10.1007/s00285-007-0116-4](https://doi.org/10.1007/s00285-007-0116-4).
- Volz, Erik M., Joel C. Miller, Alison Galvani, and Lauren Ancel Meyers (2011). "Effects of Heterogeneous and Clustered Contact Patterns on Infectious Disease Dynamics". In: *PLoS Computational Biology* 7.6. Ed. by Mark M. Tanaka, e1002042. DOI: [10.1371/journal.pcbi.1002042](https://doi.org/10.1371/journal.pcbi.1002042).
- Watts, Duncan J. and Steven H. Strogatz (1998). "Collective Dynamics of 'Small-World' Networks". In: *Nature* 393.6684, pp. 440–442. DOI: [10.1038/30918](https://doi.org/10.1038/30918).
- Wilkinson, Robert R., Frank G. Ball, and Kieran J. Sharkey (2017). "The Relationships between Message Passing, Pairwise, Kermack–McKendrick and Stochastic SIR Epidemic Models". In: *Journal of Mathematical Biology* 75.6-7. DOI: [10.1007/s00285-017-1123-8](https://doi.org/10.1007/s00285-017-1123-8).
- Wilkinson, Robert R. and Kieran J. Sharkey (2014). "Message Passing and Moment Closure for Susceptible-Infected-Recovered Epidemics on Finite Networks". In: *Physical Review E* 89.2, p. 022808. DOI: [10.1103/PhysRevE.89.022808](https://doi.org/10.1103/PhysRevE.89.022808).
- Wilson, Fiona (2015). "Romantic Relationships at Work: Why Love Can Hurt". In: *International Journal of Management Reviews* 17.1, pp. 1–19. DOI: [10.1111/ijmr.12034](https://doi.org/10.1111/ijmr.12034).
- Wong, Kerry L. M., Amy Gimma, Pietro Coletti, Christel Faes, Philippe Beutels, Niel Hens, Veronika K. Jaeger, Andre Karch, Helen Johnson, Wjohn Edmunds, and Christopher I. Jarvis (2023). "Social Contact Patterns during the COVID-19 Pandemic in 21 European Countries – Evidence from a Two-Year Study". In: *BMC Infectious Diseases* 23, p. 268. DOI: [10.1186/s12879-023-08214-y](https://doi.org/10.1186/s12879-023-08214-y).
- World Bank (2024). *World Bank Open Data*. <https://data.worldbank.org>.
- Wu, Jianyong, Radhika Dhingra, Manoj Gambhir, and Justin V. Remais (2013). "Sensitivity Analysis of Infectious Disease Models: Methods, Advances and Their Application". In: *Journal of The Royal Society Interface* 10.86, p. 20121018. DOI: [10.1098/rsif.2012.1018](https://doi.org/10.1098/rsif.2012.1018).



**Titre :** Approximation de modèles stochastiques d'épidémies sur grands graphes multi-niveaux

**Mots clés :** Processus épidémique, graphes aléatoires, deux niveaux de mélange, réduction de modèle

**Résumé :** Nous étudions un modèle *SIR* à deux niveaux de mélange, à savoir un niveau global uniformément mélangeant, et un niveau local divisé en deux couches de contacts au sein des foyers et lieux de travail, respectivement. Nous cherchons à développer des modèles réduits qui approchent bien cette dynamique épidémique, tout en étant plus maniables pour l'analyse numérique et/ou théorique.

D'abord, nous analysons l'impact épidémique de la distribution des tailles des lieux de travail. Notre étude par simulations montre que, si la moyenne de la distribution des tailles de lieux de travail est fixée, sa variance est un bon indicateur de son influence sur des caractéristiques clés de l'épidémie. Cela nous permet de proposer des stratégies de télétravail efficaces. Ensuite, nous montrons qu'un modèle *SIR* déterministe, uniformément mélangeant, calibré sur le taux de croissance épidémique fournit une approximation parcimonieuse de l'épidémie.

Néanmoins, la précision de ce modèle réduit décroît au cours du temps et n'a pas de garanties théoriques. Nous étudions donc la limite grande population du modèle stochastique à foyers et lieux de travail, que nous formalisons comme un processus à va-

leur mesure dont l'espace de types est continu. Nous établissons sa convergence vers l'unique solution déterministe d'une équation à valeur mesure. Dans le cas où les périodes infectieuses sont exponentiellement distribuées, une réduction plus forte vers un système dynamique fini-dimensionnel est obtenue.

De plus, une étude de sensibilité nous permet de comprendre l'impact des paramètres du modèle sur la performance de ces deux modèles réduits. Nous montrons que la limite grande population du modèle foyer-travail permet de bien approcher l'épidémie, même si certaines hypothèses sur le réseau de contact sont relâchées. De même, nous quantifions l'impact des paramètres épidémiques sur la capacité du modèle réduit uniformément mélangeant à prédire des caractéristiques clés de l'épidémie.

Enfin, nous considérons plus généralement des processus de population densité-dépendants. Nous établissons une formule tous-pour-un qui réduit la lignée typique d'un individu échantillonné à un processus spinal inhomogène en temps. Par ailleurs, nous quantifions par couplage la convergence en grande population d'une construction spinale.

**Title :** Approximation of stochastic models for epidemics on large multi-level graphs

**Keywords :** Epidemic processes, random graphs, two levels of mixing, model reduction

**Abstract :** We study an *SIR* model with two levels of mixing, namely a uniformly mixing global level, and a local level with two layers of household and workplace contacts, respectively. More precisely, we aim at proposing reduced models which approximate well the epidemic dynamics at hand, while being more prone to mathematical analysis and/or numerical exploration.

We investigate the epidemic impact of the workplace size distribution. Our simulation study shows that if the average workplace size is kept fixed, the variance of the workplace size distribution is a good indicator of its influence on key epidemic outcomes. In addition, this allows to design an efficient teleworking strategy. Next, we demonstrate that a deterministic, uniformly mixing *SIR* model calibrated using the epidemic growth rate yields a parsimonious approximation of the household-workplace model.

However, the accuracy of this reduced model deteriorates over time and lacks theoretical guarantees. Hence, we study the large population limit of the stochastic household-workplace model, which we

formalize as a measure-valued process with continuous state space. In a general setting, we establish convergence to the unique deterministic solution of a measure-valued equation. In the case of exponentially distributed infectious periods, a stronger reduction to a finite dimensional dynamical system is obtained.

Further, in order to gain a finer insight on the impact of the model parameters on the performance of both reduced models, we perform a sensitivity study. We show that the large population limit of the household-workplace model can approximate well the epidemic even if some assumptions on the contact network are relaxed. Similarly, we quantify the impact of epidemic parameters on the capacity of the uniformly mixing reduced model to predict key epidemic outcomes.

Finally, we consider density-dependent population processes in general. We establish a many-to-one formula which reduces the typical lineage of a sampled individual to a time-inhomogeneous spinal process. In addition, we use a coupling argument to quantify the large population convergence of a spinal process.