

**MATHEMATICAL MODELS
FOR EMERGING INFECTIONS
IN SOCIALLY STRUCTURED POPULATIONS:
THE PRESENCE OF HOUSEHOLDS
AND WORKPLACES**

LORENZO PELLIS

SUPERVISORS:

PROF. CHRISTOPHE FRASER

PROF. NEIL M. FERGUSON

IMPERIAL COLLEGE LONDON

DEPARTMENT OF INFECTIOUS DISEASE EPIDEMIOLOGY
DIVISION OF EPIDEMIOLOGY, PUBLIC HEALTH AND PRIMARY CARE
FACULTY OF MEDICINE

JULY 2009

Submitted in partial fulfillment of the Doctor of Philosophy (Ph.D.) degree

Abstract

This thesis is concerned with the description and analysis of a stochastic model for the spread of a directly transmissible infection, leading to permanent immunity after recovery, in a fully susceptible population with a social structure characterised by the presence of households and workplaces. The model considered is highly idealised, but contains the key factors affecting the spread of a directly transmissible infection, namely those environments where frequent and intense contacts are most likely.

Important analytical insights include the definition of a novel household reproduction number R_H , representing the average number of households infected by a single household, which is shown to overcome some of the limitations of a previously defined reproduction number and the development of a methodology for the approximate computation of the real-time growth rate, which is then used for the estimation of R_H from the real-time growth rate.

An efficient stochastic simulator is described and is used to gain understanding of the role that local saturation effects within workplaces play in shaping the epidemic spread and to investigate the reliability of estimates of R_0 and the average epidemic final size from the real-time growth rate when the presence of the social structure is neglected.

The methodologies are applied to the case of pandemic influenza: its relatively low infectiousness suggests that estimation of these key epidemiological quantities is surprisingly accurate when the social structure is neglected and that the additional presence of spatial constraints implying geographically localised transmission has negligible effect on the overall epidemic dynamics. Despite the lack of reliable data concerning workplaces, a realistic range of possible values for R_H is identified, but the efficacy of school closure in reducing transmission appears to be difficult to quantify because of the unknown impact it has on transmission in other workplace environments.

Contents

Abstract	3
Acknowledgements	15
Main notational conventions	19
Description and structure of this thesis	25
Research description	25
Aims and modelling framework	25
Structure of the thesis	27
1 Introduction	31
1.1 Mathematical models	31
1.2 Mathematical models in infectious disease epidemiology	32
1.2.1 Dynamical models	32
1.2.2 Purposes of epidemic modelling	33
1.2.3 Deterministic and stochastic models	34
1.2.4 Simple model and complex simulations	37
1.2.5 Network models	39
2 Background	45
2.1 Models for transmission at the individual level	45
2.1.1 The concept of disease history	45

2.1.2	Density-dependent and frequency-dependent transmission	47
2.1.3	Non-random disease history	49
2.1.4	Random disease history	51
2.2	Unstructured large populations	53
2.2.1	Basic reproduction number	53
2.2.2	Final size	55
2.2.3	Real-time growth rate	59
2.2.4	Basic reproduction number and real-time growth rate	62
2.3	Multitype models	64
2.3.1	Transmission	65
2.3.2	Threshold parameters	66
2.3.3	Final size	69
2.3.4	Real-time growth rate	69
2.3.5	Separable mixing	70
2.3.6	Correlations between infector and infectee	73
2.4	Households models	74
2.4.1	Introduction	74
2.4.2	Comparison with multitype models	75
2.4.3	Brief summary of the most important results	75
2.4.4	Model definition	77
2.4.5	Reproduction numbers	80
2.4.6	Average final size	83
2.4.7	Interventions targeted at households	84
2.4.8	Extension to variable household size	85
2.5	Other socially structured models	86
2.5.1	Metapopulation models	86
2.5.2	Overlapping groups models	87
2.5.3	Hierarchical network models	88

3	Small populations	89
3.1	Introduction	89
3.2	Consequences of Ludwig's result	92
3.3	Problem description and terminology	94
3.4	Models for the infectious contact process	96
3.4.1	Standard <i>SIR</i> epidemic model	97
3.4.2	Models with time varying infectivity	99
3.4.3	Variable susceptibility and infectivity	100
3.5	A particular case	102
3.6	Validity of the result	105
3.7	Computation of the final size distribution	107
3.8	Summary and further comments	109
4	Households-workplaces model	111
4.1	Introduction	111
4.2	Model definition	112
4.2.1	Basic model assumptions	112
4.2.2	Assumptions concerning transmission	113
4.2.3	Assumptions concerning the social structure	115
4.3	Threshold parameters	117
4.3.1	Different types of infectives	117
4.3.2	Clump reproduction number	118
4.3.3	Household reproduction number	120
4.3.4	Other reproduction numbers	123
4.3.5	Comparison between the threshold parameters	124
4.3.6	Similarities with other results	126
4.4	Average final size	126
4.5	Vaccination	127
4.6	School closure and other control policies	130

4.7	Proof of Theorem 4.5.1	131
4.8	Variable household and workplace sizes	134
4.9	Sensitivity of results to simplifying assumptions	139
4.10	Summary, comments, limitations and future research	144
5	Real-time growth rate	151
5.1	Introduction	151
5.2	Models with no local saturation	152
5.3	Households model	155
5.3.1	Real-time growth rate for households	156
5.3.2	Real-time growth rate for individuals	157
5.4	Households-workplaces model	158
5.4.1	Fixed household and workplace sizes	158
5.4.2	Variable household and workplace sizes	160
5.5	Exact dynamics in a small population	163
5.5.1	Continuous time Markov chains	163
5.5.2	Application to epidemics in a closed group	164
5.5.3	Application to the households model	166
5.5.4	Application to the households-workplaces model	169
5.6	Approximate dynamics in a small population	171
5.6.1	Epidemic in a closed group	171
5.6.2	Application to the households model	173
5.6.3	Application to the households-workplaces model	174
5.7	Numerical investigation	176
5.8	Factorisation of the global infectivity profile of a group	181
5.9	Summary and conclusions	185
6	Simulation of stochastic epidemics	187
6.1	Epidemic simulator	187

6.1.1	Introduction	187
6.1.2	Homogeneously mixing population	189
6.1.3	The simulation output	194
6.1.4	Households-workplaces model	197
6.1.5	A more realistic social structure	200
6.1.6	Efficient approximations	201
6.2	The role of workplaces in the epidemic spread	203
6.2.1	Motivation	203
6.2.2	Effects of saturation	206
6.2.3	The case of households and workplaces	209
6.2.4	Ignoring the social structure	213
6.3	Summary and conclusions	228
7	Application to influenza	233
7.1	Approximate method for the estimation of R_H	233
7.1.1	Upper bound estimation	234
7.1.2	Lower bound estimation	237
7.1.3	The case of influenza	238
7.2	Spatial constraints on the social structure	243
7.2.1	Motivation	243
7.2.2	The social structure	243
7.3	School closure	249
7.3.1	School closure only	250
7.3.2	School and workplace closure	252
7.4	Summary and conclusions	253
8	Conclusions and future research	259
8.1	A look backwards	259
8.2	A look forward	261

List of Figures

2.1	Epidemic average final size in an infinite population	58
2.2	Social structure and epidemic spread in the households model	78
3.1	Real infections and infectious contacts in a group of size 3	95
3.2	All possible epidemics in a group of size 3: rank and generations	103
4.1	Social structure and epidemic spread in the households-workplaces model .	114
4.2	Construction of R_H : a two-type next generation matrix for households . .	122
4.3	Level curves of R_* and R_H	125
4.4	Secure vaccination coverage: reduction in R_H	129
4.5	Effects of vaccination on final size and infection trees	132
4.6	Average final size with frequency and density dependent transmission . . .	137
5.1	Between-states transition rates in a Markovian epidemic model	165
5.2	All states and transition rates of a Markovian epidemic model	166
5.3	Transient states and transition rates of a Markovian epidemic model	167
5.4	Analytical approximation to the rate of infection of the secondary case . .	178
5.5	Analytical approximation to the rate of new infections for $R_0 = 1.5$	179
5.6	Analytical approximation to the rate of new infections for $R_0 = 1.5$	180
6.1	Simulator: linked list of events and hash table	193
6.2	Simulation output: r , R_0 and R_e	195
6.3	Simulation output: μ_H and R_H	199

6.4	Fixed R_H : R_w and r for the <i>exp</i> disease history (extreme range)	210
6.5	Fixed R_H : R_w for the <i>gamma</i> and the <i>TVI</i> disease history (extreme range)	211
6.6	Fixed R_H : comparison of r for all disease history (extreme range)	212
6.7	Fixed R_H : R_w and r for the <i>exp</i> disease history (normal range)	213
6.8	$R_0(r)$ for <i>TVI</i> (stratified)	214
6.9	$R_0(r)$ for <i>TVI</i> (all sizes)	215
6.10	$R_0(r)$ for <i>TVI</i> (approximation)	216
6.11	$R_0(r)$ for <i>exp</i> (stratified)	217
6.12	$R_0(r)$ for <i>exp</i> (approximation)	217
6.13	$R_0(r)$ for <i>TVIexp</i> (stratified)	219
6.14	$R_0(r)$ for <i>TVIexp</i> (approximation)	219
6.15	$afs(r)$ for <i>TVI</i> (stratified)	221
6.16	$afs(r)$ for <i>TVI</i> with constant R_H (stratified)	221
6.17	$afs(r)$ for <i>TVI</i> (approximation)	222
6.18	$afs(r)$ for <i>TVI</i> with constant R_H (approximation)	223
6.19	$afs(r)$ for <i>exp</i> (stratified)	224
6.20	$afs(r)$ for <i>exp</i> (all sizes)	225
6.21	$afs(r)$ for <i>exp</i> (approximation)	226
6.22	$afs(r)$ for <i>TVIexp</i> (stratified)	227
6.23	$afs(r)$ for <i>TVIexp</i> (approximation)	227
7.1	$R_H(r)$ for schools of size 20 and workplaces of size 1	240
7.2	$R_H(r)$ for schools of size 20 and workplaces of size 4	241
7.3	$R_H(r)$ for schools and workplaces of size 20 (population of children)	242
7.4	Commuting distances on a random network	244
7.5	Example of commuting distances on a moderately localised spatial network	245
7.6	Example of commuting distances on a highly localised spatial network	246
7.7	Epidemic on random and spatial network (with global infections)	247
7.8	Epidemic on random and spatial network (without global infections)	248

7.9	Effect of closing schools without affecting workplace transmission	251
7.10	Effect of closing school and stopping workplace transmission	252

Acknowledgements

It is almost impossible to name and thank all those that have helped me during this time without the risk of neglecting to acknowledge some: partly because there are many of them; partly because the exhausting task of writing a thesis cannot be disentangled from the rest of one's life and any form of support, even the most indirect, plays an important role in the final outcome. Although this process offered many important opportunities for maturation, it was nevertheless a daunting process, which I would not have been able to face without the help of all the people around me.

To all of you go my most profound thanks. Although it is not easy for me to accept, I have to admit that this thesis would have not been possible without your help, and is therefore yours at least as much as mine.

I am sure that many names will be missing, and I apologise for that, but I still would like to try to mention explicitly some of them.

The greatest thanks go to my supervisors Prof. Christophe Fraser and Prof. Neil Ferguson.

Thank you, Christophe, for all the care you showed for me, both from the academic and the human point of view: thanks for the confidence you had in me, right from the first moment we met; the extremely competent and stimulating scientific supervision; the thorough corrections and comments on the thesis drafts you found time to write when you were most busy, and the unsociable hours you were available to meet me at; the moral support I received from you when most needed; the precious suggestions, always at the right moment; the view on life you often shared with me; the honesty you showed when less nice things needed to be said; the answer, always ready, to all those questions that mattered so much to me and that were totally unrelated to my research project. You have been and will remain for me much more than a supervisor.

Thank you, Neil, for your full support: from the funding opportunity that allowed me to start working with you, to the time you were able to find for me, in your extremely busy

life; from the efficiency you showed in helping me during those quick meetings (few words always spot on) to the suggestions you gave me when I showed interest in topics other than my project; from your long-term view on research in general and the research field in particular, which I found very enriching, to your extremely focused solutions to contingent problems I needed to solve in virtually no time. You have been and will remain an amazing example of scientific authority, efficient supervisor, team leader and far-sighted researcher.

Special thanks also go to all my work colleagues who helped me with mathematical and modelling corrections, helpful comments and support. In particular: Dr. Pete Dodd, for the extremely competent mathematical and linguistic suggestions, for the constant moral support and for representing a continuous stimulus for research ideas, collaboration opportunities, or simply opinion exchanges; Dr. Paul Parham, for the always-ready linguistic help, the patience you had (and have) in checking and re-checking bits I am never sure about, and the ever-present moral support; Dr. James Truscott, for the patience you showed in listening to all my non-sense as well as for all those one-shot insights you repeatedly gave me just passing next to my desk (you now probably think twice before exiting your office, for the fear of being stopped once more); Dr. Simon Cauchemez, for all the mathematical and modelling suggestions and for the comprehension towards all the doubts on my future I talked with you about; and Dr. Chris Rhodes and Prof. Henrik Jensen, for the time spent listening to me and offering suggestions and advice.

Many others have assisted me in this laborious task, and although their help was less related to the academic aspects of this thesis, it was by no means less important.

The greatest thanks go undoubtedly to my fiancée Laura, who had and still has the unfortunate role of sharing with me all the moments when I am most stressed, boring and difficult to bear. Your patience has been unlimited, your support immeasurable and I will never be able to thank you enough for that.

I will also be unable to thank enough my parents and my entire family which, despite the distance, have been always present when support, help or understanding was most needed.

I would also like to thank all those in my department that have been always ready to listen to me: Dr. Déirdre Hollingsworth, Dr. Tini Garske, Ms Rosalind Eggo, Miss Íde Cremin, Dr. Judith Legrand, Ms Kimberly Marsh, Dr. Federico Calboli and many others, including Ms Andrea Johnson, Dr. Steven Dunstan, Mr Dipen Mandalia, Ms Francesca Tracey, Ms Hafiza Bibi, Ms Karen Harvey, Ms Katherine England, Ms Jane Lillywhite, Dr. Emily Lyons and the rest of the administrative staff.

I would also like to thank my university professors in Italy, among which Prof. Lucio Torelli, Prof. Aldo Nocentini, Prof. Alfredo Bellen, Prof. Marino Zennaro, Prof. Daniele Del Santo and Prof. Emilia Mezzetti, who supported my choice of starting this new experience abroad, but who have been ready to listen to me and help me in whatever way they could every time I asked them.

Finally I would like to thank all my friends, for their help and patience: among the ones I left in Italy, Laura Derganz, Fabio Barnaba, Antonella Poli and the others; among the ones I met here, Dr. Thomas Alarcon, Dr. Renaud Lambiotte and Dr. Elsa Arcaute from the Math Institute; among the new friends I met in Oxford, Dr. Benjamin Vaughan, Guido Klingbeil and Ornella Cominetti; and also Saverio Reina, Cesare Tronci, Marco Pautasso, David and Gebina, Paul, Matt and all the others from Clayponds, the orchestra, the choir, St. Mary's Church and the Ealing Chamber Music Club, too many to name.

Thanks to you all for mattering so much to me.

Main notational conventions used in the thesis

t	Absolute time, set to $t = 0$ at the beginning of the epidemic.
τ	Time since infection, i.e. a relative time variable which, for each infective, is set to $\tau = 0$ at the time of infection.
n	Size of a finite closed group of individuals. When the population is large, N is used instead.
β or $\beta(\tau)$	Constant or time varying rate at which a single infective makes infectious contacts with all other individuals, often referred to as one-to-all infection rate. $\beta(\tau)$ is also called infectivity profile.
Φ	Total infectivity (a random variable, in general) from a single infective to all other individuals. It is given by β times the length of the infectious period in the standard <i>SIR</i> model, or more generally by the area under the curve $\beta(\tau)$ (possibly a random trajectory).
R_0	Basic reproduction number, given by $R_0 = \mathbb{E}[\Phi]$.
λ or $\lambda(\tau)$	Constant or time varying rate at which a single infective makes infectious contacts with any other selected individual, often referred to as one-to-one infection rate.
Λ	Total infectivity (a random variable, in general) from a single infective to any other selected individual. It is given by λ times the length of the infectious period in the standard <i>SIR</i> model, or more generally by the area under the curve $\lambda(\tau)$ (possibly a random trajectory). In a group of size n , $\lambda = \beta/(n - 1)$ or $\lambda(\tau) = \beta(\tau)/(n - 1)$.
T_{max}	Upper bound for the length of the infectious period in the time-since-infection model, i.e. $\beta(\tau) = \omega(\tau) = 0$ for all $\tau > T_{max}$.
$\omega(\tau)$	Generation time distribution, defined as $\beta(\tau)/R_0$, i.e. normalised to have integral one. It represents the distribution of times between the infection of an individual and times at which he or she makes infectious contacts with other individuals.

q	Escaping probability for the simple Reed-Frost model. Also probability of not being vaccinated in the households or households-workplaces model.
Q	Random escaping probability for the randomised Reed-Frost model. Also the infinitesimal generator of a Markov chain (Q -matrix).
I	Random variable for the length of the infectious period in the standard <i>SIR</i> model. In the standard <i>SIR</i> model $\Lambda = \lambda I$, $\Phi = \beta I$.
ι	Average length of the infectious period in the standard <i>SIR</i> model, i.e. $\iota = \mathbb{E}[I]$.
σ	Standard deviation of the length of the infectious period in the standard <i>SIR</i> model, i.e. $\sigma^2 = \text{Var}(I)$.
E	Random variable for the length of the latent period added to the standard <i>SIR</i> model in order to obtain the standard <i>SEIR</i> model.
e	Average length of the latent period in the <i>SEIR</i> model, i.e. $e = \mathbb{E}[E]$.
ξ	Random offspring of a branching process.
X	Number of susceptibles in the population.
Y	Number of infectives in the population.
Z	Random proportional final size of an epidemic, i.e. random fraction ($Z \in [0, 1]$) of the initial susceptible population that is ultimately infected in the epidemic. Sometimes used for the number of recovered/immune individuals.
z	Average proportional final size of an epidemic, i.e. $z = \mathbb{E}[Z]$, expressed as a fraction of the initial susceptible population.
χ	Fraction of the population that is initially infected. When the epidemic in an infinite population starts with a finite number of initial infectives, as in most cases, $\chi = 0$.

ε	Fraction of the initial susceptible population that escapes infection: $\varepsilon = 1 - z$.
N	Total population size in a large population.
r	Real-time growth rate, also known as Malthusian parameter.
ν	Recovery rate, when the Markovian case of the standard <i>SIR</i> model with exponentially distributed length of infectious period is assumed.
T_g	Generation time, defined as $\int_0^{+\infty} \omega(\tau) d\tau$ in the model with time varying infectivity.
$L_f(s)$	Laplace transform of the function f calculated in s , i.e. $L_f(s) = \int_0^{+\infty} f(x) e^{-sx} dx$.
$K = (k_{ij})$	Next generation matrix.
$\rho(M)$	Spectral radius of the square matrix M .
a_i	Susceptibility of individuals of type i . Sometimes, susceptibility of individual i .
h_i	Proportion of individuals of type i in the population.
α_i	Availability-susceptibility of individuals of type i , i.e. $\alpha_i = h_i a_i$.
b_j or $b_j(\tau)$	Infectivity (or infectivity profile) of individuals of type j .
c_j	Contact rate of individuals of type j .
T	Type reproduction number.
S	Random size of a local susceptibility set.
μ, μ_H, μ_W	Average size of an epidemic, excluding the initial infective(s), in a generic group, a household or a workplace, respectively.
n_H, n_W	Sizes of households or workplaces.
m_H, m_W	Numbers of households or workplaces.
h_n	Proportion of households (or workplaces) of size n , i.e. probability that a randomly selected household (or workplace) has size n .
π_n	Probability that the household (or workplace) of a randomly selected individual has size n .

R_*	Household reproduction number in the households model. Clump reproduction number in the households-workplaces model.
R_H	Household reproduction number for the households-workplaces model.
R_I	Individual reproduction number.
R_{ij}	Average number of households infected through route i from a household that was infected through route j . Possible routes are: local ($i = L$), i.e. through a workplace, or global ($i = G$).
p_{crit}	Critical vaccination coverage for individuals, i.e. the (minimal) proportion of the population of individuals that needs to be vaccinated in order to prevent an epidemic.
p_{crit}^*	Critical vaccination coverage for households in the households model, i.e. the (minimal) fraction of households that needs to be vaccinated in order to prevent an epidemic (vaccination of a household is interpreted as vaccination of all its members). In the households-workplaces model it would refer to a critical vaccination coverage for clumps, but such a definition is not useful.
p_{crit}^H	Critical vaccination coverage for households in the households-workplaces model, i.e. the (minimal) fraction of households that needs to be vaccinated in order to prevent an epidemic (vaccination of a household is interpreted as vaccination of all its members).
p	Probability of an individual being successfully vaccinated (fully protected).
q	Probability of an individual not being vaccinated: $q = 1 - p$. Also escaping probability of the Reed-Frost model.
P	Random proportion of the population that is vaccinated.
Θ	Heaviside function: $\Theta_{[a,b]}(x) = 1$ when $x \in [a, b]$ and 0 otherwise. Also indicator function of a specific event.

$\gamma, \gamma_H, \gamma_W$	Rates at which new cases appear during an epidemic in a group, a household or a workplace, respectively.
τ_H, τ_W	Time-since-infection of a household or a workplace .
μ_i, μ_i^H, μ_i^W	Average number of cases occurring in generation i according to the Reed-Frost model in a closed group, a household or a workplace, respectively.
$\omega_{[i]}(\tau)$	i -th autoconvolution of the generation time distribution $\omega(\tau)$, i.e. $\omega_{[i]}(\tau) = \omega * \omega * \dots * \omega(\tau)$, i times.
α	Shape parameter of a gamma distribution.
δ	Scale parameter of a gamma distribution.
A	Random susceptibility.
B or $B(\tau)$	Random infectivity or random infectivity profile.
$\Omega(\tau)$	Random trajectory of the generation time distribution.
U	Generic random vector describing an epidemic outcome.
ζ	Generic measure.

Description and structure of this thesis

Research description

The main original contributions I brought to the field lie in the specific area of models for human directly transmissible infections. In particular, the models I considered are those able to capture some of the forms of heterogeneity in contact patterns that are due to the presence of a social structure and that are deemed to be relevant in shaping the spread of the infection.

Original contributions mainly involve the gain in understanding of the causal determinants of the behaviour of a model with households and workplaces, the definition of a new reproduction number with clear biological meaning for it, the study of its properties and the analysis of the effects of fundamental modelling assumptions on model outcomes, of the role of local saturation effects on the infection spread, of the qualitative impact of school closure in mitigating an epidemic and finally of the major factors that limit quantitative model predictability.

Aims and modelling framework

The specific aims of this thesis are:

1. the definition and description of a model for the spread of directly transmissible infections leading to permanent immunity after recovery in a population socially structured in households and workplaces: possible examples are given by influenza, severe acute respiratory syndrome (SARS), smallpox, measles and other childhood infections, respiratory syncytial virus (RSV), etc.;
2. the identification of key factors affecting the spread of epidemics;
3. the understanding of the role played by the presence of workplaces in the population, by their size and the magnitude of transmission occurring within them, on the overall spread of infections;

4. the investigation of the likely effect of closing schools as a form of control policy in the case of an outbreak of an emerging infection, such as influenza.

It is convenient to describe from the beginning the general framework to which results, comments and discussions presented in this thesis have to be interpreted, except when otherwise stated.

Following Anderson & May (1991), the framework most commonly adopted when discussing infectious diseases is based on the distinction between microparasites and macroparasites (Anderson & May, 1991; Diekmann & Heesterbeek, 2000; Keeling & Rohani, 2008). *Microparasites* can be thought of as those pathogens characterised by small dimensions and very high reproduction rates (usually viruses or bacteria) and for which the pathogen abundance within the infected host is not counted in discrete units, but is rather thought as a continuous variable. *Macroparasites*, instead, have larger size (usually multicellular organisms), longer and more complex life cycle and are reasonably counted in discrete units: the classical example is provided by worms.

In both cases, the pathogen abundance affects the severity of disease symptoms and the host infectivity, but from an epidemiological point of view there is an important distinction. On the one hand, the small dimensions and high replication rates characteristic of microparasites suggest that, after infection, the process occurring within an infected host develops independently of external events and subsequent re-infections of the same infected host (*superinfections*) have little or no impact on the progression of the infection experienced by the host. On the other hand, macroparasites are affected by external conditions and by the details of the pathogen life cycle: in particular, pathogen reproduction and repeated infections from outside have major effects on the infection stage and progression.

Of course, the distinction between the two classes is not a clear cut, and some pathogens do not comfortably fit in either of those. However, although the biological distinction between microparasites and macroparasites is unavoidably blurred, from the mathematical perspective the distinction between the two classes is clear: whatever the infectious agent, when the dynamics within an infected host are assumed to be decoupled from those of

the epidemic at the population level, the pathogen is a microparasite; otherwise it is a macroparasite. The epidemiology of microparasites is therefore much easier to study mathematically, as an infection event triggers an autonomous process within the host - a clock can be associated to each infective, which starts ticking at the time of infection and follows the within-host dynamics, regardless of outside events and subsequent reinfections. The distinction is therefore not so much a biological classification, but rather a modelling choice: the same pathogen may be either treated as a microparasite or a macroparasite, depending on the modelling assumptions which, in addition to the biology of the pathogen, may also depend on the particular research question addressed and the level of detail in (and quality of) the data available about the pathogen itself.

In this thesis, I deal exclusively with simple mathematical models for directly transmissible infections caused by microparasites. I also focus primarily on transmissible diseases leading to permanent immunity after recovery, for which the presence of households and workplaces is relevant: for example influenza, measles, chickenpox, SARS, RSV, etc.

Since the attention is mainly focused on single outbreaks of emerging infections, the population is usually considered to be large and fully susceptible and demographic effects are neglected, as they occur on a longer time-scale than that of the epidemic itself and therefore have limited effect on the infection spread.

Structure of the thesis

The Introduction presents a general overview of mathematical models and their application to infectious disease epidemiology, a brief description of the deterministic and stochastic modelling approaches and of different models that are not discussed in depth in this thesis.

Chapter 2 reviews some already known results, which build up in complexity to provide the tools for the further results presented in the next chapters. In particular it describes various modelling assumptions on disease transmission, asymptotic results in large populations for single-type models, multi-type models, households models and offers a glance

at other more complex models.

Both the Introduction and Chapter 2 can be skipped by the reader familiar with the subject, although the section referring to households models offers a deep analysis of a very specific model, which is then directly extended to the households-workplaces model of Chapter 4. It therefore provides a useful intermediate step along the path of increasing complexity.

Chapters 3, 4, 5, 6 and 7 contain mostly original contributions.

Chapter 3 focuses on small populations: although results are in general known in the stochastic epidemic community and can be found in the literature, a different point of view is proposed, with some original insights and a highlighted dependence of model properties on the fundamental assumptions.

Chapter 4 concerns the main model discussed in this thesis: a model with households and workplaces. The model has already been considered in Ball & Neal (2002), but here a novel household reproduction number is defined and the relationship it has with the critical vaccination coverage is discussed.

Chapter 5 contains a thorough discussion concerning the exact epidemic dynamics in a small population for the model with constant recovery rate (Markovian case) and the approximate epidemic dynamics in a small population for models with time-varying infectivity (non-Markovian). The dynamics in small populations are then used to compute the real-time growth rate both in the households and the households-workplaces models.

Chapter 6 reports the construction of a large scale individual-based stochastic simulation and uses it to explore the role that workplaces have in shaping the epidemic spread, under different modelling assumptions for the infectivity profile of individuals.

Chapter 7 focus specifically on influenza. Various modelling techniques are used to: quantify the household reproduction number, thus providing insight into the effort needed

to stop the epidemic when control policies are targeted at households; explore the effect of imposing some spatial constraints on the social structure; explore the effect of school closure as a form of control policy.

Finally, Chapter 8 is devoted to comments and future research trends.

Chapter 1

Introduction

1.1 Mathematical models

Throughout the centuries, but in particular in the last one, mathematics has been spreading like an infection in almost all the subjects one can think of, from physics to engineering, from finance to philosophy, from social sciences to psychology, from biology to epidemiology. Nowadays it permeates research as well as industry and services, and appears ubiquitously in our everyday life.

The main reason for this phenomenon is probably that, whenever quantitative information is needed, a form of measure is necessary and this is usually expressed in numbers. But, arguably, there is a more important reason: mathematics, in its every aspect, is a form of abstraction from reality that can be used to predict something unknown from something else which is known.

One of the most striking manifestations of this power is expressed in the construction of a mathematical model. A clear example is given by dynamical systems: knowledge (or presumed knowledge) of the rules governing a physical phenomenon allows its description, for example, with a system of equations. Given the system state in one point in time (e.g. at present), the equations can be solved for every time point in the future, and the solution can be interpreted as the future evolution of the physical phenomenon.

A mathematical model of a process or a phenomenon is a conceptual model used to represent the reality using the mathematical language. A mathematical model must be a simplified representation of reality for two main reasons: first of all, the real world is far too complex and rich of interconnected phenomena to allow for a full description; second, in order to be useful in practice, the model must be tractable in some way (either analytically or through a simulation).

However, the fundamental idea at the base of a mathematical model is that there is no need to model the full complexity of the real world: some factors play a more important role than others and one of the main purposes of a model is to point them out. Clearly the levels of importance of different factors range across the whole spectrum and therefore a model can be indefinitely refined by adding more and more of these. The model is then usually validated against available data: this procedure can highlight whether the assumptions made are effectively suitable or consistent, and if it is necessary to begin (or continue) a process of model refinement.

There is no single recipe to determine or prioritise the most “relevant” aspects of the process, in particular because their importance depends also on the particular question that one wants to answer: the decision concerning what to include, and therefore the kind of model to adopt, is entirely in the hands of the modeller and different modellers make different choices. For this reason, defining a model is not merely following a sequence of well defined tasks, which might be accomplished mechanically or blindly: it requires creativity and insight. And it is therefore extremely stimulating.

1.2 Mathematical models in infectious disease epidemiology

1.2.1 Dynamical models

There is a clearly marked difference in the field of epidemiology between infectious and non-infectious diseases. With non-infectious diseases, the models involved are mainly of

statistical type, since the main priority is to infer a causal interaction between variables from the observation of an association between these variables.

In infectious disease epidemiology, the modelling approach is more of a dynamical type: the causal relationship between various elements is assumed *a priori* and the state of the system described affects its future evolution. When the interest is focused on the dynamics of the system, this approach usually leads to differential or difference equations, the main mathematical tool that can “project” in the future a known (or presumed so) initial condition. Even when the main interest lies in parameter estimation, for which a wide variety of statistical tools is used, a dynamical model for transmission is still assumed, and its parameters are the target of the estimation procedure.

1.2.2 Purposes of epidemic modelling

Except for some important but sporadic attempts, as in the case of Bernoulli’s mathematical study of the efficacy of variolation techniques against smallpox in 1766 (e.g. see Bailey, 1975), it is only in the last century that the application of mathematical models to the field of infectious disease epidemiology started receiving attention in the literature (Bailey, 1975). In particular during the last few decades of the century (the HIV epidemic in the 1980s was arguably a fundamental factor in this process) the field has grown more and more conscious about the role that models play as fundamental investigative tools. Nowadays, it is widely recognised how important they are, for example, in:

1. understanding the mechanisms behind the spread of diseases (Anderson & May, 1991; Andersson & Britton, 2000; Diekmann & Heesterbeek, 2000);
2. highlighting the most important factors that drive the infection process (e.g. see Fraser *et al.*, 2004);
3. investigating the effectiveness of different control policies (e.g. see Longini *et al.*, 2004; Ferguson *et al.*, 2005);
4. assessing the efficacy of vaccines and prophylactic treatments (Halloran *et al.*, 2007);

5. designing more efficient observational studies (Datta *et al.*, 1999; Yang *et al.*, 2006);
6. planning mass vaccination campaigns (Grassly *et al.*, 2006) and
7. exploring future what-if scenarios (Ferguson *et al.*, 2001).

1.2.3 Deterministic and stochastic models

One possible model classification, which is relevant for the field of infectious disease epidemiology, consists in distinguishing between deterministic and stochastic models.

Deterministic models

Deterministic models are characterised by the fact that, once suitable initial conditions are specified, the future evolution of the system is uniquely determined.

They are usually easier to deal with than their stochastic counterparts and, for this reason, they are the first to appear in the field of epidemic modelling. Although some stochastic models had been developed by McKendrick or by Reed and Frost in the 1920s (e.g. see McKendrick, 1926, or Fine, 1977), the attention of the literature was mainly focused on deterministic models (Hamer, 1906; Ross, 1911; see Hethcote, 2000; Bailey, 1975, and references therein), which culminated in the milestone work by Kermack & McKendrick (1927).

Despite the subsequent development of stochastic models, deterministic models still remain the most widely used nowadays for quantitative practical purposes and when quick answers are needed, because of their mathematical tractability and the flexibility they show in allowing the model to be enriched by more realistic features that characterise the population and the infection under investigation (see, for example, Anderson & May, 1991; Diekmann & Heesterbeek, 2000; Halloran *et al.*, 1994; Brisson *et al.*, 2000). Examples of such additional features include: “passive immunity, gradual loss of vaccine and disease-acquired immunity, stages of infection, vertical transmission, disease vectors, macropar-

asitic loads, age structure, social and sexual mixing groups, spatial spread, vaccination, quarantine, and chemotherapy” (Hethcote, 2000).

Even when the interest is mainly focused on the theoretical properties of the model, deterministic models are usually easier to deal with than their stochastic counterparts, since they often lead to differential equations, a very mature branch of mathematics in which theory and numerical methods are well established and readily available. The review (Hethcote, 2000) contains numerous references to analytical results concerning asymptotic properties of compartmental deterministic models, involving both ordinary and partial differential equations, and, in particular, results about local and global stability analysis of model steady states.

Despite the vast literature and the numerous applications to practical problems, deterministic models have some limitations. First of all, they are usually based on some unrealistic assumptions, the most crucial of which is often the oversimplified description of the interaction between individuals (Becker *et al.*, 1995; Edmunds *et al.*, 1997). Second, they are not representative of the system behaviour when numbers are small and they are therefore unable to capture some particular effects which are known to be important in epidemic modelling, e.g. the fact that outbreaks may die out because of random events before they become established (e.g. see Andersson & Britton, 2000) or that the time they require to become established is highly dependent on random events occurring when the number of infectives is limited (Barbour, 1975; Svensson & Scalia-Tomba, 2001; Andersson & Britton, 2000).

Quite descriptive of such limitations is the case of the model describing the spatial spread of rabid foxes in the UK considered in Murray *et al.* (1986) (see also Murray, 2002). After a first epidemic wave and the apparent disease extinction, the model predicts a new epidemic wave, although smaller. The authors comment on the fact that a deeper investigation would require a stochastic model to be used instead of a deterministic one, but they do not go further. Instead, the reasons behind the new epidemic wave have been discussed in light of stochastic considerations in Mollison (1991) and referred to as the *attofox phenomenon*: the deterministic model used was unable to deal properly with the

extinction of the infection and a small fraction of fox (atto = 10^{-18}) was still present in the population and caused a resurgence in the number of infected cases and a new epidemic wave.

Stochastic models

Stochastic models are characterised by the fact that each single event is attributed a probability of occurrence. Events may or may not occur, but are always counted in discrete units, and therefore the number of infected cases is always an integer value.

This provides a more reasonable representation of reality at the individual level and therefore is much more suitable to describe the system behaviour when the number of actors involved is small. In particular, stochastic models are able to capture important phenomena that are expected to occur when dealing with small numbers. Some of them have been cited above: extinction and random delays before an epidemic becomes established. Others may be more complex and involve external forces, e.g. casual reintroductions of the disease in the population, random environmental variations, etc.

The increased realism of stochastic models versus deterministic ones is counterbalanced by an increased level of complexity. Full stochastic models generally involve the study of the evolution in time of the probabilities that the system is in each possible state (see e.g. Keeling, 2008; Dodd & Ferguson, 2007) or, when the attention is not focused on the full dynamics, at least of the full distribution of quantities of interest. In this last case, powerful techniques can involve, for example, the use of moment generation functions of random variables (Newman, 2002) or results concerning convergence of random processes (Andersson & Britton, 2000).

Although when numbers are small stochastic models need to be used, when numbers become large, correlations between random variables become weak and the law of large numbers can be invoked to show that the average stochastic dynamics are well approximated by the deterministic ones, thus allowing the simpler deterministic models to be used instead. Although the idea of a deterministic approximation is somewhat intuitive,

the mathematical theory supporting it is not at all trivial: see Andersson & Britton (2000); Kurtz (1971), and references therein for the Markovian case (i.e. constant recovery rate; see Section 2.1.4) and Kurtz (1981) for extensions to different distributions of infectious periods.

1.2.4 Simple model and complex simulations

The simplest possible description of the contact patterns between individuals is the homogeneous one: each individual has the same chance to contact any other individuals in the population. This *homogeneous mixing* assumption, originated from mass-action descriptions of kinetics reactions, is used in epidemic models exclusively for mathematical convenience. It is in fact widely recognised that the heterogeneity in contact patterns between individuals, due to age differences or different levels of susceptibility and infectivity, are important factors affecting the spread of the infection (e.g. see Bansal *et al.*, 2007 and references therein; see also Andersson & Britton, 2000; Diekmann & Heesterbeek, 2000; Anderson & May, 1991). Furthermore, the spatial distribution of individuals and the presence of a social structure in the population play also a crucial role in the epidemic spread, by adding various forms of heterogeneity in contact patterns and local saturation effects (Riley, 2007; Riley & Ferguson, 2006; Becker *et al.*, 1995; Edmunds *et al.*, 1997).

Besides increased realism, more complex models than those assuming homogeneous mixing are also necessary to investigate the effects of targeted control policies, which are usually less expensive and more efficient than control policies applied to the whole population (e.g. see Roberts, 2004; Aldis & Roberts, 2005; Becker *et al.*, 2005; Longini *et al.*, 2005; Ferguson *et al.*, 2005, 2006; Roberts *et al.*, 2007). In other cases, targeted control policies may not be particularly efficient, but administratively convenient: for example, it may be easier to deliver vaccination to entire households instead of randomly selected individuals (Becker & Dietz, 1995; Becker *et al.*, 1995; Becker & Utev, 1998; Wu *et al.*, 2006; Fraser, 2007). Furthermore, during an emerging outbreak, control policies like quarantine and prophylactic treatment are often targeted at households (Wu *et al.*, 2006; Fraser, 2007),

while social distancing measures can involve, among other actions, school and workplace closure (e.g. Ferguson *et al.*, 2005, 2006; Cauchemez *et al.*, 2008).

Unfortunately, the increased realism gained by adding a more detailed social structure to a model often makes it analytically intractable and one has to rely on complex microsimulations (Longini *et al.*, 2004, 2005; Ferguson *et al.*, 2005, 2006; Germann *et al.*, 2006; Riley & Ferguson, 2006) which can be difficult to parameterise (Elveback *et al.*, 1971, 1976), require a computationally intensive sensitivity analysis and might hinder the understanding of the causal determinants of model behaviour (Ferguson *et al.*, 2003). Therefore, despite the increasing computational power available to researchers today, analytically tractable models remain invaluable tools in understanding the role that different factors play in the spread of the infection. Furthermore, the fact that no or very little computational power is required for their analysis and application makes them very attractive for policy information when quick results are needed, as for instance in the case of emerging infection outbreaks (Roberts, 2004; Aldis & Roberts, 2005; Roberts *et al.*, 2007).

In the exploration of epidemic models that are simple enough to allow for some analytical tractability but at the same time are more realistic than oversimplified models, two main research areas seem to be identifiable: *network models* and *models with multiple levels of mixing*. The distinction between the two is not clearly marked, in particular since any form of interaction between individuals can be represented through a suitable network: homogeneous mixing, for example, can be represented by a complete network, where each individual is connected to any other; homogeneous mixing where no infective can infect more than k susceptibles (e.g. number of secondary cases binomially distributed between 0 and k) can be seen as a random network where all nodes have degree k . It is questionable, however, to what extent it is worth using a network approach in these limiting cases: e.g. the last example has been treated in Diekmann & Heesterbeek, 2000, p. 269, but the authors describe the interaction as a “network with hardly any structure”.

Given that a rough distinction can be made between these two areas, a very brief discussion (not even close to any possible sort of completeness) about the network modelling

approach is proposed here. The remainder of the thesis, instead, is concerned with a much deeper presentation of epidemic models with two levels of mixing.

1.2.5 Network models

Network models represent an extremely flexible modelling tool, in which the population is represented as a graph: vertices represent individuals and edges represent acquaintances between them which offer a possibility for transmission; sometimes, alternatively, edges are used directly to represent infectious contacts, which result in transmission as soon as one individual is infectious and the other is susceptible. Here, edges are thought as acquaintances through which the infection may or may not spread.

In a network model, it is possible to define for each individual a set of individuals (his or her *neighbours*) to which the infection can be transmitted and a set of individuals from which the infection can be acquired, thus avoiding the unrealistic homogeneous mixing assumption (see, for example, Keeling & Eames, 2005). Most of the time, the two sets coincide, i.e. only symmetric contacts are considered (e.g. two-way conversations, handshakes, etc.), and interactions between individuals can therefore be modelled using undirected networks. However, this is not always the case, as the model can include also some asymmetric forms of transmission (e.g. blood transfusions), which require directed networks to be used instead.

Apart from some simple cases, it is usually impractical to describe a real social structure through a network, because it would require knowledge of every individual and every relationship in the population. Even if a complete description were possible, it is in general not easy to understand how the final outcome on an epidemic would change if the network were to change, even slightly (Keeling & Eames, 2005). Additionally, real social networks are dynamic, not static – people form new connections throughout life, and lose others.

Different idealized networks have thus been considered in modelling studies in order to understand how general properties of networks influence epidemic dynamics (e.g. see Keeling & Eames, 2005; Keeling & Rohani, 2008). Various properties of the network are

deemed to be relevant for epidemic spread. The two main ones considered here are: the *average path length*, defined as the expected minimum number of edges required to connect two randomly chosen vertices, and the *clustering*, i.e. the probability that two acquaintances of the same individual are also acquaintances of one another (Keeling & Eames, 2005).

Some of the most relevant examples of idealised networks are random graphs, lattices and small worlds (other are spatial or scale-free networks).

Random graphs

Random graphs are formed by connecting each pair of nodes independently, with a certain probability. The spatial position of the vertices is irrelevant. Random graphs are characterized by short average path length and little clustering (Keeling & Eames, 2005; Keeling & Rohani, 2008). A real human network usually shows a structure in which the first property is satisfied but the second one is not, i.e. real networks are highly clustered (see Watts & Strogatz, 1998; Newman *et al.*, 2000; Moore & Newman, 2000).

Lattices

Lattices are networks in which the vertices are positioned on a regular grid and connected only to the nearest neighbours (Keeling & Eames, 2005). For example, a two-dimensional square lattice is represented by nodes occupying all positions on the Cartesian plane with integer coordinates: each node is then connected with the four closest nodes, i.e. parallel to the main axes. Since lattices show a high clustering, an epidemic simulated on such a network spreads in roughly circular wave-like fronts of infection (Keeling & Eames, 2005) and takes a long time to reach a susceptible individual occupying a position far from the point where the epidemic started.

Small world networks

“*Small world*” is a term describing a network showing high clustering and small average path length (Watts & Strogatz, 1998; Newman *et al.*, 2000; Moore & Newman, 2000; Keeling & Eames, 2005). Therefore, small worlds are far more suitable than the previous two kinds of networks to describe real-world social networks.

A possible construction algorithm which leads to a network with these properties has been suggested by Watts & Strogatz (1998), when they first introduced the concept of small-world network. Starting from a lattice or any other network with natural high clustering, edges are selected and rewired with a certain probability by keeping fixed the node at one side of the edge and selecting a new node for the other side, at random from the entire graph.

A remarkably low number of rewired edges is sufficient to dramatically reduce the average path length of the network, while a large number of rewired edges is necessary to reduce sensibly the clustering of the network. Therefore, intermediate networks, in which some but not too many edges have been rewired, show simultaneously the high clustering and short average path length that characterise real-world social networks.

From an epidemiological point of view, a disease starting from one point initially presents a wave-like front, but very soon the presence of long-range random links allows the infection to be suddenly seeded in other regions of the network, where new wave-like fronts start. This reduces dramatically the time the infection needs to sweep through the population (Keeling & Eames, 2005; Keeling & Rohani, 2008).

Comparison with homogeneous mixing models

A possible comparison between an epidemic in a homogeneous mixing model and an epidemic spreading through a network is obtained when parameters are tuned such that the initial case infects the same average number of secondary cases in the two models. In a large susceptible population, the number of secondary cases when homogeneous mixing is assumed is Poisson distributed, while on the network it is binomially distributed: given

the same mean, however, the binomial distribution has a smaller variance and therefore a smaller probability that no infection occurs. Hence, in its very early stage, an epidemic is less likely to fade out than when homogeneous mixing is assumed (Keeling & Eames, 2005; Keeling & Rohani, 2008). On the other hand, the clustering properties of some networks cause a very rapid saturation of the local environment around an infective, thus reducing the mean number of secondary cases generated by an infectious individual and increasing the time the infection requires to spread out (Keeling, 2005; Keeling & Eames, 2005; Keeling & Rohani, 2008).

If the tuning between the two models is obtained by forcing the epidemic to spread at the same speed, then the transmission parameter in the network model needs to be higher than in the previous comparison, to compensate for the effects of local saturation. Therefore, the probability of epidemic extinction is even less likely than when matching the average number of initial infections.

Another fundamental difference between a homogeneous mixing model and a network model is that the former always exhibit an exponential increase in the number of infected cases, while in the latter, repeated contacts between already infected individuals, which tend to occur when clustering is sufficiently high, may cause the lack of an exponential growth in the number of cases.

Further comments

In general, the presence of exponential phases is strictly linked with the concepts of threshold parameters and reproduction numbers, which are arguably the most important concepts in infectious disease epidemiology. However, real-world epidemics do not always show an exponentially growing phase. Therefore, networks could represent an important modelling tool to investigate many issues of practical interest that are difficult to capture within other modelling frameworks.

Much more could be said about such a vast research area: however, given that the main topics of this thesis are models with two levels of mixing, the choice here has been to

discuss only those concepts that will appear in the remainder of the thesis. For everything else concerning networks, the reader is referred to the very comprehensive reviews given by Albert & Barabási (2002) and Newman (2003).

Chapter 2

Background

In this chapter some well established results from the literature are recalled, both for completeness and because they represent the “bricks” upon which the more complex results of the next sections are based. The reader familiar with the subject can skip this chapter, as the following ones will reference to its relevant points when needed. However, even the familiar reader might benefit from a quick look at Section 2.1.3 and 2.1.4, on the different assumptions concerning the possible disease histories considered in this thesis, and at Section 2.4 about households models, which directly generalises to the households-workplaces model of Chapter 4.

2.1 Models for transmission at the individual level

2.1.1 The concept of disease history

When an individual is infected, the pathogen (microparasite, in this case) typically undergoes a phase of reproduction which triggers a reaction from the immune system. A competition takes place, where usually the immune system requires some time to overcome the pathogen. During this period, the host is infectious and can transmit the infection to

other hosts. If and when the immune system gains final control, the host is not infectious anymore and is here considered to be totally immune to another infection from the same pathogen.

The transmission process therefore depends at least on the pathogen-host competition and on the interaction between an infected host and another one. Further complexities, like the contribution of the surrounding environment to the transmission process, will not be considered here (see Grassly & Fraser, 2008).

It is often extremely difficult to disentangle such components and for this reason, when studying the dynamics at the population level, it is often preferred to take a more pragmatic approach and assume that the infectiousness of an individual and the interaction with others are captured in what is here called a *disease history*.

There is usually little consensus among researchers about which disease history is the most suitable for a particular model, even when the interest is focused on a specific pathogen in a specific population. Furthermore, different assumptions concerning the disease history can sometimes have a substantial impact on a particular outcome of interest, while in other cases they can have little or no impact at all. However, given the fact that such a disease history is a basic ingredient of an epidemic model, researchers usually make an initial assumption about it and then carry on their model analysis. Sensitivity analysis for this assumption is rarely performed, especially since the disease histories considered in the literature can differ not only in the numerical value of some parameters but also in their fundamental structure. For this reason it is often not clear how much the results depend on the particular choice of the disease history. Particular care has been taken in this thesis to unravel as much as possible the implications of this fundamental choice.

An overview of most of the disease histories appearing in the literature is presented below.

The transmissible disease, unless otherwise specified, is considered to be a human infection and therefore the host will be referred to as an individual.

2.1.2 Density-dependent and frequency-dependent transmission

The mathematical notation appearing in the literature is still inconsistent, mainly because different research groups adopt different fundamental assumptions. In particular, two rather distinct approaches are commonly considered. They are based on two different assumptions referred to as *density-dependent* and *frequency-dependent transmission*.

Consider a population of size n and define *infectious contact* as a contact between two individuals that results in an infection whenever one is infectious and the other is susceptible. Assume that individuals mix homogeneously, i.e. each individual has the same chance of contacting any other individual in the population.

When density-dependent transmission is assumed, the average number of infectious contacts an infective makes with any other individual increases linearly with the population size. This framework has also been referred to as *pseudo mass-action* (de Jong *et al.*, 1995). Since under this assumption the probability that each infective infects a given susceptible is kept constant, i.e. independent of the population size, it is convenient to parameterise the model in terms of an infective-to-susceptible infection rate. This is, for example, the case of Ball and co-authors (e.g. Ball, 1986; Ball *et al.*, 1997; Ball & Neal, 2002; Ball & Lyne, 2002; Ball *et al.*, 2004) whenever they consider a single population (as opposed to a sequence of populations with increasing size).

When frequency-dependent transmission is assumed, the average number of infectious contacts an infective makes is independent of the total population size. In this framework, a single infective has a probability to encounter a susceptible that is proportional to the density of the susceptibles in the population. This assumption occurs frequently in the study of reaction kinetics, and is known as the *law of mass-action* (e.g. see Murray, 2002). In this case, it is convenient to parameterise the model in terms of the rate at which an infectious individual makes infectious contacts with *any* other individual in the population (see notation at the end of this section). This is, for example, the framework considered in Diekmann & Heesterbeek (2000).

The two different frameworks represent two extreme cases and it can be argued that

neither of those is particularly representative of reality. Intermediate behaviours, where the dependence on the population size is more complex than linear, have also been explored (McCallum *et al.*, 2001; Hethcote & Driessche, 1991, and references therein), although they may appear to be somewhat artificial.

Although different opinions concerning which model should be used may arise when homogeneous mixing is considered in a small group (e.g. a household), when an infinite population limit is considered by letting the population size increase, researchers usually agree on the fact that the average number of infectious contacts from an infective cannot increase indefinitely and that therefore the frequency-dependent assumption is the one to consider. Ball and co-authors, despite using density-dependent transmission when considering single populations, naturally switch to a frequency-dependent framework when investigating sequences of populations with size tending to infinity.

Although the results for small groups presented here hold for both the above frameworks, the frequency-dependent assumption is the one that is generally considered, except when otherwise stated. The motivation lies in the empirical evidence that published data seem to convey, according to de Jong *et al.* (1995) and Hethcote (2000) for large and moderately large populations, and the suggestion from a more recent household study (Cauchemez *et al.*, 2004) that the mixing patterns resemble more those of the frequency-dependent framework than the density-dependent one even in very small groups like households.

However, since there is still inconsistency in the literature and an overview of various models for the disease transmission from different research groups is recalled below, some effort has been put into maintaining as much uniformity as possible in the notation adopted here (see also p. 19 for a summary table of notation). In general, the rate at which infectious contacts occur from an infective to all other individuals (described hereafter in short as *one-to-all*) will be denoted by β , whether it is constant or time-dependent. The infectious-to-susceptible infectious contact rate (in short *one-to-one*) will be denoted by λ , whether it is constant or time-dependent. Therefore, if n is the size of a population for

which homogeneous mixing is assumed, it follows $\lambda = \beta/(n - 1)$. When the population is large, n will be usually substituted by N and the approximation $\lambda = \beta/N$ will be often considered instead. Note that, when density-dependent contact rate is considered, λ is assumed to be constant, i.e. β is proportional to the population size. Vice versa, when frequency-dependent contact rate is considered, β is assumed to be constant, while λ is inversely proportional to the population size.

For completeness, it is worth enriching the notation by denoting with Φ the one-to-all total infectivity (random, in general), i.e. the area with average R_0 under the whole curve associated with the one-to-all infection rate β , and with Λ the one-to-one total (random) infectivity associated with λ and with average $R_0/(n - 1)$.

2.1.3 Non-random disease history

In an unstructured population, if each individual that is infected experiences the same disease history as any other infective, the disease history is here referred to as *non-random*. Some basic non-random disease histories are now described, together with the transmission model associated with them. All the following transmission models assume homogeneous mixing between individuals and assume that, after recovery, an individual is immune and does not participate further to the spread of the epidemic.

Reed-Frost model

The Reed-Frost model was first proposed by Lowell Reed and Wade Hampton Frost in a series of lectures in 1928 (see Bailey, 1975, pp. 12-13, or Fine, 1977) but remained initially unpublished. Despite being well known nowadays and having been discussed by many authors (Ball, 1983a; Scalia-Tomba, 1985, 1986; Andersson & Britton, 2000, e.g.), its definition is briefly reported here for convenience.

When infected, an individual is assumed to experience a latent period (infected but not infectious) of a fixed length, usually taken as the time-unit in the model, after which all the transmission is concentrated in a single point in time. Each susceptible independently

escapes infection from each infective with fixed probability q . This leads to an epidemic that spreads in discrete generations and for which all infection events are independent of one another.

Despite the unrealistic temporal dynamics, the model may be suitable for infections characterised by a short infectious period compared to the latent period (e.g. measles but not influenza), and in populations that are sufficiently small to prevent different generations from overlapping (e.g. households). However, even when generations overlap, it is shown in Section 3.2 that the model can still be used to compute the final size probability distributions.

Time varying infectivity

More realistic temporal dynamics than those of the Reed-Frost model are obtained by assuming that, after infection, an individual makes infectious contacts with other individuals (one-to-all, see above) at the points of a Poisson process with time-inhomogeneous infection rate described by a function $\beta(\tau)$, identical for each infective, where τ represents the time elapsed since the infection of the individual. $\beta(\tau)$ is assumed to satisfy the conditions $\beta(\tau) \geq 0, \forall \tau > 0, \beta(\tau) = 0, \forall \tau < 0$. It is biologically reasonable that $\int_0^{+\infty} \beta(\tau) d\tau$ be finite, as it represents the total infectivity spread by the individual, universally denoted by R_0 . This implies that $\beta(\tau) \rightarrow 0$ as $\tau \rightarrow \infty$, but does not imply the presence of a particular recovery time. Infectives can then be thought of as never recovering or they can be thought of as recovering at a certain time T_{max} after infection, such that $\beta(\tau) \equiv 0, \forall \tau > T_{max}$.

The *infectivity profile* $\beta(\tau)$ is commonly thought to be unimodal, but can also have more complex shapes: from a biological point of view, in fact, it is closely related to the viral shedding profile which is usually unimodal, but can also range from bimodal, as in the case of HIV (Wawer *et al.*, 2005; Hollingsworth *et al.*, 2008; Grassly & Fraser, 2008), to multimodal, as in the case of malaria (Grassly & Fraser, 2008).

A possible latent period of length e is naturally included by requiring that $\beta(\tau) = 0, \forall \tau \in [0, e]$.

2.1.4 Random disease history

It is recognised that individuals can have different reactions to the pathogen and be more or less infectious when infected. Except from some particular conditions, as in the case of coinfections or various forms of immunodepression, there is little hope of having prior information about how an individual will react to infection. A possible solution is to assume a *random* disease history.

Extended Reed-Frost models

The Reed-Frost model can be extended to the so-called *randomised Reed-Frost* model, where the one-to-one escaping probability Q is randomly drawn from a specified distribution (e.g. see Von Bahr & Martin-Löf, 1980; Picard & Lefèvre, 1990). A further generalisation of the randomised Reed-Frost model is the so-called *collective Reed-Frost* model (Picard & Lefèvre, 1990; Lefevre & Utev, 1999) and a very general approach has also been considered by Scalia-Tomba (1990).

Standard *SIR* model

In the *standard SIR model* (Andersson & Britton, 2000; Ball, 1986), infectives are assumed to make infectious contacts with each other selected individuals (one-to-one) at the points of a homogeneous Poisson process with constant rate λ during infectious periods with lengths given by independent realisations of a random variable (r.v.) I . For biological reasons, I is usually assumed to have at least finite mean and variance.

The Poisson processes are assumed to be independent between individuals and with respect to the length of the infectious periods. In the case where the r.v. I is exponentially distributed, the model reduces to the special case of the so-called *general stochastic epidemic model* (Andersson & Britton, 2000), a name that now seems unsuited, given the numerous generalisation that have been considered in the literature.

When a latent period needs to be included, it can be done by specifying another r.v. E and assuming that, upon infection, individuals experience latent periods the lengths of which are given by independent and identically distributed copies of E . However, the additional presence of a latent period is rarely considered. The reason, carefully explained in Section 3.2, is that most of the results are actually independent of its presence and particular form.

Time varying infectivity

The standard *SIR* model is the most commonly considered when interest is placed on how the variability in the total infectivity affects the epidemic, since the total infectivity spread by an infective towards a given susceptible is given by the r.v. $\Lambda = \lambda I$. The dynamics, however, can be still seen as unrealistic, since one might expect that the infectivity of an individual will change over time: in particular an individual is usually not infectious immediately after having been infected, but undergoes a latent stage, after which the infectivity builds up as the disease progresses.

A very general model that can be considered is therefore the one that assumes that, at time τ after the infection, each infective makes infectious contact with other individuals (one-to-all) at the points of an inhomogeneous Poisson process with rate $\Phi\omega(\tau)$, where $\omega(\tau)$ is a normalised distribution, called *generation time distribution* (see Section 2.2.4 below, or Fraser, 2007; Svensson, 2007), and Φ is a r.v. describing the total (one-to-all) infectivity of an individual.

This model does not include all the previous cases (for example, a random length I of infectious period would require the function $\omega(\tau)$ to be random, as done in Svensson, 2007) but it is sufficiently general to include the benefits of a time-since-infection modelling approach and the variability in infectivity across individuals, given by Λ (one-to-one) or Φ (one-to-all), which is explained in Section 3.7 affecting the final size distribution of an epidemic in a small population.

2.2 Unstructured large populations

Since an infection event requires an *infecter* and an *infectee*, epidemic models are characterised by an intrinsic non-linearity, because new infections occur at a rate which depends in general both on the number of infectives and the number of susceptibles present in the population. For this reason, the mathematical analysis of epidemic models is challenging and assumptions that simplify the mathematical tractability are often considered.

A well established simplifying assumption is to consider a large population and to approximate the initial phase of the epidemic with a linear model where the number of susceptibles remains constant at the initial value (see, for example, Andersson & Britton, 2000; Diekmann & Heesterbeek, 2000). Most of the fundamental results in the field of epidemic modelling (e.g. the concept of R_0) rely on this assumption and some of them are recalled here for reference and as basis for the further extensions.

2.2.1 Basic reproduction number

The basic reproduction number (Heesterbeek, 2002) is probably the most important concept in infectious disease epidemiology. It is defined as the average number of secondary cases generated by a “typical” infectious case in a totally susceptible population and is almost universally denoted by R_0 .

The concept appeared first in demography and ecology (Heesterbeek, 2002), where it is defined as the average number of female offspring of a female individual. The parallelism with epidemic modelling is clear, under the caveat that a population can in theory expand indefinitely, while the number of infected individuals cannot be larger than the total population size. But under the infinite population assumption, the two frameworks coincide and the concept of R_0 for epidemics can be formally defined.

Branching process approximation

It has already been observed in the introduction that a stochastic framework is more complex but more suitable than the deterministic one for the description of the infection

process when the numbers involved are small.

Upon infection, an individual makes infectious contacts with other individuals at different points in time. Some of the infectious contacts (those with susceptible individuals) result in disease transmission. If the interest is not in the times of the infections but only in their number, the infection process can be modelled as a discrete-step *branching process* (see e.g. Grimmett & Welsh, 1986), where each infective is replaced by a random number ξ of new infectives, considered to constitute the next generation. In general, as the epidemic progresses, since more and more contacts occur with already infected individuals and do not result in real infections, the distribution of ξ changes across different generations. However, in the infinite population limit and in the early phase of the epidemic (all susceptible but a finite number of infective or recovered individuals), such a distribution becomes independent of the generation. This type of branching process is called a (*simple*) *Galton-Watson process* (e.g. see Grimmett & Welsh, 1986) and the distribution of ξ is called the *offspring distribution*. Finally, the basic reproduction number is defined as $R_0 = \mathbb{E}[\xi]$.

In a large but finite population, the infinite population approximation is exact up to the moment of appearance of the first *ghost* (Andersson & Britton, 2000), i.e. the first individual that fails to be infected when contacted because of a previous infection event. When the saturation of susceptibles is taken into account and the distribution of ξ changes generation by generation, $\mathbb{E}[\xi]$ is often referred to as the *effective reproduction number* and denoted by R_e .

Threshold condition

According to standard branching process theory (e.g. see Grimmett & Welsh, 1986), the population of a Galton-Watson process either goes extinct in finite time or grows unbounded. In an infinite population the former case corresponds to an epidemic that dies out after a finite number of cases, i.e. zero fraction of the population; the latter to an epidemic that infects a positive fraction of the population. Here the two different out-

comes are referred to as *small* and *large* epidemics, respectively. When the population is finite, such a distinction becomes blurred: the smaller the population, the less neat the distinction between a negligible and a major fraction of the population infected.

Standard branching process theory also teaches how such a dichotomy in the behaviour of the Galton-Watson process is dictated exclusively by the average of the offspring distribution. The consequence of this is the *threshold property* of the basic reproduction number for epidemics: if $R_0 < 1$, no large epidemic can occur; if $R_0 > 1$, large epidemics occur with positive probability, although such a probability is in general smaller than 1 as random events can cause the extinction of the infection before the epidemic becomes established.

Note that, despite the use of the term “small” for epidemics with $R_0 < 1$, such epidemics can still be matter of concern for the health care system, especially when R_0 is close to 1 (for example, see Jansen *et al.*, 2003, about measles outbreaks in the UK).

Critical vaccination coverage

In addition to its threshold property, R_0 is important for the simple relationship it has with the critical vaccination coverage, i.e. with the fraction of the population one has to vaccinate in order to control the spread of the infection. In fact, the vaccination of a fraction

$$p_{crit} = 1 - \frac{1}{R_0} \quad (2.2.1)$$

of the population causes each of the R_0 infectious contacts of an average individual to hit a susceptible with probability $1 - p_{crit} = 1/R_0$, and thus causing the value of R_0 after the vaccination to be reduced to 1.

Other control policies have also the effect of reducing R_0 , although such a reduction is much more difficult to compute and depends on further assumptions and model choices.

2.2.2 Final size

The final size of an epidemic is another quantity of interest in epidemic modelling and its importance lies in the fact that it represents a very natural measure of how severe

an epidemic is. Computation of the final size of an epidemic, provided the necessary information about the spread of the epidemic is given, is particularly important and not at all a trivial problem. An even greater importance of this quantity derives from the fact that, under some mild conditions, this outcome is very robust with respect to variations in the model assumptions and is not affected by the actual dynamics of the spread of the infection (see Chapter 3 and Pellis *et al.*, 2008).

Average final size

Given a stochastic model for the spread of the infection, the fraction of the population that is ultimately infected is a random variable $Z \in [0, 1]$. A possible quantity of interest is its average value, $z = \mathbb{E}[Z]$.

Consider a large homogeneously mixing population and assume for the time being that all individuals have the same infectivity (the following result, however, holds under surprisingly general conditions; see e.g. Grassly & Fraser, 2008). Let the population size increase while keeping fixed the number of initial cases. Then, in the limit of an infinite population, the average final size (as a fraction of the total population) is either 0, when $R_0 < 1$, or, when $R_0 > 1$, it is 0 with probability equal to the probability of extinction of the correspondent branching process and, conditional on non-extinction, is given by the largest root of the transcendental equation

$$1 - z = e^{-R_0 z} \quad (2.2.2)$$

(Andersson & Britton, 2000). Note how the only parameter on which the final size depends is the basic reproduction number. When the epidemic starts with a positive fraction χ of the population initially infected (an infinite number of initial cases, in the infinite population limit), then the final fraction of the initially susceptible population that is ultimately infected, is given by the only positive solution z of the equation

$$1 - z = e^{-R_0(\chi+z)}. \quad (2.2.3)$$

If an equivalent deterministic model is used instead of a stochastic one, the same equation

gives also the final size in the deterministic case.

Equations (2.2.2) and (2.2.3) can be read as “balance equations”, where the left-hand side represents the average fraction of the population that ultimately escapes infection, while the right-hand side represents the probability that an initial susceptible escapes the total average infection “pressure” (i.e. the total force of infection integrated over the time during which it is exerted) due to all individuals infected in the entire epidemic: in fact, denoting by N the total (large) population size, each of the Nz cases infected during the epidemic exerts a total infection pressure of R_0/N on each single susceptible and, given that the infection process is modelled as a Poisson process, e^{-R_0z} represents the probability that that susceptible ultimately escapes infection. Note that the times at which different cases are infected is irrelevant, under suitable assumptions discussed thoroughly in Chapter 3, and their infectivity is added together because the sum of independent Poisson processes is a Poisson process with the rate given by the sum of all the rates.

Final size distribution

If the interest is focused not only on the average final size but also on its full distribution, then it has been shown (e.g. see Andersson & Britton, 2000) that, in a very large population of size N , the final size distribution conditional on non-extinction is approximately Gaussian around the final size z of the deterministic model (Equation (2.2.3)), with variance given by:

$$\frac{1}{N} \frac{\varepsilon(1 - \varepsilon) + \beta^2 \sigma^2 (\chi + z) \varepsilon^2}{(1 - \beta \iota \varepsilon)^2}, \quad (2.2.4)$$

where β represents the one-to-all infection rate, ι and σ^2 denote respectively the mean and the variance of the r.v. I describing the length of the infectious period, and

$$\varepsilon = 1 - z = e^{-R_0(\chi+z)}$$

represents the average fraction of the initial susceptible population which escapes infection. The initial fraction of the population infected χ could be 0, when the initial infectives represent only a negligible fraction of the total population size.

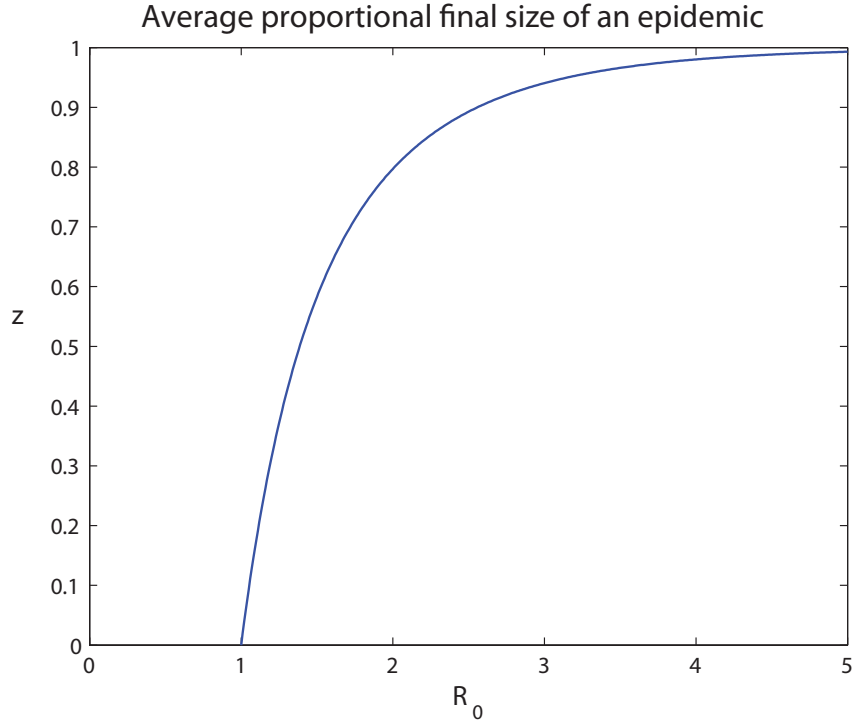


Figure 2.1: Average proportion of the population ultimately infected, in the limit of an infinite population and conditional on a large epidemic occurring, plotted as a function of R_0 . When $R_0 \leq 1$, no large epidemics are possible.

The reason why in an infinite population, the deterministic result coincides with the average of the stochastic one, despite the epidemic being a non-linear process, is not at all trivial and relies on the fact that at each point in time a large stochastic epidemic, once established, is normally distributed around the values predicted by the deterministic model (see Andersson & Britton, 2000, or Kurtz, 1971, 1981), intuitively because the infinite population limit allows correlations between infectives and susceptibles to be neglected.

Basic reproduction number and final size

As already mentioned, an important property of the average final size in a simple homogeneously mixing population is that it depends only on the value of R_0 and not on other details of the model.

Figure 2.1 shows that R_0 and the average final size z are monotonically related and

therefore, since it is very natural to use the average final size as a measure of how severe an epidemic is, R_0 turns out to be an equivalent indicator of the epidemic severity.

2.2.3 Real-time growth rate

The basic reproduction number and the final size of an epidemic are extremely useful in providing information about the probability of a large outbreak occurring, of the effort required to control or reduce the spread of the infection and as a measure of the severity of the epidemic. Their main strength is arguably their relative independence of the details of the individual-to-individual transmission and especially on the dynamics of the spread itself. In other words, they contain no information about time or about the speed of the spread.

However, this also represents a severe limitation, because the dynamics of the spread are not only theoretically interesting but also invaluable in practice, as they provide information concerning the length of the epidemic, the time allowed to control it at the source, the effectiveness of responsive control interventions, and so on.

Lotka-Euler equation

The simplest possible piece of information concerning the dynamics of an epidemic is represented by the real-time growth rate.

The main result for the computation of the real-time growth rate r for a homogeneous mixing model is here presented and discussed (see also Diekmann & Heesterbeek, 2000). Consider, for the time being, that the disease history is characterised by a time-varying infectivity profile $\beta(\tau)$, equal for every infective, where τ represents the time since infection. For this model,

$$R_0 = \int_0^{+\infty} \beta(\tau) d\tau \quad (2.2.5)$$

(assumed to be finite because of its biological meaning). Assume that $R_0 > 1$ and consider a deterministic epidemic during the exponentially growing phase. The incidence $y(t)$ is therefore increasing exponentially with an unknown real-time growth rate r , i.e. $y(t) \propto e^{rt}$.

Now, if at time t the number of susceptibles is given by $X(t)$, the incidence at time t is given by the contribution, for all $\tau > 0$, of the force of infection $\beta(\tau)y(t - \tau)/N$ due to all individuals that have been infected τ units of time before. Therefore,

$$y(t) = \frac{X(t)}{N} \int_0^{+\infty} \beta(\tau)y(t - \tau)d\tau. \quad (2.2.6)$$

In a large but realistic population, the exponentially growing phase is considered to be that time window during which:

1. the number of cases infected is sufficiently large that the deterministic behaviour well approximates the stochastic one;
2. the time elapsed from the beginning of the epidemic is sufficiently long that the initial transient phase, when random events matter, is “forgotten”. In other words, the deterministic approximation holds at time t but also at all times $t - \tau$ that influence the behaviour at t , however large is the value of the delay τ : this is the reason why the upper limit of the integral is $+\infty$ (in practice, any time T_{max} such that the infectivity profile can be considered zero after T_{max} suffices);
3. the fraction of infected population, however, is still negligible compared to the whole population, so that saturation effects can be ignored and the problem linearised.

Since saturation effects are not experienced, $X(t) \approx N$ and, by substituting $y(t - \tau) = ke^{r(t-\tau)}$, where k is a constant. The result is:

$$1 = \int_0^{+\infty} \beta(\tau)e^{-r\tau}d\tau. \quad (2.2.7)$$

This is usually referred to as the *Lotka-Euler equation* (Wallinga & Lipsitch, 2007). Note that the right-hand side of this equation is nothing other than the Laplace transform of β as a function of r and therefore the properties it satisfies are well known. The Laplace transform will be here denoted as $L_\beta(r)$.

Numerical solution

Except for some particular cases, e.g. when the infectivity decreases exponentially ($\beta(\tau) = \beta e^{-\nu\tau}$), for which Equation (2.2.7) can be solved analytically, in general there is no explicit solution. According to Diekmann & Heesterbeek (2000), the fact that this expression for r is more implicit than the usual expressions for R_0 (i.e. Equation (2.2.5)) is one of the main reasons for the large success of the basic reproduction number, despite the fact that the threshold conditions $R_0 > 1$ and $R_0 < 1$ are always equivalent to the threshold conditions $r > 0$ and $r < 0$, respectively (see Diekmann & Heesterbeek, 2000).

However, if the interest is specifically focused on r , the Lotka-Euler equation needs to be solved numerically. This does not represent a problem in general as, when $R_0 > 1$ (which is the most interesting case, since there is no exponentially growing phase when only small epidemics can occur), $L_\beta(r)$ is greater than 1 for $r = 0$ and tends to 0 as $r \rightarrow +\infty$. Furthermore, it can be shown (see Diekmann & Heesterbeek, 2000, and references therein) that $L_\beta(r)$ is a convex function of r , for $r > 0$, and therefore that there is a unique solution of the Lotka-Euler equation that can be found iteratively with the usual numerical methods for finding the zeros of non-linear equations (e.g. bisection, Newton method, secant method, etc.; see any basic textbook in numerical analysis or Press, 2007).

The other non-random disease history commonly appearing in the literature (see Section 2.1.3) is that of the Reed-Frost model, although such a model is often avoided as too unrealistic when the main interest is on the dynamics of the spread. However, the same result described before holds for the Reed-Frost model, with the caveat that $\beta(\tau)$ is not a proper function but a δ -distribution.

The result can be extended to random disease histories as well, but such an extension is postponed to Chapter 5 as part of a wider discussion on the real-time growth rate of an epidemic. Furthermore, it is convenient for such an extension to exploit results concerning the multidimensional case (Section 2.3.2).

2.2.4 Basic reproduction number and real-time growth rate

In the case of the model with time-varying infectivity $\beta(\tau)$, the relationship between R_0 and r is given by Equations (2.2.5) and (2.2.7), and is completely determined by the infectivity profile $\beta(\tau)$. It is often convenient to normalise the infectivity profile, thus defining the *generation time distribution*

$$\omega(\tau) = \frac{1}{R_0} \beta(\tau). \quad (2.2.8)$$

The relationship between R_0 and r is then summarised in the following equation:

$$\frac{1}{R_0} = \int_0^{+\infty} \omega(\tau) e^{-r\tau} d\tau. \quad (2.2.9)$$

The generation time distribution represents the distribution of time intervals between the infection of a case and the infectious contacts he or she makes with other individuals. At the beginning of an epidemic in a large population, i.e. when saturation effects can be ignored and the model linearised, all infectious contacts result in real infections, and $\omega(\tau)$ therefore represents also the distribution of times between infection of an infector and infection of each of his or her infectees (if any). The average of this distribution is often referred to as the *generation time*,

$$T_g = \int_0^{+\infty} \tau \omega(\tau) d\tau, \quad (2.2.10)$$

and represents a crude measure of how quickly new cases are generated and therefore how fast is the spread of the infection.

Given that one of the few pieces of information that is usually readily available during the spread of an epidemic is the number of reported cases over time, the exponential growth can usually be estimated in real-time. However, information about $\omega(\tau)$ is needed in order to estimate R_0 using Equation (2.2.9) and the reconstruction of $\omega(\tau)$ from data is not at all a trivial process (e.g. see Svensson, 2007).

Wallinga & Lipsitch (2007) discussed thoroughly the influence of $\omega(\tau)$ in shaping the relationship between r and R_0 . The main idea is the following.

Since precise information about the shape of $\omega(\tau)$ is not always available, the family of all infectivity profiles with the same average T_g associates to the same real-time growth rate r a whole range of values of R_0 . Although, for fixed r and T_g , distributions $\omega(\tau)$ can always be found for which R_0 is as close to 1 as desired (e.g. by allowing $\beta(\tau) \rightarrow +\infty$ as $\tau \rightarrow 0^+$), a reasonable lower bound for R_0 is given by an exponentially decaying profile $\omega(\tau) = \nu e^{-\nu\tau}$, for which

$$R_0 = 1 + rT_g.$$

As the generation time distribution, from widely spread in time, becomes more and more peaked around the average T_g , the value of R_0 increases and converges to the value

$$R_0 = e^{rT_g},$$

obtained when $\omega(\tau)$ is a Dirac δ -function peaked at T_g . The proof that this distribution provides the maximum value of R_0 is a simple application of Jensen's inequality (see any standard probability book, e.g. Grimmett & Stirzaker, 2001):

Theorem 2.2.1 (Jensen's inequality, probabilistic version). *If $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, X is a random variable on Ω and if ϕ is a measurable convex function on the real axis, then:*

$$\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]. \quad (2.2.11)$$

When $\phi(\sigma) = e^{-r\sigma}$ (convex) and $\omega(\tau)$ is the density function of X , the inequality becomes:

$$e^{-rT_g} \leq \int_0^{+\infty} \omega(\tau) e^{-r\tau} d\tau,$$

i.e.

$$R_0(\omega) \leq e^{rT_g},$$

which holds for any density distribution $\omega(\tau)$ with average T_g .

2.3 Multitype models

The assumption that every individual has the same chance to infect any other individual during the infectious period is clearly unrealistic, since various forms of heterogeneity can be observed in real life: individuals of different ages or with particular characteristics (e.g. immunodepression, risky behaviours) can have different levels of susceptibility to the infection; different individuals can spread the infection more or less than others (e.g. super-spreaders, sexually active individuals), or experience the disease in milder or more severe forms; finally, individuals with different characteristics or ages can mix in different ways with individuals of any other type (e.g. children mix more intensively among themselves than adults do).

In order to include these forms of heterogeneity in epidemic models, one possible strategy is to use *multitype models*. Multitype models assume that individuals can be categorised according to those characteristics that are deemed to be relevant for the infection process and that a description of how individuals of each group infect individuals of each other group can be provided.

Since the mathematical analysis is severely limited by the non-linearity induced by saturation effects, the usual requirement in these multitype models is that the number of individuals of each type is large, so that depletion of susceptibles is not perceived in any of the groups (Diekmann & Heesterbeek, 2000).

Analogously to the single-type models, when the epidemic starts with only a few cases, initial random events drive the whole epidemic process, which can go extinct or result in a major outbreak. A deterministic approximation can still be considered but, as for the single-type model, it is appropriate only when numbers in each group are sufficiently large to invoke the law of large numbers. Deterministic results, therefore, usually refer to the time window already considered in Section 2.2.3. Although not often commented in the literature, however, note that the more groups are considered, the narrower such a time window could be, thus requiring unrealistically large populations in order for the model to be somehow descriptive of the real epidemic spread.

2.3.1 Transmission

For these multitype models, the mathematical submodels describing the individual to individual transmission are trivial extensions of the ones discussed in Section 2.1.3 and 2.1.4.

Among the non-random disease histories, the literature presents examples of:

1. the *multitype Reed-Frost model*, obtained when the escaping probability q of the Reed-Frost model is substituted by q_{ij} , the probability that an individual of type j escapes infection from an individual of type i . A model of this type has been discussed, for example, in Scalia-Tomba (1986);
2. the multitype model with time varying infectivity, obtained as a extension from the single-type model with varying infectivity by defining $\beta_{ij}(\tau)$ as the rate of infection from an infective of type j to a susceptible of type i at time τ after the infection of the j -case. Models of this type are extensively used in Diekmann & Heesterbeek (2000).

Among the random disease histories:

1. the multitype randomised Reed-Frost model, considered for example in Neal (2006) (other very general forms of Reed-Frost model have been considered in Scalia-Tomba, 1990);
2. the multitype *SIR* model, considered for example in Andersson & Britton (2000), and references therein.
3. Finally, multitype randomised models with time varying infectivity seem to be less common in the literature, although they can be included in the very general approach used by Svensson (2007).

2.3.2 Threshold parameters

Basic reproduction number

The basic reproduction number represents the average number of cases generated by a single case in a totally susceptible population. When different types are present, the natural framework extension is obtained by defining the *next generation matrix* (Diekmann *et al.*, 1990; Diekmann & Heesterbeek, 2000)

$$K = (k_{ij}),$$

where k_{ij} represents the average number of cases of type i generated by a single case of type j in a totally susceptible population. The basic reproduction number is then defined as the *dominant eigenvalue* of the next generation matrix (Diekmann & Heesterbeek, 2000; Diekmann *et al.*, 1990).

The reason for such a definition is carefully explained in Diekmann & Heesterbeek (2000): the idea is that, as the epidemic advances, the proportions of cases of each type in each generation stabilises on the components of the eigenvector associated with the dominant eigenvalue and the initial configuration is quickly forgotten.

More formally, given the vector norm $\|v\| = \sum_{i=1}^n |v_i|$, where the v_i 's are the n components of the generic vector v , the associated matrix norm of a generic matrix M is given by

$$\|M\| = \sup_{\|v\| \neq 0} \frac{\|Mv\|}{\|v\|} = \sup_{\|v\|=1} \|Mv\|.$$

The *spectral radius* of a square matrix M is then defined by

$$\rho(M) = \sup\{|\vartheta| : \vartheta \text{ is an eigenvalue of } M\}.$$

An equivalent characterisation is given by

$$\rho(M) = \lim_{n \rightarrow \infty} \|M^n\|^{\frac{1}{n}},$$

which appears to provide a suitable definition of R_0 when applied to the next generation matrix K , as it represents the “long-term average per generation multiplication number”

(Diekmann & Heesterbeek, 2000). Therefore, in general, the definition of the basic reproduction number is given by

$$R_0 = \rho(K).$$

The *Perron-Frobenius theory* of positive matrices provides the mathematical conditions for the dominant eigenvalue to be positive and unique, and to coincide with the spectral radius. In particular it is based on the following definitions.

Definition 2.3.1. A matrix $M = (m_{ij})$ such that $m_{ij} \geq 0, \forall i, j$, is said to be *non-negative* and such a property is denoted by $M \geq 0$.

Definition 2.3.2. A non-negative square matrix $M = (m_{ij})$ is said to be *irreducible* if $\forall i, j, \exists n = n(i, j) \in \mathbb{N}$ such that $(M^n)_{ij} > 0$, i.e. the element in position (i, j) of the matrix M^n is strictly positive. Note that the integer n may be different for different pairs (i, j) .

Definition 2.3.3. A non-negative square matrix $M = (m_{ij})$ is said to be *primitive* if $\exists n \in \mathbb{N}$ such that $(M^n)_{ij} > 0$, i.e. the element in position (i, j) of the matrix M^n is strictly positive. Note that the integer n is independent of the choice of the pair (i, j) . A primitive matrix is also irreducible.

Theorem 2.3.4. Let M be a non-negative square matrix. Then the spectral radius $\bar{\vartheta} = \rho(M)$ is an eigenvalue of M and is dominant in the sense that $|\vartheta| \leq \bar{\vartheta}$ for all other eigenvalues ϑ of M .

In addition to being non-negative, the requirement that the matrix M is primitive guarantees that there are no other eigenvalues ϑ of M such that $|\vartheta| = \bar{\vartheta}$, and therefore that $\bar{\vartheta}$ can be called *the* dominant eigenvalue of M .

From a biological point of view, requiring that the next generation matrix K be irreducible is equivalent to requiring that, if the epidemic starts with an individual of type j , after a finite number n of generations there will be cases of a given type i in the population. In other words, the different types “mix well” and cases of each type will eventually be generated by cases of any type. As opposed, a *reducible matrix* is one where the set of

possible types can be split into subsets that do not interfere with each other and therefore the system can be reduced to more decoupled systems: in this case it may well happen that an epidemic spreads widely among some types of individuals but without affecting individuals of some other types.

Requiring that K is primitive is equivalent to requiring that the different types not only mix well, but are such that, whatever the type (or types) appearing at the beginning of the epidemic, at a certain generation n , all types will be present altogether.

The property of “well-mixing” given by irreducibility and the fact that there is only one strictly dominant eigenvalue R_0 (due to K being primitive) guarantee that, when $R_0 > 1$, generation after generation, the distribution of cases converges to the only eigenvector associated to R_0 . (In a formal mathematical sense, there are some initial conditions for which this statement is not true; however, such particular conditions are unstable and will not occur in practice, because of unavoidable stochastic fluctuations.) It can be shown that this eigenvector has all components strictly positive: such components represent the fractions of individuals of each type appearing in each generation, after sufficient time has passed that such fractions have stabilised and the initial conditions have been “forgotten”.

As in the case of a single type, also for the multi-type models R_0 discriminates between large and small epidemics and is directly related, with the same equation as in the single-type case (Equation (2.2.1)), to the critical vaccination coverage for individuals, i.e. the fraction of the population that needs to be vaccinated in order to prevent or halt the spread of the infection. Such vaccination is assumed to be delivered randomly in the population, i.e. independently of the individuals’ type.

Other reproduction numbers

Although the basic reproduction number R_0 remains the most relevant concept in modern epidemiology, other threshold parameters can in general be defined and can sometimes be more useful. This is the case, for example, of the *type reproduction number* T defined in Roberts & Heesterbeek (2003) and Heesterbeek & Roberts (2007), which is much more

informative than R_0 on the effort required to control the spread when control policies are targeted at a single type instead of all types indiscriminately. Given a type j , the type reproduction number for type j , T_j , is defined as the average number of cases of type j caused by a single case of type j when all possible routes of transmission are considered: the new cases can be generated directly by infection from the single case or through any chain of any length involving intermediate infectives of any other type.

2.3.3 Final size

Equation (2.2.2) for the average final size conditional on a large outbreak having occurred can be extended to the multitype model with l different types. Let χ_i represent the fraction of the subpopulation of type i initially infected (possibly 0 when, in the infinite population limit, only a finite number of individuals of type i is infected at the beginning of the epidemic) and let $z_i \in [0, 1]$ represent the average proportion of initial susceptibles of type i (i.e. not counting the fraction χ_i) that are ultimately infected in the epidemic. Then the z_i 's satisfy the set of l coupled transcendental equations (Andersson & Britton, 2000)

$$1 - z_i = e^{-\sum_{j=1}^l (h_j/h_i)k_{ij}(\chi_j + z_j)}, \quad i = 1, \dots, l, \quad (2.3.1)$$

where h_i represents the fraction of all the initial susceptibles consisting of individuals of type i . The overall average final size z is then given by $z = \sum_{i=1}^l h_i z_i$.

As for the single type case, the result can be read as a set of “balance equations”: for each i , the left-hand side of the correspondent equation represents the fraction of subpopulation i that escapes infection and the right-hand side represents the probability that a single individual of type i survives the infection pressure due to all individuals of all other types that are ultimately infected during the epidemic.

2.3.4 Real-time growth rate

The Lotka-Euler equation for the real-time growth rate for the transmission model characterised by a non-random time-varying infectivity (Section 2.1.3), can be extended to the

multitype case thanks to the argument provided by Diekmann & Heesterbeek (2000). The multidimensional equivalent of the operator “Laplace transform of $\beta(\tau)$ ” as a function of r , $L_\beta(r)$ (Section 2.2.3), is given by the matrix operator

$$K_r = \left(\int_0^{+\infty} \beta_{ij}(\tau) e^{-r\tau} d\tau \right) = (L_{\beta_{ij}}(r)), \quad (2.3.2)$$

In the single type model, the real-time growth rate r for the epidemic is the implicit solution of the equation $L_\beta(r) = 1$. In the multitype case, r is obtained by requiring that the dominant eigenvalue of K_r , $\rho(K_r)$, satisfy the equation

$$\rho(K_r) = 1. \quad (2.3.3)$$

The same technical condition assumed for the next generation matrix K , namely K primitive, are assumed to hold for K_r too, implying the dominant eigenvalue to be unique and coincide with the spectral radius.

It is shown in Diekmann & Heesterbeek (2000) that, in a deterministic framework,

$$R_0 > 1 \iff r > 0. \quad (2.3.4)$$

In a stochastic framework, even when $R_0 > 1$, extinction might occur. However, when the interest is focused on the real-time growth rate, it is implicitly assumed that an exponentially growing phase is being considered and therefore that $R_0 > 1$.

As in the single-type case (Section 2.2.3), Equation (2.3.3) is guaranteed to have a solution, and that this solution is unique, because $\rho(K_r)$ is a monotonic decreasing function of r and that $\rho(K_0) = R_0 > 1$ and $\lim_{r \rightarrow \infty} \rho(K_r) = 0$ (Diekmann & Heesterbeek, 2000). Therefore, efficient iterative techniques can be used to solve it numerically.

2.3.5 Separable mixing

A particularly simple case of multitype model is obtain when the elements k_{ij} of the next generation matrix can be factorised in the product of two terms $k_{ij} = \alpha_i b_j$, each of which depends only on one of the two types involved in the infection process and not on the

other. The model obtained in this case is usually referred to as a multitype model with *separable mixing* (Diekmann *et al.*, 1990; Diekmann & Heesterbeek, 2000) (or, sometimes, *proportionate mixing*, as in Andersson & Britton, 2000, although the same term is also used in Diekmann *et al.*, 1990; Diekmann & Heesterbeek, 2000, for something slightly more specific).

The multitype model turns out to be particularly simplified when separable mixing is assumed, because the factorisation $k_{ij} = \alpha_i b_j$ implies that the matrix K has rank one and the only positive eigenvalue (all the others are zero) is $R_0 = \sum_i \alpha_i b_i$. The result is in essence a weighted form of homogeneous mixing, where in each generation, the fraction of new cases generated by an individual in class j is proportional to b_j and the fraction of new cases of type i infected is proportional to α_i .

At this point, it is convenient to distinguish between two different sources of heterogeneity, due to the fact that the spread of a directly transmissible disease is based on two different processes: a within-host process, involving the interaction between the individual and the pathogen, and a between-hosts process, involving contacts between an infectious individual and a susceptible one.

Heterogeneities can affect either of the two processes, or both.

When heterogeneities involve only the interaction between host and pathogen, individuals may differ from one another in their susceptibility and/or infectivity, but the mixing pattern remains homogeneous, i.e. each individual has still the same probability to contact any other individual. The model so obtained is always characterised by separable mixing, as k_{ij} is now the product of the infectivity b_j of type j and the susceptibility α_i of type i . To be more precise, the availability of individuals of type i also plays a role, so that $\alpha_i = h_i a_i$, where h_i represents the fraction of type i individuals in the population and a_i measures the actual susceptibility of type i . In each generation, the pool of newly infected individuals is partitioned into types proportionally to the availability-susceptibility vector $(\alpha_1, \alpha_2, \dots)$, which is in fact the eigenvector relative to the dominant eigenvalue R_0 .

When heterogeneities in mixing patterns (host-host interaction) are allowed in addi-

tion to those within host, then the multitype model may or may not be characterised by separable mixing (Diekmann *et al.*, 1990; Diekmann & Heesterbeek, 2000).

An important example of separable mixing with heterogeneous contact patterns, in addition to the many examples appearing in Diekmann & Heesterbeek (2000), is given in Anderson & May (1991) concerning the spread of sexually transmitted disease in a population with different levels of sexual activity. Given that an individual that is s times more active than another is also s times more susceptible and s times more infectious, if there are l different groups with sexual activity distributed according to the vector $c = (c_1, \dots, c_l)$ and the proportions of individuals of each type is given by $h = (h_1, h_2, \dots, h_l)$, then the probability that a random infection hits an individual in group j is given by

$$\frac{c_j h_j}{\sum_i c_i h_i} = \frac{c_j h_j}{\mathbb{E}[c]},$$

where the average $\mathbb{E}[c]$ is weighted according to the availability of different types. The average number of infections generated by an infective in group j is proportional to c_j and therefore

$$R_0 \propto \frac{\sum_j c_j^2 h_j}{\sum_j c_j h_j} = \frac{\mathbb{E}[c^2]}{\mathbb{E}[c]} = \mathbb{E}[c] + \frac{\text{Var}(c)}{\mathbb{E}[c]}. \quad (2.3.5)$$

As it is often the case in multitype models, a positive correlation between susceptibility and infectivity within the same type increases R_0 as the most susceptible individuals, which in this case are also more infectious than others, tend to be infected earlier in the epidemic (e.g. see Bansal *et al.*, 2007). In the example of sexually transmitted diseases, this phenomenon is captured in the term of Equation (2.3.5) containing the variance of the sexual activity.

As for R_0 , the computation of the real-time growth rate r is much simpler in the case of separable mixing. In fact, the condition that $\rho(K_r) = 1$ reduces to

$$1 = \sum_i \alpha_i \left(\int_0^{+\infty} \beta_i(\tau) e^{-r\tau} d\tau \right) = \int_0^{+\infty} \left(\sum_i \alpha_i \beta_i(\tau) \right) e^{-r\tau} d\tau = \int_0^{+\infty} \beta(\tau) e^{-r\tau} d\tau,$$

where $\beta(\tau)$ represents the infectious profile of an average individual, where the average is weighted according to the availability-susceptibility of different cases. This is indeed

the right average to consider as during the exponentially growing phase, the (constant) proportion of infectives of each type that occur in an infinitesimal interval of time is given by their relative availability-susceptibility (these proportions represent the elements of the eigenvector of K_r relative to the dominant eigenvalue 1).

This consideration intuitively allows the extension of the Lotka-Euler equation (Equation (2.2.7)) to random disease histories (although technical conditions are in general required as the number of “types”, i.e. possible realisations of the disease history, may not be finite). In this case, all individuals have the same susceptibility and each possible infectivity profile is decided at the time of infection, according to a specified probability distribution. The real-time growth rate can then be obtained by applying the Lotka-Euler equation to the average infectious profile.

2.3.6 Correlations between infector and infectee

Note that, in the classical multitype model, either with or without separable mixing, correlations within susceptibility and infectivity within the same individual are allowed, but it is in general assumed that there is no correlation between the infectivity of the infector and the infectivity of the infectee. In other words, an individual of type j infects individuals of type i according to k_{ij} , independently of the identity of his or her infector. Stated yet differently, each individual can be attributed a type before the epidemic starts, as it is the case, for example (but not only), when the categorisation in types depends on some static characteristics of the individual.

Although this is mathematically convenient, it is not at all easy to find real examples where this assumption is not violated, as often pathogen evolution causes the identity of the infector to affect the future infection outcome of the infectee. In particular, see Section 3.6 for a thorough discussion about these assumptions and their implications.

When this assumption is relaxed, it is still possible to compute R_0 and therefore obtain results about the probability of a large outbreak and the effort required to prevent it. However, the analysis of the epidemic final outcome becomes markedly more complex

(Kendall & Saunders, 1983; Svensson & Scalia-Tomba, 2001; Ball & Britton, 2007). For example, in general it is not possible to determine the average final size of each type, although the overall final size can still be computed. See Kendall & Saunders (1983), Svensson & Scalia-Tomba (2001) and Ball & Britton (2007) for more results.

2.4 Households models

2.4.1 Introduction

Multitype models allow the introduction of important forms of heterogeneity in the model. However, when interest is focused on directly transmitted infections, it is natural to assume that the household represents the most fundamental unit of transmission, followed by the workplace/school structure and then by local communities, towns, cities, and so on (Longini & Koopman, 1982; Edmunds *et al.*, 1997). Such a priori assumption has been supported by complex microsimulations (Elveback *et al.*, 1971, 1976) and has been recently confirmed more quantitatively from the behaviour of similar but larger and more robustly parameterised studies (Longini *et al.*, 2004, 2005; Ferguson *et al.*, 2005, 2006; Germann *et al.*, 2006; Riley & Ferguson, 2006). Moreover, the household is often the fundamental unit at which control policies are targeted, both for practical reasons (people that want or are asked to limit their movements are more likely to stay at home than anywhere else) and structural reasons (decisions about action or compliance to control policies are usually taken by a family as a unit, and most households, though not all, are inhabited by families).

Although it can be argued that a significant proportion of the population is institutionalised and therefore does not live in what is commonly considered to be a household (e.g. long-term care, hospitals, army, student halls, etc.; see, for example, the comments in House & Keeling, 2008), in an attempt to model a more realistic social structure while still allowing for some analytical tractability, a possible alternative to the use of multitype models is achieved by assuming that the population is partitioned into households. The

simplest mixing pattern one can assume is therefore that of homogeneous mixing within each household, superimposed on the homogeneous mixing (usually at a much smaller rate) occurring in the population at large (Ball *et al.*, 1997).

These so-called *households models* have been studied extensively during the last 15 years (Becker & Dietz, 1995; Ball *et al.*, 1997, 2004), although the idea goes back to Bartoszyński (1972).

2.4.2 Comparison with multitype models

The main difference between household models and multitype models is that in the case of households models, even in a large population and at the beginning of an epidemic, the infection process occurring in a household cannot be linearised since the size of the household is small and the local depletion of susceptibles cannot be ignored.

This does not mean that multitype models are unsuitable to describe a social structure: the papers by Roberts (2004), Aldis & Roberts (2005) and Roberts *et al.* (2007) are clear examples of successful application of these models to a socially structured population. They however neglect saturation effects in general, although an “artificial” form of saturation can be introduced, as done in Roberts (2004), Aldis & Roberts (2005) and Roberts *et al.* (2007), by allowing a primary case in a household to infect some secondary cases, but not allowing the within-household epidemic to proceed any further.

On the other hand, the flexibility that households models show in dealing with saturation effects makes them mathematically more challenging, but arguably more realistic. In any case, the comparison between the two different modelling approaches is still an open research question, given their fundamentally different mathematical descriptions, and the problem of which model type to use, and when, still needs to be dealt with satisfactorily.

2.4.3 Brief summary of the most important results

In this section, a particular households model is described and analysed in detail. The main reason for being so specific in the description of a model that has already been studied

carefully in the literature (e.g. Ball *et al.*, 1997) is that the extension to the model with households and workplaces of Chapter 4 is more straightforward.

However, slightly different models have also been considered (Longini & Koopman, 1982; Longini *et al.*, 1982; Becker & Dietz, 1995; Becker *et al.*, 1995; Becker & Dietz, 1996; Wu *et al.*, 2006; Fraser, 2007; House & Keeling, 2008) and it is therefore worth describing some of the common features they share.

Among many important results of all these households models, the one that is most relevant for this study is the definition of a reproduction number, denoted by R_* , representing the average number of households infected by a “typical” infected household in a totally susceptible population (Ball *et al.*, 1997; Wu *et al.*, 2006; Fraser, 2007). This reproduction number for households satisfies the usual threshold property: when $R_* > 1$, large epidemics can occur; when $R_* < 1$ no large epidemic can occur. Its analytical expression is usually of the form:

$$R_* = R_g(1 + \mu), \quad (2.4.1)$$

where R_g represents the average number of infections generated by an infective outside their household and $1 + \mu$ represents the average size of an epidemic in a household, given by the sum of the primary case that imported the infection from outside (in a large population and at the beginning of the epidemic, multiple reintroductions are rare) and the average number of other cases μ . This average takes into account the fact that larger households are more likely to be infected than smaller ones.

As for R_0 in the homogeneous mixing case (Equation (2.2.1)), R_* is linked to a critical vaccination coverage by the formula:

$$p_{crit}^* = 1 - \frac{1}{R_*}. \quad (2.4.2)$$

In this case, however, p_{crit}^* refers to households: interpreting vaccination of a household as vaccination of all its members, p_{crit}^* represents the fraction of households one needs to vaccinate in order to avoid a large epidemic (Becker & Dietz, 1995; Ball *et al.*, 1997; Fraser, 2007). Equation (2.4.2) holds also when households have variable sizes, as long as R_* is

computed appropriately (larger households are more likely to be infected) and the households to vaccinate are selected at random. Since vaccination strategies or other control policies like quarantine and prophylactic treatment are more likely to be targeted to entire households than to randomly selected individuals (Becker & Dietz, 1995; Becker & Utev, 1998; Wu *et al.*, 2006; Fraser, 2007), R_* might, in some contexts, be a more useful parameter to estimate than R_0 .

Although not often stated explicitly, the factorisation of R_* in the form (2.4.1) is a non-trivial result and depends on some model assumptions, which are thoroughly discussed in Section 4.9 (a proof of the result, called Wald's identity for epidemics, is also reported in Section 4.9). Here, it is implicitly assumed that these assumptions are satisfied.

2.4.4 Model definition

Let the population be partitioned into m groups, each of size n : we call them households. Each person belongs to one and only one household. The total population size is $N = mn$ (see Figure 2.2).

Each infectious individual has probability p_H to infect each other susceptible individual in his household and a probability p_G , usually much smaller, to infect each other susceptible individual in the population (including individuals from his or her household). These are here referred to as *local* and *global infections* respectively. A *local* (or *global*) *infectious contact* is instead defined as a contact between two individuals which results in a local (or global) infection whenever the first individual is infectious and the second one is susceptible. A household is said to be *susceptible* if all the members are susceptible and *infectious* when at least one of its members is infectious. The epidemic process within an infected household, which is driven mainly (but not necessarily only) by local infections, is called *local epidemic*. A local epidemic ends when there are no more infectious members in the household.

Assume the population is wholly susceptible and the epidemic starts when one individual is infected. It is not investigated how this individual has been infected.

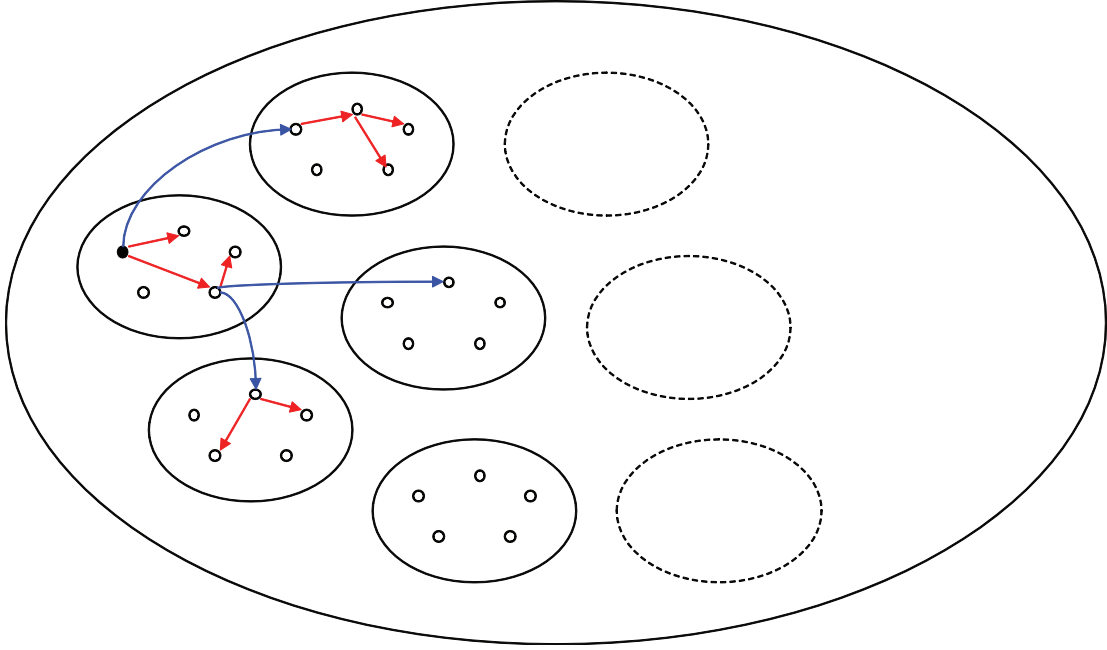


Figure 2.2: Schematic representation of a households model with households of size $n = 5$. The black dot represents the initial infective in the population, the red arrows represent local contacts and the blue arrows global contacts.

Infinite population limit

Assume now that the number of households m (and hence the population size N) is large. Therefore, in the early stage of the epidemic (i.e. when the proportion of infected households is small), the following observations hold:

1. the probability p_G of an infectious individual globally infecting a randomly chosen susceptible in the population is small;
2. the probability of an infectious individual making infectious contacts with an already infected individual is small, because the number of susceptibles is large;
3. the probability of an infectious person making global contacts towards other members of the same household is small;
4. the probability that two different people that are infected globally belong to the same household is small.

In the limit of an infinite number of households (and hence an infinite population), the previous events happen with probability tending to zero.

In this framework, the practical consequences are the following:

1. (from 3 and 4) once a household is infected, a local epidemic develops only through local contacts;
2. (from 4) any global infection hits a previously uninfected household;
3. (from 4) once a local epidemic has finished, no other global infection can hit that household, even if there are still susceptible members who escaped infection.

For the sake of completeness, a household in which a local epidemic has finished will be called *immune*. Note that this concept is well defined only in the limit of an infinite number of households and in the early phase of the epidemic (i.e. when the proportion of infected households is zero): in fact, when a positive fraction of the population has been infected, there is a positive probability of a reintroduction of the infection in an already infected household. The same household therefore will behave (in an infectious sense) differently from when it experienced the first epidemic wave and the new behaviour will be highly dependent on how many susceptibles escaped infection the first time.

Primary and secondary cases

Once a susceptible in a previously uninfected household is infected, he or she starts a household epidemic. Following Ball *et al.* (1997), the first infective is referred to as a *primary case* and each other person who is ultimately infected in this local epidemic is defined to be a *secondary case*. Note that this definition of secondary case is different from the usual one and therefore a basic reproduction number for the proliferation of individuals in this case is expected to be different from R_0 : as in Ball *et al.* (1997), it will be denoted by R_I . Note however that in the model considered in Becker & Dietz (1995) the two definitions coincide under the hypothesis that, in the case of a highly infectious disease, a

primary case immediately infects all the other member of the household, who are therefore all secondary cases in the common sense.

2.4.5 Reproduction numbers

As in the case of multitype models, the household model allows more than one reproduction number to be defined. Some of them are described here and compared.

Household reproduction number

Let μ_H represent as the average number of secondary cases in a household, i.e. the average number of initial susceptibles who are infected in a local epidemic started by one single infective. Numerical values for μ_H may be computed in terms of more basic parameters once assumptions are made on the local infection process and the individuals' disease history (see Section 3.7) but no closed formula for the final size distribution of an epidemic in a finite population (or at least for its average) seems to exist.

Also, let R_g denote the average number of global infections made by a single infectious individual throughout his or her infectious period in a totally susceptible population. Again, R_g may be a function of some more basic parameters concerning the infection process and the individual level.

Different studies concerning a basic reproduction number *for household* have been published (e.g. Becker & Dietz, 1995; Ball *et al.*, 1997, 2004; Becker *et al.*, 2005) and in the already established notation it is denoted here by R_* . R_* is defined as the average number of households infected by a “typical” household in a totally susceptible population. Wald's identity for epidemics (see Ball, 1986, or Section 4.9) leads to the already known result that

$$R_* = R_g(1 + \mu_H). \quad (2.4.3)$$

Note the linear dependence of R_* on R_g , which is a consequence of studying the interaction among the highest level structure present in the model, i.e. the household. The whole complexity of the local epidemic process is captured into μ_H .

Individual reproduction numbers

As mentioned above, another threshold parameter R_I , referring not to households but to individuals, can also be obtained. Although its usefulness is limited in the context of this thesis, its derivation and properties are presented anyway, as the same arguments are used in Sections 4.3 and 4.5 to derive and discuss the properties of a reproduction number for households in a model with households and workplaces.

The construction of R_I is the following. Recall from the considerations of Section 2.4.4 that, in a large population and in the initial phase of the epidemic, both primary and secondary cases generate primary cases (through global contacts) but only primary cases generate secondary cases (in the sense of Section 2.4.4 above). This two-type branching process is driven by the next generation matrix (see Section 2.3.2):

$$K = \begin{pmatrix} R_g & R_g \\ \mu_H & 0 \end{pmatrix} \quad (2.4.4)$$

where the element in position (i, j) represents the average number of individuals of type i infected by an individual of type j throughout his infectious period. R_I is defined as the dominant eigenvalue of this matrix and is therefore:

$$R_I = \frac{R_g}{2} + \frac{1}{2} \sqrt{R_g^2 + 4\mu_H R_g}. \quad (2.4.5)$$

Note that R_I has a more complex expression than R_* and does not depend linearly on R_g , the reason being that the attention is focused on the interaction between lower level structures (i.e. the individual) and therefore local and global infectivity are mixed in a complex way.

The two reproduction numbers are related according to the following

Theorem 2.4.1. *For all non-negative values of R_g and μ_H ,*

- $R_* < 1 \Leftrightarrow R_I < 1$
- $R_* = 1 \Leftrightarrow R_I = 1$

- $R_* > 1 \Leftrightarrow R_I > 1$,

i.e. R_* and R_I share the same threshold at 1 and are consistent in their threshold property.

Proof. The trivial case of $\mu_H = 0$ gives $R_* = R_I = R_g$. Instead, when $\mu_H > 0$, imposing $R_* = 1$ is equivalent to imposing $\mu_H = (1 - R_g)/R_g$, which implies $R_g < 1$. R_I is therefore given by

$$R_I = \frac{R_g}{2} + \frac{1}{2}\sqrt{(R_g - 2)^2} = \frac{R_g}{2} + \frac{1}{2}(2 - R_g) = 1.$$

When $R_* < 1$, then $\mu_H < (1 - R_g)/R_g$, which, since $\mu_H > 0$, still implies $R_g < 1$. $R_I < 1$ follows as the square root is a monotonic increasing function of its argument. $\square$

It has been noted above that this individual reproduction number is different from R_0 , except in some limiting cases.

The basic reproduction number R_0 can still be defined in this model, although care needs to be taken in describing what a “typical” infectious case is. Some authors (e.g. Ball *et al.*, 1997) consider a “typical” case as a randomly selected individual in a fully susceptible population. In this case, when $R_0 > 1$, the infection quickly spreads in the household of the initial case, but may still be unable to spread widely between households and may not result in a large epidemic: i.e. R_0 is not a useful threshold parameter.

If one wants to maintain the threshold property of R_0 , a possible approach is to let the infection spread through the household of the first randomly selected case in the population and then to define a “typical” case by averaging across all the individuals infected in the household. Unfortunately, with such a definition, the non-linearity induced by the quick depletion of susceptibles within the households makes R_0 analytically intractable, because the number of new cases generated by an infective varies according to how many susceptibles are still present in the household and is therefore difficult to compute. R_* does not suffer from this problem since it refers to entire households (the complexity due to the local depletion of susceptibles is captured in μ_H) and the household-to-household infection process becomes linear in the limit of an infinite number of households and in the initial phase of the epidemic (Ball *et al.*, 1997; Fraser, 2007).

2.4.6 Average final size

As far as the average final size conditional on a large epidemic occurring is concerned, an analytical result first appeared in Ball *et al.* (1997), although rather technical in nature. However, subsequent investigations (Ball & Neal, 2002) highlighted a more intuitive form, based on the concept of local susceptibility set. The *local susceptibility set* of individual i is a random set consisting of individual i plus all individuals that, if infected through a global contact, would infect individual i through a chain of local contacts (i.e. household contacts). Therefore, the probability of i escaping infection equals the probability of all individuals in i 's local susceptibility set escaping global infections.

Given that individuals are all identical, the only relevant piece of information about the local susceptibility set of individual i is its (random) size S_i and, $\forall i$, S_i is independent and identically distributed according to a random variable S . In addition, given that the transmission process is symmetric, S_i has the same distribution of a local epidemic final size distribution started by a single initial infective (included), the average of which is given by $\mathbb{E}[S] = 1 + \mu_H$. No closed formula for such a distribution seems to exist, but a computational technique for its numerical computation, developed by Ball (1986) (for other more technical but more efficient ones, see Lefèvre & Picard, 1990, 1999) is reported in Section 3.7.

In the spirit of Sections 2.2.2 and 2.3.3, the equation for average proportion $z \in [0, 1]$ of initial susceptibles who are ultimately infected in a large epidemic spreading in an infinite population of households (assume the epidemic starts with only a finite number of infectives) can be read as a balance equation between the fraction of the population escaping the infection and the probability that an initial susceptible survives the infection pressure of all cases infected throughout the epidemic. If $f_S(s) = \mathbb{E}[s^S]$ denotes the probability generating function of S , then the probability that each individual of the local susceptibility set S_i of a selected individual i escapes infection is given by $f_S(e^{-R_g z})$, as each individual in S_i escapes global infections independently with probability $e^{-R_g z}$. Therefore z is the

(largest) solution of the balance equation:

$$1 - z = f_S(e^{-R_g z}). \quad (2.4.6)$$

Further analytical results are available on the finer structure of the final size, i.e. on how infected cases are distributed into households, but they are not reported here; see Ball & Neal (2008) for further details.

2.4.7 Interventions targeted at households

In a homogeneously mixing model, one of the most practically useful properties of R_0 is the simple relationship it has with the critical vaccination coverage, i.e. the proportion of individuals in the population to vaccinate in order to avoid major epidemics:

$$p_{crit} = 1 - \frac{1}{R_0}.$$

In a households model, since R_* refers to interaction between households, it is strictly linked with the proportion of households that has to be vaccinated in order to avoid major epidemics. In fact, assume each household is successfully vaccinated (i.e. totally immunised) with probability p and is left unvaccinated with probability $q = 1 - p$; then any global contact hits an unvaccinated household with probability q . Therefore, after the vaccination, R_g is reduced to $R_g^V = q R_g$ and (μ_H is not affected) R_* is reduced to $R_*^V = q R_*$. To avoid major epidemics, the vaccination should reduce R_*^V below 1 and the critical vaccination coverage for household is obtained when $R_*^V = 1$:

$$p_{crit}^* = 1 - \frac{1}{R_*}.$$

Unlike R_* , R_I refers to individuals and hence should be related to the critical vaccination coverage for individuals; however, successful vaccination of each individual with probability p has an effect on both R_g and μ_H . As before, $R_g^V = q R_g$ because any global contact hits an unvaccinated individual with probability q , but μ_H^V has no simple relationship with μ_H . However, Theorem 4.5.1 in Section 4.5 states that $\mu_H^V \leq q \mu_H$ (*herd immunity effect*; see

e.g. Andersson & Britton (2000)), and hence a vaccination coverage $1 - 1/R_I$ reduces R_I to $R_I^V \leq 1$, i.e. in the worst case it is just sufficient to avoid major epidemics but in other cases could bring R_I^V even below 1. These comments are not developed any further here, but are delayed until Section 4.5, since the interest focuses more on intervention strategies targeted at households rather than on individuals.

When households have different sizes, the same result for the critical vaccination coverage for households holds, provided that households are selected at random (i.e. independently of their size), and fully vaccinated.

However, as larger households are more susceptible and more infectious, vaccination strategies preferentially targeting larger households generally achieve better results. A thorough discussion is delayed again until Section 4.8. Note in any case that vaccination of entire households is in general administratively convenient, but not optimal in term of using a scarce vaccine. The so-called *equalising strategy* (e.g. see Ball *et al.*, 1997), which consists in leaving the same number of unvaccinated individuals in each household, performs better as it fully exploits the herd immunity effect within households. Furthermore, House & Keeling (2008) discuss the benefits obtained by targeting children, a form of vaccination which is at the same time practically viable, very close to the equalising strategy and also aimed at the class of the most strongly mixing individuals.

Other forms of intervention, different from vaccination, are more difficult to analyse, given that time has not been explicitly considered in this households model; see also Section 4.6.

2.4.8 Extension to variable household size

The model can be readily extended to variable household sizes. The only difference from the fixed size case is that the parameter μ_H , the average final sizes of a household epidemic, takes into account this distribution. In particular, let h_n be the probability that a randomly selected household has size n . Then the probability that a randomly chosen individual

belongs to a household of size n is

$$\pi_n = \frac{nh_n}{\sum_n nh_n}.$$

Note that the probabilities that the selected individual belongs to households of different sizes are weighted according to the number of individuals they contain and this takes into account the fact that larger structures are more likely to be infected than smaller ones; distributions like this are sometimes referred to as *size-biased distributions* (see Becker & Dietz, 1995; Ball *et al.*, 1997). Let $\mu_H^{(n)}$ be the average final size of an epidemic in a household of size n , starting with a single initial case (not counted) and $n-1$ susceptibles. Then μ_H is defined as

$$\mu_H = \sum_n \mu_H^{(n)} \pi_n.$$

The model can be thought of as a form of multitype model with separable mixing, where households of size n have susceptibility proportional to n but infectivity proportional to $\mu_H^{(n)}$. When the within household transmission tends to infinity (e.g. as in Becker & Dietz, 1995), then all secondary cases are infected immediately and the infectivity of the household is proportional to n . In this limiting case, the formula for R_* has the same structure of that for R_0 of Equation (2.3.5).

2.5 Other socially structured models

The households model is not the only type of model which explicitly takes into account the presence of a social structure. It has already been commented on the fact that multitype models can be used with the same purpose (see Section 2.4.2 and Roberts, 2004; Roberts *et al.*, 2007; Aldis & Roberts, 2005). However, other possible model choices are briefly discussed here.

2.5.1 Metapopulation models

The terms *metapopulation models* or *patch models* usually refer to models characterised by the presence of a large population partitioned into a fixed number of subpopulations. Ho-

homogeneous mixing is assumed in each subpopulation and a matrix of coefficients defines the intensity of the coupling between the subpopulations. While the number of subpopulations is usually assumed to be relatively small, the size of each subpopulation is considered large. In addition, the between-population coupling is often assumed to be much weaker than the within-subpopulation infectivity, and the within-subpopulation dynamics can therefore be treated as those of a single large population (Section 2.2): R_0^j , the basic reproduction number of subpopulation j , determines if an epidemic infecting a major fraction of population j can occur and with which probability and the real-time growth rate and the average final size can be studied using the standard techniques. Note that, once a large subepidemic is taking place, further reintroductions from other subpopulations have negligible impact on it.

Metapopulation models have been extensively used in the literature: a clear example is given by Rvachev & Longini Jr (1985). A stochastic extension of the model has also been considered (see Colizza *et al.*, 2007).

Note that, apart from a rather different interpretation and derivation, a metapopulation model is formally equivalent to a multitype model where the elements $k_{ij}, \forall i \neq j$, of the next generation matrix are much smaller than the elements $k_{ii}, \forall i$, on the diagonal.

2.5.2 Overlapping groups models

A direct extension of the households model is achieved when individuals belong to more than one type of mixing group and different groups are allowed to overlap (Ball & Neal, 2002). The levels of mixing are still considered to be only two, where the first is given by any form of local mixing (within any type of group included in the model), superimposed on a second one, typically a background global homogeneous mixing in the population at large. This is much more than a simple refinement, as it allows for new possible epidemic scenarios to appear. For example, in the households model, an epidemic remains contained within a household unless one of its members transmits the infection to another individual in the population. In an overlapping group model, the infection can be transferred from

household to household via local contacts within other groups and can spread widely even if no global infection occurs. The local transmission does not need to explicitly take into account a specified social structure, but can be also represented by a generic network, which is still superimposed to the background global homogeneous mixing: this is the case, for example, of Ball & Neal (2008).

The simplest form of overlapping groups model that can be considered is the one containing only two types of local groups and is the main topic of this thesis: in particular, such groups are thought to be households and workplaces/schools. However, when studied from a theoretical point of view, the identification of such groups is simply a matter of convenience.

2.5.3 Hierarchical network models

Of course, overlapping partitions can be combined with more than one level of mixing to obtain even more general and complex models. Hierarchical social structures, where for example individuals are gathered in households, households in communities, communities in cities and cities in countries are also possible. However, given the level of complexity, little theoretical work has been done on this topic and model properties are usually explored using numerical simulations.

Among the hierarchical models that tend to provide a balance between parsimony and sufficient complexity to induce curious behaviours, a relevant example is given by Watts *et al.* (2005); other examples, all based on the same model structure, are given in Grabowski & Kosiński (2004, 2005, 2006) and Grabowski & Rosińska (2006).

Finally, when the model aim is that of being as realistic as possible and of including all the details that are deemed relevant for transmission, mathematical analysis becomes impossible and large scale individual-based microsimulations represent the only viable resource (Elveback *et al.*, 1976; Longini *et al.*, 2004, 2005; Ferguson *et al.*, 2005, 2006; Germann *et al.*, 2006; Riley & Ferguson, 2006).

Chapter 3

Small populations

Since both the households model (Section 2.4) and the households-workplaces models of Chapter 4 are characterised by the presence of small groups where saturation effects play an important role in the spread of the infection, it is convenient to discuss in detail the properties and the results concerning epidemics in small populations.

3.1 Introduction

Despite the analytical complexity due to the intrinsic non-linear behaviour of the infection process, which cannot be avoided when the population considered is small, many remarkable results have been obtained in the past (Ball, 1983b, 1986; Longini *et al.*, 1982; Picard & Lefèvre, 1990; Diekmann *et al.*, 1990; Anderson & May, 1991; Ball *et al.*, 1997; Diekmann & Heesterbeek, 2000; Andersson & Britton, 2000; Hethcote, 2000).

Some of these results are derived by exploiting “tricks” that allow the transformation of the actual infection process into another one, which does not correspond to what can be observed in reality, but that has the advantage of being analytically tractable. Some examples are the Sellke construction (Sellke, 1983) and the Scalia-Tomba imbedded representation (Scalia-Tomba, 1985, 1990), based on the concept of infection pressure. Another one is the possibility of ignoring the temporal dynamics of the epidemic process under cer-

tain suitable conditions. To our knowledge, this was first suggested by Ludwig (in Ludwig, 1975, but already appearing in Ludwig, 1974).

On the one hand, the Sellke construction and the imbedded representation by Scalia-Tomba has been well referenced and has been often reported and explained when used (for example in Andersson & Britton, 2000, or Ball *et al.*, 1997). On the other hand, the idea proposed by Ludwig is often not acknowledged and its presence behind many important results is hidden by chains of subsequent references; there are of course some exceptions, when authors have clearly stated its importance (Hethcote & Tudor, 1980; Longini & Koopman, 1982; Longini *et al.*, 1982; Scalia-Tomba, 1986; Picard & Lefèvre, 1990; Andersson, 1999). The main reason for this lack of referencing to Ludwig's idea is that the result is well-known among stochastic modellers. However, the result has many important epidemiological implications which may be of interest to a broader audience as well.

Naively speaking, Ludwig's "trick" consists in observing that, for a disease leading to permanent immunity after recovering, it is possible to associate to an epidemic process, occurring in continuous time, another (fictitious) contact process in discrete-time steps that has the same final size distribution (Ludwig, 1975). In other words, as far as the final size distribution is concerned, the absolute times at which various events occur can be neglected.

Such a new contact process is often referred to in the literature as "generation-based description of an epidemic". However, as will be emphasised below, the word "generation" may be misleading. For this reason, it is worth introducing the concept of *rank* of an infective, as Ludwig did in his paper (Ludwig, 1975): initial infectives are assigned rank zero; all those individuals that have an infectious contact with an initial infective (independently of whether they have already been infected or not by somebody else) are assigned rank one. Analogously, individuals are assigned rank $n + 1$ ($n \geq 1$) if they avoid infectious contacts with infectives of rank less than n and have an infectious contact (not necessarily an effective infection) with an infective of rank n . Ludwig argues that this rank-based process has the same final size distribution as the real epidemic process, since "an infective is an infective regardless of how his rank is assigned" (Ludwig, 1975).

Two different problems initially motivated this study. The first one refers to the fact that, since generations overlap during a real-time epidemic, an individual experiences different forces of infection over time in the real-time process and in the fictitious rank-based process, and therefore it is not at all intuitive why the two processes should lead to the same final size distribution (Diekmann & Heesterbeek, 2000, p. 27-28). Both processes are commonly considered in the literature (e.g. Diekmann & Heesterbeek, 2000), but the reasons and the conditions that allow one to switch between the two are often not clearly stated or lost in references to previous results and in mathematical details.

The second problem concerns an epidemic in a household where individuals are also allowed to be infected from outside. Ball *et al.* (1997) obtained a remarkable result for the average final size of an epidemic in a households model (Section 2.4), in the limit of an infinite population. The result is based upon the following:

Theorem 3.1.1. *Consider a group consisting of n susceptibles and a initial infectives in which the susceptibles can be infected both within and from outside the group. Assume that each susceptible escapes infection from outside independently of each other and with probability π . Then the final size distribution is the same as that of a group with no infections from outside, consisting of $n - Y$ susceptibles and $a + Y$ initial infectives, where Y is a binomial r.v. on n trials and probability of success $1 - \pi$.*

Theorem 3.1.1 follows directly from the previous problem, but it is even less intuitive, as it may well happen that a new individual is infected from outside when a first epidemic wave has already finished sweeping through the household. In this case, the number of new infections generated by the infectious case depends on how many susceptibles escaped the previous epidemic. Furthermore, concepts like that of “generations of infectives” require more careful consideration in this setting. Although apparently different, both the problems can be understood in terms of Ludwig’s result, as already suggested in the study by Longini & Koopman (1982), where a situation similar to that of Theorem 3.1.1 was first considered (see below).

3.2 Consequences of Ludwig's result

Although the argument that Ludwig presents in his paper is simple to understand and might even be considered “obvious” (Ludwig, 1975), the results based on it have grown more and more complex and, if considered alone, they are far from being intuitive. Some of them are described here.

A first direct consequence of Ludwig's argument is that the presence of a latent period of any form does not affect the final size distribution of an epidemic, despite having an impact on the system dynamics. For this reason, most stochastic epidemic models considered in the literature are of the *SIR* type, i.e. assume that individuals are immediately infectious after infection without experiencing a latent stage (e.g. see Ball, 1986; Andersson & Britton, 2000).

A second important consequence of Ludwig's argument is that final size results obtained for the Reed-Frost model can be extended to a wide variety of other models, despite the unrealistic temporal dynamics. The other models described in Sections 2.1.3 and 2.1.4 (which are also revisited in Sections 3.4.1 and 3.4.2 below), for example, present more complex and in general more realistic temporal dynamics, but the argument given by Ludwig leads to the conclusion that the final size distributions of these models can be obtained by studying a corresponding Reed-Frost model (or its extension to variable susceptibility and infectivity) and this justifies the great attention the Reed-Frost model received in the literature.

A third consequence concerns the models with a social structure. Longini and Koopman (Longini & Koopman, 1982) proposed a model that has been referred to as the independent-households model (Demiris & O'Neill, 2005). The attention is focused on a single household and the epidemic within the household is modelled using a Reed-Frost model enriched by the possibility that individuals are also infected from outside (community transmission). The probability of infection from outside is assumed to be the same for every susceptible and to be constant in time, i.e. epidemics in different households develop independently of one another.

The model has been widely used in estimating the secondary attack rate, i.e. the probability of transmission within the household (Longini *et al.*, 1982), the probability of community transmission (Longini *et al.*, 1982), the efficacy of vaccine and prophylaxis in reducing susceptibility, infectivity and development of symptoms after infection (Halloran, 2003; Halloran *et al.*, 2007). It has also been used in comparing different designs for vaccine efficacy studies (Datta *et al.*, 1999; Yang *et al.*, 2006).

Since the independent-households model is basically of Reed-Frost type, all the conclusions drawn from it would have to be taken with extreme caution as they would be highly dependent on the often unrealistic temporal dynamics of the Reed-Frost model. In particular, results might be reasonable for measles but not for influenza, because of the short latent period and the trend that different generations have to quickly overlap (Hope Simpson & Sutherland, 1954). Ludwig's argument drastically increases the robustness of all these results based on final size inference, as it implies that the final size distribution obtained from the Reed-Frost model is the same as that of other models with different temporal dynamics, irrespectively of the fact that different generations can overlap.

Furthermore, the assumption of the independent-households model that each susceptible has the same constant probability of being infected from outside is realistic only in a very large population of households (Demiris & O'Neill, 2005). Therefore, the validity of the above results would be limited to the case of household epidemics developing independently of one another. However, Ludwig's argument proves again to be a crucial result of robustness.

In fact, the above results obtained from final size inference should be based, instead of on the independent-households model, on the households model with two levels of mixing described in Section 2.4 (see also Ball *et al.*, 1997), which is more realistic as it allows the probability of infection from outside to vary throughout the epidemic. Despite the complexity, important results have been obtained from this model, among which is the one, already referred to above, concerning the average final size in the limit of an infinite number of households (Ball *et al.*, 1997). This result is based on Theorem 3.1.1 and on the

fact that, despite the probability of a susceptible individual being infected from outside the household changes over the course of the epidemic, the probability of being ultimately infected during the whole epidemic is equal for every susceptible and depends on the final size of the whole epidemic but not on its dynamics (see Section 3.8).

Again, the result has its roots in Ludwig's argument, as it heavily exploits the possibility of ignoring the temporal dynamics of the epidemic within and outside the households. The conclusion is that the results concerning the final size of the independent-households model are the same as those of the more realistic model with two levels of mixing. This extends the validity of all the important results of final size inference, which are based on the independent-households model, to the more realistic case of the model with two levels of mixing, in which household epidemics in general do not develop independently of one another.

3.3 Problem description and terminology

For a disease leading to permanent immunity after recovery, instead of the epidemic process itself, Ludwig (1975) suggested to study a different contact process. This new process is based on the concept of *infectious contact* and on the notion of *rank* of an infective. An *infectious contact* is defined as a contact from an infectious individual towards any other individual, that would lead to an infection if the second individual is susceptible at that time.

The *rank* of an infective is constructed in the following fashion: initial infectives are assigned rank zero; all individuals that experience an infectious contact from an initial infective during their infectious period are assigned rank one. Note that some of these infectious contacts result in an effective infection, while others may not, namely those directed towards individuals who have already been infected by some other non-initial infectives before experiencing their infectious contact from an initial one.

In Figure 3.1, for example, individual 1 is the only initial infective and therefore has rank 0, individual 2 has rank 1; finally, individual 3 has rank 1, despite having been actually

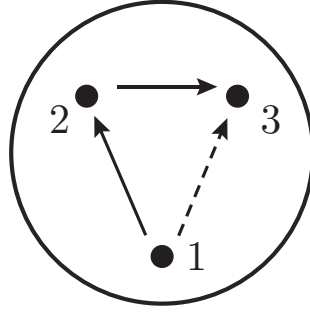


Figure 3.1: Schematic representation of an epidemic in a population of size 3: bold arrows represent infections, dashed arrows represents infectious contacts that fail to do not result in an infection. Individual 1 is the primary case.

infected by individual 2, because of the infectious contact from individual 1. Analogously to the assignment of rank 1, an individual is assigned rank $n + 1$ ($n \geq 1$) if he or she avoids infectious contacts from infectives of rank less than n and has an infectious contact (not necessarily an effective infection) from an infective of rank n .

Note that, when an individual experiences more than one infectious contact, only the first one is deemed to lead to an infection: in Figure 3.1, the infectious contact between individuals 1 and 3 can only have occurred after the infectious contact with individual 2, otherwise the real infection would have come from individual 1. Therefore the real infection process is highly dependent on the times at which events occur. On the other hand, the rank-based construction depends only on the presence or absence of infectious contacts and not on their times of occurrence.

It is worth at this point defining some terminology used in the remainder of the chapter: in general, the word *process* will be used to refer to a set of events and their times of occurrence; other words like *network*, *tree*, etc. will not contain any information about times of occurrence. A *infectious contact process* is the process describing all the infectious contacts between individuals and their times of occurrence: some of them, namely those from infectious individuals towards susceptible ones, will lead to an infection, but others may not. The *infection process* is the reduction of the infectious contact process only to those contacts that cause infection and their time of occurrence. The *infection tree* is what

is obtained from the infection process once the times of infection are ignored. Infected individuals can then be gathered in *generations*. All these concepts are related to what can be observed in a real-time epidemic.

If, from the infectious contact process, individuals are assigned a rank according to Ludwig's construction and only infectious contacts that are relevant for the rank assignment and their times of occurrence are considered, it is obtain what is here called the *rank-based process*. Ignoring times will generate a *rank-based contact tree*. Once this tree is defined, the individuals in it can be gathered in "generations". Note that these "generations" may in general be different from those obtained from the infection process (see Figure 3.1). For the sake of clarity, the term *generations* will be reserved for the real generations of infectives, i.e. those obtained from the infection tree, and the term rank will be used when referring to Ludwig's construction.

Such precise definitions are necessary because the infection tree and the rank-based contact tree may be topologically different; therefore, individuals with the same rank are not necessarily in the same generation. However, in the literature, the rank-based construction is almost always referred as the "generation-based description of the epidemic". Analogously, sentences like "studying an epidemic in generations" are ambiguous, since they probably refer to the ranks obtained from Ludwig's construction and not to the trivial process of grouping in generations (in the real-time sense) the individuals infected in the real-time infection process. Furthermore, when discrete-generation models like the Reed-Frost one are used, the two constructions coincide and distinctions in terminology are not needed.

3.4 Models for the infectious contact process

In this section, all the various models for the infectious contact process between individuals (i.e. disease histories) described in Sections 2.1.3, 2.1.4 and 2.3.1 are reviewed in the light of Ludwig's result. Common assumptions considered here for all these models are the following: the disease leads to permanent immunity after recovery; the population consists

of n individuals, labelled with numbers from 1 to n and individual 1 starts being infectious at time $t = 0$. It is not investigated how he or she acquired infection.

3.4.1 Standard *SIR* epidemic model

Recall from Section 2.1.4 or from Andersson & Britton (2000) that in the *standard SIR epidemic model*, individuals mix homogeneously and, upon infection, they experience infectious periods that are independent and identically distributed according to a random variable I . During the infectious period, each infective makes infectious contact with any other given individual at the points of a homogeneous Poisson process with rate λ . After the infection period has terminated, the individual recovers and does not participate to the epidemic spread any further.

Note that, according to this definition of the model (other authors have proposed slightly different ones), the individuals' behaviour is described only during their infection period, since this is the only part of individuals' behaviour that is relevant for the spread of the epidemic. However, it is clear that they interact with each other also before being infected and (assuming the disease does not cause death) after recovery.

Assume that the disease is sufficiently mild that it does not influence the social interactions of an individual, i.e. infected individuals do not behave differently from susceptible or recovered ones (this assumption can be relaxed, see the comment at end of Sec. 3.4.2). For each pair of individuals (i, j) , it is then possible to construct a sequence $(\tau_{ij}^{(k)})_{k \in \mathbb{N}}$ of all the times of occurrence of the contacts from i to j after time $t = 0$. If i is infected at time T_i and his or her infectious period lasts for a time interval I_i , a function $p_i(t)$ can be associate to i , with a constant value p for $t \in [T_i, T_i + I_i]$ and 0 otherwise, which describes the probability that the infection is transmitted across a contact.

From the set of all the times of contacts, the values of I_i 's associated to each individual and a sequence of realisations of independent Bernoulli trials with success probability p (used to describe which contacts would be able to transmit infection) it is possible to follow the epidemic in real time, observe $\forall i$ if i acquires infection and, in this case, store the time

T_i of infection.

Note that, so far, the development of the epidemic is observed in calendar time and that no assumption has been specified to describe how contacts occur in time. It is now assumed that, for each pair (i, j) , contacts from i to j occur at the points of a Poisson process with rate c . Poisson processes relative to other pairs are mutually independent by assumption and independent of the sequence of Bernoulli trials and of the length of the individual's infectious periods. This assumption allows great simplification, thanks to the fact that the intervals between subsequent contacts from i to j are exponentially distributed and therefore satisfy the memoryless property: for this reason, the probability that a contact occurs at a time τ after the infection of i is exponentially distributed with parameter c , independently of the absolute time T_i of his infection. This property allows to switch from a calendar time perspective to a time since infection one.

The standard *SIR* model is now a consequence of the above more elementary assumptions since, during his infectious period, i makes infectious contacts towards j at the point of a homogeneous Poisson process with rate $\lambda = pc$.

Note that, if more than one infectious contact occurs from i to j during i 's infectious period, only the first one may be relevant: all the subsequent ones are superfluous, since j has been already infected by the first contact from i or was already infected before. Therefore, the full epidemic process can be described by the set $\{\tau_{ij}, i, j = 1, \dots, n, i \neq j\}$, where $\tau_{ij} > 0$ represent the time of the first infectious contact from i to j since the (hypothetical) infection if i and are drawn from an exponential distribution with parameter λ . If i ultimately escapes infection, T_i is not defined and all the times $\tau_{ij}, j = 1, \dots, n, j \neq i$, are superfluous.

It may be convenient to represent all these times in a $n \times n$ matrix with entries in $\mathbb{R}^+ \cup \{*\}$, where $\mathbb{R}^+$ represents the strictly positive real numbers and the symbol $*$ indicates an element that has no meaning for the epidemic process, for example all the elements on the main diagonal. Once a matrix (τ_{ij}) and the vector (I_i) representing the duration of the (hypothetical) infectious periods are specified, all the elements τ_{ij} s.t. $\tau_{ij} > I_i$ can be

substituted with $*$ because they refer to infectious contacts from i occurring after he or she has recovered.

The matrix so obtained represents the full infectious contact process (in the terminology of Section 3.3). Once it is specified, both the real time infection process and Ludwig's rank-based process can be constructed and, from these, the infection tree and the rank-based contact tree can be considered. Alternatively, a graphical representation of the full infectious contact network can be obtained by drawing an arrow for every non- $*$ element in the matrix (τ_{ij}) : from it, the rank-based contact tree (but not the infection tree any longer) can then be constructed. Figure 3.1 in Section 3.3 exploits a graphical representation of both the infection tree (in bold arrows) and the infectious contact network (in dashed arrows) on the same set of nodes in order to provide, through an example, an intuitive explanation of Ludwig's result.

3.4.2 Models with time varying infectivity

In the simplest case, referred to as non-random in Section 2.1.3, all individuals mix homogeneously and, at time τ after infection, they make infectious contact with any other given individual (one-to-one) at a rate $\lambda(\tau)$, where the function $\lambda(\tau)$ is the same for each infective. Similarly to the case of the *SIR* model treated in the previous section, the social interaction is not modelled before infection or after recovery.

However, as in the previous section, one can assume that contacts between individuals are still made at the points of a homogeneous Poisson process with rate c , while now the probability of transmission across a contact changes as a function $p(\tau)$ of time since infection, equal for each infective. As before, thanks to the memoryless property of the inter-contact times, it is possible to switch from an absolute time perspective to a time-since-infection one and the usual time-since-infection model is recovered by defining $\lambda(\tau) = cp(\tau)$.

The main difference from the standard *SIR* model is that $\lambda(\tau)$ may be non-zero $\forall \tau > 0$. Strictly speaking, in this case the infectives never recover, although biological arguments

imply that $\int_0^{+\infty} \lambda(\tau) d\tau$ is finite and therefore that the infectivity $\lambda(\tau) \rightarrow 0$ sufficiently fast as τ increases. Even if there is no recovery, for each pair (i, j) , there is still a positive probability

$$q = e^{-\int_0^{\infty} \lambda(\tau) d\tau}$$

that j escapes infectious contacts from i . The same matrix (τ_{ij}) of the previous section can be defined: $\forall i, \tau_{ii} = *$ and $\forall i, \forall j, \tau_{ij} = *$ with probability q and with probability $p = 1 - q$ it is a real positive number drawn from the density distribution

$$\frac{\lambda(\tau) e^{-\int_0^{\tau} \lambda(s) ds}}{\int_0^{\infty} \lambda(\tau) e^{-\int_0^{\tau} \lambda(s) ds} d\tau} = \frac{\lambda(\tau) e^{-\int_0^{\tau} \lambda(s) ds}}{1 - q},$$

i.e. from the normalised unconditional probability density of j experiencing an infectious contact from i at time τ after i 's infection.

Note that the assumption that the disease is mild enough not to interfere with the contact process can be relaxed: in fact, we could consider a contact rate $c(\tau)$ that changes in time, therefore making the contact process non-homogeneous. However, since the contact rate changes only as a function of time since infection, the switching between an absolute time perspective and the time-since-infection one is still possible since, after infection of i , the first infectious contact towards j occurs at the first point of a non-homogeneous Poisson process with rate $\lambda(\tau) = p(\tau)c(\tau)$.

3.4.3 Variable susceptibility and infectivity

The approach adopted in the previous section can be extended to allow for many forms of heterogeneity by setting, for each pair (i, j) , a function of time since infection $\lambda_{ij}(\tau)$, being it deterministic or randomly drawn from a suitable set of non-negative functions.

Such a generalisation allows Ludwig's result to be applicable to most of the models for diseases leading to permanent immunity after recovery appearing in the literature of stochastic modelling. Firstly, it includes all multitype models where the infectivity and susceptibility depend on some static characteristics of the individuals: in this case, individuals of type k are characterized by a susceptibility a_k and an infectivity profile $b_k(\tau)$. Note

that, since the type of an individual is independent of the infection process, in particular the infectivity of an individual is independent of the time of infection and the identity of the infector.

A second class of models that can be considered is that characterised by randomly drawn susceptibility and infectivity. More formally, this is the case when, for each individual i , the susceptibility is a random variable A_i , taking values in a suitable set $\mathcal{A} \subseteq \mathbb{R}^+$, and the infectivity profile since infection is a random trajectory $B_i(\tau)$, drawn from a set of possible disease histories $\mathcal{B}$ (non negative integrable functions on $\mathbb{R}^+$). Ludwig's result holds whenever, for each individual i , it is possible to generate *a priori* (i.e. before the epidemic starts) the realizations a_i and $b_i(\tau)$ of A_i and $B_i(\tau)$: therefore, $\forall i, A_i$ and $B_i(\tau)$ are required to be independent of calendar time and $\forall i \neq j, B_i(\tau)$ and $B_j(\tau)$ must be independent of each other, so that no event occurring during the epidemic can affect such distributions.

A third kind of models is obtained when the infection rate between individuals is determined by the underlying (static) social structure (e.g. households models, overlapping group models, etc.). The simplest example is that of a households model where, denoting by $\lambda_G(\tau)$ the infectious profile from an infective towards each susceptible in the population (global infectivity) and $\lambda_H(\tau)$ the additional infectivity towards each susceptible within the same household (household infectivity),

$$\lambda_{ij}(\tau) = \begin{cases} \lambda_G(\tau) + \lambda_H(\tau) & \text{if } i \text{ and } j \text{ share the same household,} \\ \lambda_G(\tau) & \text{otherwise.} \end{cases}$$

A fourth very general type of models is obtain by combining all the previous cases, i.e. by allowing the sets A and B from which susceptibility and infectivity are randomly drawn to depend on the individuals involved in the infection process. In this case, $\lambda_{ij}(\tau) = a_j b_i(\tau)$, where $a_j \in A_{ij}$ and $b_i(\tau) \in B_{ij}$. This case includes households models (see Section 2.4 or Ball *et al.*, 1997) or more general overlapping groups models (e.g. the households-workplaces model of Chapter 4; see also Ball & Neal, 2002). The households model is a

particular case where $A_{ij} = 1$ and

$$B_{ij} = \begin{cases} \{\lambda_G(\tau) + \lambda_H(\tau)\} & \text{if } i \text{ and } j \text{ share the same household,} \\ \{\lambda_G(\tau)\} & \text{otherwise.} \end{cases}$$

For all these models, analogously to what has been done in the previous section, it is first necessary to draw the functions $\lambda_{ij}(\tau), \forall i \neq j$, then define $q_{ij} = \int_0^\infty \lambda_{ij}(\tau) d\tau$, and finally set $\forall i, \tau_{ii} = *$ and $\forall i, \forall j, i \neq j, \tau_{ij} = *$ with probability q_{ij} and, with probability $p_{ij} = 1 - q_{ij}$, randomly draw it from the density distribution

$$\frac{\lambda_{ij}(\tau) e^{-\int_0^\tau \lambda_{ij}(s) ds}}{\int_0^\infty \lambda_{ij}(\tau) e^{-\int_0^\tau \lambda_{ij}(s) ds} d\tau} = \frac{\lambda_{ij}(\tau) e^{-\int_0^\tau \lambda_{ij}(s) ds}}{1 - q_{ij}}.$$

3.5 A particular case

A simple example in a population of size $n = 3$ is considered in this section to remark that both the real-time and the rank-based process have the same final size distribution but that the rank-based process is much easier to study. Although the following example refers to the standard *SIR* epidemic model discussed in Section 3.4.1, any consideration, apart from the technical details about how the probability of various epidemics is computed, holds for all the models of Section 3.4.

In the graphical representation of Figure 3.1, bold arrows represent real infections and dashed arrows represent infectious contacts that do not lead to an infection. Consider an epidemic with final size 3. On the graph with nodes 1, 2 and 3, two bold arrows must reach nodes 2 and 3 and the possible configurations are (1, 2), (1, 3), (1, 2), (2, 3) and (1, 3), (3, 2). For each combination of bold arrows, all the remaining four directed links between nodes can be occupied or not by a dashed arrow, giving a total of $3 \cdot 2^4$ possible graphs. However, any dashed arrow towards node 1 is superfluous for the epidemic, and is therefore convenient to ignore the presence or absence of these arrows, thus reducing the number of graphs to 12, each of which actually represents four graphs. The 12 graphs are represented in Figure 3.2. In the terminology of Section 3.3, they are all the relevant infectious contact networks.

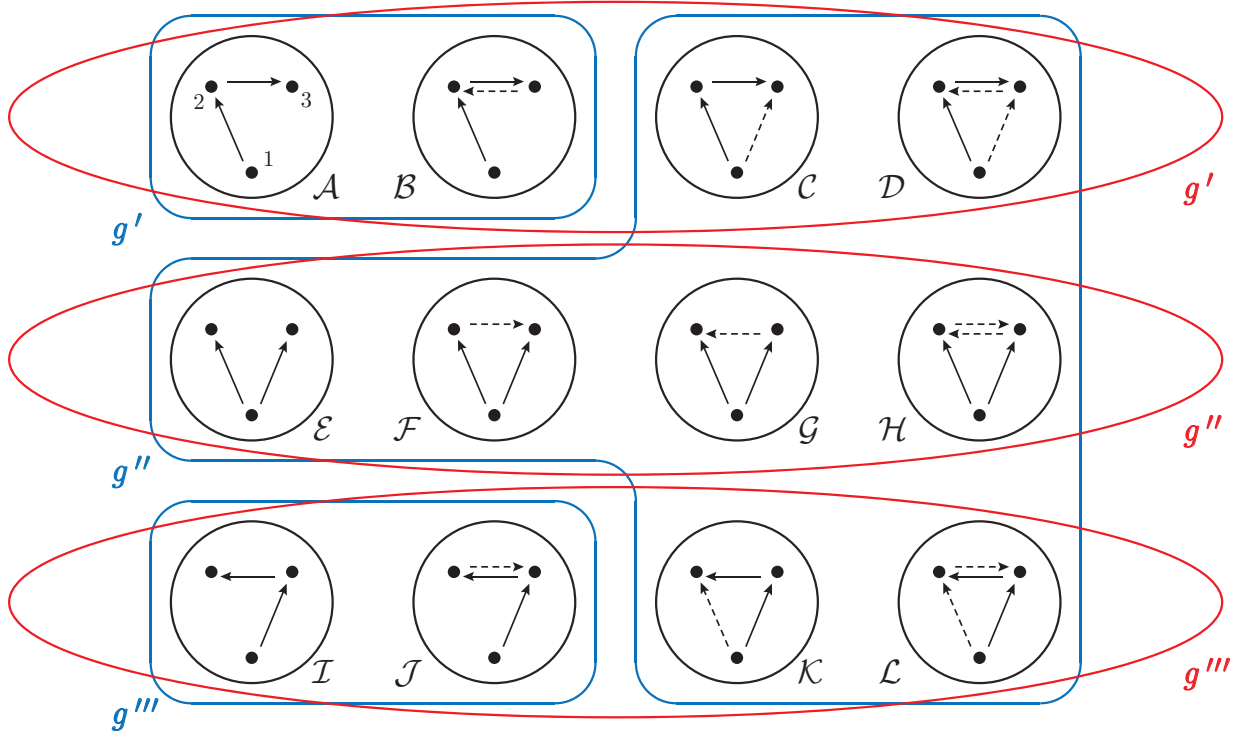


Figure 3.2: All the possible (relevant) infectious contact networks giving final size $f = 3$ in a population of size $n = 3$. Individual identification labels are reported only in $\mathcal{A}$ but are the same for all the other cases. Epidemics leading to the same list $g' = (\{1\}, \{2\}, \{3\})$, $g'' = (\{1\}, \{2, 3\}, \{\})$ or $g''' = (\{1\}, \{3\}, \{2\})$ of individuals in each generation according to the real-time approach are grouped in red. Epidemics leading to the same list g', g'' or g''' of individuals with the same rank according to the rank based process are grouped in blue.

Assuming the *SIR* infectious contact model of Section 3.4.1, it is possible to explicitly calculate the probability of each infectious contact network in Figure 3.1 to occur. For example, conditioning on the values of I_1, I_2, I_3 ,

$$\begin{aligned}
 P(\mathcal{A} \mid I_1, I_2, I_3) &= \\
 &= \mathbb{P}(\tau_{12} < I_1, \tau_{13} \geq I_1, \tau_{23} < I_2, \tau_{32} \geq I_3, \text{ any value of } \tau_{21} \text{ and } \tau_{31} \mid I_1, I_2, I_3) \\
 &= \mathbb{P}(\tau_{12} < I_1 \mid I_1, I_2, I_3) \mathbb{P}(\tau_{13} \geq I_1 \mid I_1, I_2, I_3) \times \\
 &\quad \times \mathbb{P}(\tau_{23} < I_2 \mid I_1, I_2, I_3) \mathbb{P}(\tau_{32} \geq I_3 \mid I_1, I_2, I_3) \\
 &= \left(\int_0^{I_1} \beta e^{-\beta \tau_{12}} d\tau_{12} \right) \left(\int_{I_1}^{\infty} \beta e^{-\beta \tau_{13}} d\tau_{13} \right) \left(\int_0^{I_2} \beta e^{-\beta \tau_{23}} d\tau_{23} \right) \left(\int_{I_3}^{\infty} \beta e^{-\beta \tau_{32}} d\tau_{32} \right)
 \end{aligned}$$

and, if f_I is the probability density function of the random variable I ,

$$\mathbb{P}(\mathcal{A}) = \int_0^\infty \int_0^\infty \int_0^\infty \mathbb{P}(\mathcal{A} \mid I_1, I_2, I_3) f_I(I_1) f_I(I_2) f_I(I_3) dI_1 dI_2 dI_3$$

If we define $g' = (\{1\}, \{2\}, \{3\})$, $g'' = (\{1\}, \{2, 3\}, \{\})$, $g''' = (\{1\}, \{3\}, \{2\})$, the probabilities of observing generations g' , g'' and g''' in the real-time approach are:

$$\begin{aligned} \mathbb{P}(g') &= \mathbb{P}(\mathcal{A} \cup \mathcal{B} \cup \mathcal{C} \cup \mathcal{D}) \\ \mathbb{P}(g'') &= \mathbb{P}(\mathcal{E} \cup \mathcal{F} \cup \mathcal{G} \cup \mathcal{H}) \\ \mathbb{P}(g''') &= \mathbb{P}(\mathcal{I} \cup \mathcal{J} \cup \mathcal{K} \cup \mathcal{L}) \end{aligned}$$

On the other hand, individuals with the same rank are grouped in g' , g'' and g''' in the rank-based approach with probabilities:

$$\begin{aligned} \mathbb{P}(g') &= \mathbb{P}(\mathcal{A} \cup \mathcal{B}) \\ \mathbb{P}(g'') &= \mathbb{P}(\mathcal{C} \cup \mathcal{D} \cup \mathcal{E} \cup \mathcal{F} \cup \mathcal{G} \cup \mathcal{H} \cup \mathcal{K} \cup \mathcal{L}) \\ \mathbb{P}(g''') &= \mathbb{P}(\mathcal{I} \cup \mathcal{J}). \end{aligned}$$

The two approaches give different results because of where the contribution of events $\mathcal{C}$, $\mathcal{D}$, $\mathcal{K}$ and $\mathcal{L}$ is counted. Focus the attention on the infectious contact network $\mathcal{C}$: conditioning on I_1, I_2, I_3 ,

$$\begin{aligned} \mathbb{P}(\mathcal{C} \mid I_1, I_2, I_3) &= \mathbb{P}(\tau_{12} < I_1, \tau_{13} < I_1, \tau_{23} < I_2, \tau_{32} \geq I_3, \\ &\quad \text{any } \tau_{21}, \text{ any } \tau_{31} \text{ and } \tau_{12} + \tau_{23} < \tau_{13} \mid I_1, I_2, I_3). \end{aligned}$$

This probability is difficult to compute because of the dependence among the times τ_{ij} introduced through the last condition $\tau_{12} + \tau_{23} < \tau_{13}$. Note however that the infectious contact network $\mathcal{F}$ has probability

$$\begin{aligned} \mathbb{P}(\mathcal{F} \mid I_1, I_2, I_3) &= \mathbb{P}(\tau_{12} < I_1, \tau_{13} < I_1, \tau_{23} < I_2, \tau_{32} \geq I_3, \\ &\quad \text{any } \tau_{21}, \text{ any } \tau_{31} \text{ and } \tau_{12} + \tau_{23} \geq \tau_{13} \mid I_1, I_2, I_3) \end{aligned}$$

and therefore, by the law of total probabilities,

$$\mathbb{P}(\mathcal{C} \cup \mathcal{F} \mid I_1, I_2, I_3) = \mathbb{P}(\tau_{12} < I_1, \tau_{13} < I_1, \tau_{23} < I_2, \tau_{32} \geq I_3, \\ \text{any } \tau_{21} \text{ and any } \tau_{31} \mid I_1, I_2, I_3)$$

which is easier to compute.

Analogously, the probabilities of $\mathcal{G} \cup \mathcal{K}$ and of $\mathcal{D} \cup \mathcal{H} \cup \mathcal{L}$ are easier to compute than the probabilities of the single infectious contact networks. However, the final size is the same regardless on how it is studied.

3.6 Validity of the result

The general framework presented in this chapter allows to analyse Ludwig's result in depth, highlighting what is preserved and what is not when switching from the real-time to the rank-based process. Thanks to this approach, it is easily seen on which assumptions the result is based and hence for which models it does not hold.

First of all, recall that the result is specified for models of diseases that lead to permanent immunity after recovery and therefore cases of waning or no immunity are not considered in this context. However, there are some exceptions also in the case of permanent immunity.

Strictly speaking, Ludwig's rank-based construction, as described in his paper, can always be performed retrospectively after an epidemic has finished, provided that the infectious contact process is somehow known: it is sufficient to consider all the individuals that have been ultimately infected and attribute them the proper rank value. It is clear that, since by definition the rank is assigned only to individuals that have been infected, the final sizes of the real-time and the rank-based processes are the same; furthermore, since this attribution of the rank values can be done whatever the outcome of the epidemic, it is also clear that the probabilities of occurrence of each particular final size have to be the same for both processes. However, this observation is not useful if the final size distribution

cannot be computed easily in either of the two processes. Since the real-time process is rarely analytically tractable, the hope lies on the tractability of the rank-based process.

It has been highlighted in Section 3.4.1 how this is the case when individuals' behaviour can be studied *a priori*, i.e. before the epidemic is observed. It has also been pointed out that the key idea is to switch from a calendar time perspective to a time-since-infection one and in particular that this operation can be performed under the assumption that contacts between individuals occur at the points of a homogeneous Poisson process or, alternatively, that the non-homogeneity of the Poisson process is only due to factors that influence it as a function of time since infection.

If the contact rate changes as a function of calendar time the simplification is not possible: this may be the case when social external factors affecting individuals' behaviour are taken into account. Some examples include: individuals mixing during the day but not during the night; social patterns changing between weekdays and weekends; individuals spending some time with some individuals (e.g. at home) and then some other time with others (e.g. outside or at work); the beginning of the autumn season with children going to school and mixing more closely; the implementation of a control policy increasing social distancing; and so on.

If individuals' behaviour is modelled only after infection and the remaining contacts are not explicitly modelled (as it is done in most cases), the possibility of referring to a time-since-infection perspective is not derived, but directly assumed. Even in this case, however, the real-time infection process and the rank-based contact process are not equivalent. First, the two processes differ in the absolute time of infection of an individual (e.g. in Figure 3.1, the real-time infection process sees the infection of 3 before what would occur by considering the rank-based one): therefore, the equivalence of the final size distribution for the two is again compromised when the infectivity of individuals is allowed to depend on calendar time, in addition to the time since infection.

Second, the two processes differ in who infects whom (what is referred as the *book-keeping* in Diekmann & Heesterbeek, 2000) and therefore Ludwig's result cannot hold in general if there is a correlation between infector and infectee, i.e. if the infectivity of an in-

dividual is affected by some characteristics of the individual that infected him or her. This is the case, for example, when two or more strains of the same pathogen circulate in the same population and lead to different levels of infectivity and some sort of cross-immunity: in this case, the identity of the infector determines the type of strain (and thus the future infectivity) of the infectee. See Kendall & Saunders (1983); Svensson & Scalia-Tomba (2001); Ball & Britton (2007) for some theoretical work on this “infecter-dependent”-type model.

Note that these objections are not minor limitations to the applicability of the result to practical situations: models in infectious disease epidemiology often require external forcing to reproduce observed patterns and prediction of the likely impact of different control policies requires modelling sudden changes in the external conditions. Furthermore, some degree of dose-dependence in transmission or evolutionary changes in the pathogen’s strain structure under selection pressure may sometimes play an important role when considering specific diseases.

However, although in many practical contexts the assumptions of independence on calendar time and the uncorrelated infectivity between individuals are violated, when the population is relatively small or the duration of the epidemic is relatively short, the models described in Section 3.4 might provide good approximations of a real-world infection process.

3.7 Computation of the final size distribution

Ludwig’s construction described above suggests that, provided some simplifying assumptions hold, the actual disease history of the infectives is irrelevant if only information about the final size distribution is required. The only relevant information required is the distribution of total infectiousness of individuals. Therefore, the standard *SIR* model for the disease history (Section 2.1.4) is sufficiently general to study the final size distribution of an epidemic, independently of the presence or absence of a latent period or the actual shape

of the infectivity profile. This is particularly important, as a computational technique for computing the full final size distribution has been developed in Ball (1986) (see also Andersson & Britton, 2000). Any distribution for the length of the infectious period in the standard *SIR* model (i.e. any distribution for the total infectivity of individuals) can be considered, provided that a moment generating function of such a distribution exists and that “pathological” cases (e.g. latent periods with an infinite average) are excluded.

The theorem on which the technique is based is reported here. Although the notation is different, a proof can be found in Ball, 1986 or in Andersson & Britton, 2000.

Theorem 3.7.1. *Consider a standard SIR epidemic model in a population of size n with a initial infectives. Define $u = n - a$ as the number of initial susceptibles and denote by I the random variable describing the duration of the infectious period. Furthermore, assume homogeneous mixing and frequency dependent contact rate with an infective-to-susceptible (one-to-one) contact rate $\lambda = \beta/(n - 1)$. Let P_k^u be the probability that the final size of the epidemic (excluding the initial cases) is exactly k , $0 \leq k \leq u$. Then:*

$$\sum_{k=0}^l \binom{u-k}{l-k} P_k^u / [L_I(\lambda(u-l))]^{k+a} = \binom{u}{l}, \quad 0 \leq l \leq u, \quad (3.7.1)$$

where $L_I(s) = \mathbb{E}[e^{-sI}]$ represents the Laplace transform of I , i.e. the moment generating function of $-I$.

As Equations (3.7.1) form a linear triangular system, very efficient computational techniques exist to find its numerical solution. However, the practical usefulness of the result is limited to population sizes around 50 or 70, due to rounding errors in calculations of the binomial coefficients involved. Luckily, asymptotic results for the final size in large populations (e.g. Andersson & Britton, 2000) appear to be quite accurate for population sizes as small as hundreds of individuals, thus covering almost the whole range of possible sizes of interest.

3.8 Summary and further comments

In addition to drawing attention to Ludwig's result, this chapter offered a new derivation of it from a different perspective and a new convenient graphical way to summarise enough information of the contact process to allow the derivation of both the real-time infection tree and to assign ranks to individuals according to Ludwig's construction (see Figures 3.1 and 3.2). It has been pointed out how some ambiguities in terminology can represent a source of confusion in the literature. Finally, the insights that can be gained from Ludwig's result have been thoroughly commented, in particular highlighting the assumptions under which it does not hold.

In summary, Ludwig's rank-based construction allows the computation of final size probabilities by ignoring the dynamics of epidemic spread and such a simplification is possible only when the infectious behaviour of each individual, in case of infection, can be provided *a priori*, i.e. before the epidemic starts. This is not possible when the infectious life of an individual depends on the time of infection or the identity of the infector. These requirements usually translate in the assumptions of a static surrounding environment and of an uncorrelated infectivity of different infectives. Many real-life examples when these assumptions are violated have been provided.

The general approach proposed in this chapter allows the investigation of more complex situations than those originally considered by Ludwig, in particular that of Theorem 3.1.1 described in Section 3.1.

In fact, since the final size distribution can be studied by considering the full infectious contact networks, which do not depend on the absolute time at which events occur, even in the case of multiple infections of the same household from outside at different times, the probability distribution of the sets of individuals ultimately infected or of the final size distribution can still be computed combining the probabilities of various networks occurring with those of various individuals being infected from outside and taking the union of the outgoing connected components of these individuals. Furthermore, the assumption of constant probability π of infection from outside is also justified in light of the possibility

to ignore the temporal dynamics of the epidemic in the total population. In particular, π can be calculated from the total force of infection due to all the individuals ultimately infected in the population, independently of when they are infected.

This argument increases even further the relevance of Ludwig's result, which was already clear from the fact that it allows the results concerning the final size of the Reed-Frost model and its generalizations (multitype, randomised, collective Reed-Frost models, Reed-Frost models with infection from outside: Longini & Koopman, 1982; Picard & Lefèvre, 1990; Neal, 2006) to be extended to many other models with more realistic temporal dynamics. It is therefore a fundamental result of robustness with direct and indirect consequences in the field of mathematical models for infectious disease epidemiology.

Chapter 4

Households-workplaces model

4.1 Introduction

Although based on a previously defined model (Ball & Neal, 2002), this chapter contains the core of the original analytical work of my research and is summarised in Pellis *et al.* (2009). The following chapters expand from it and contain both analytical and computational original work. The main purpose of adding the presence of workplaces into a households model is to investigate their contribution in shaping the spread of the infection, in particular by disentangling the workplace transmission from other forms of between-household transmission. From a more practical perspective, such a model can also be used to investigate the effect of closing schools as a form of non-pharmaceutical intervention aimed at increasing social distancing.

School closure may appear to be a drastic control measure, because of its economical implications, but it is clearly administratively viable, socially acceptable and has been implemented in the past: for example, during the 1918 influenza pandemic (e.g. Markel *et al.*, 2007), or in 2003 in Singapore (Ministry of Education and Ministry of Health, 2003) and in Hong Kong (Chan-Yeung & Yu, 2003).

In general, however, it has always been a controversial control policy, mainly because of the difficulties in assessing its efficacy, which may be too limited for the associated

economical costs (Cauchemez *et al.*, 2008). Such difficulties arise mainly from the limited data and the unclear and not easily quantifiable change of behaviour subsequent to the control policy implementation.

Actually, the social structure modelled in this chapter consists only of households and workplaces and does not distinguish between adults and children: in this sense, schools are not explicitly described. However, closing schools may in general result in some adults staying at home to take care of their children. Therefore investigating school closure by analysing the effects of closing all workplaces, a policy that can be studied in the model, could arguably be not too rough an approximation (some refinements are considered in Section 7.3).

4.2 Model definition

4.2.1 Basic model assumptions

The model is a straightforward generalisation of the households model described in Section 2.4. The generic model assumptions are as follows: individuals can be infected and then recover; immunity is lifelong; no disease-induced death is assumed, in the sense that each recovered individual still mixes with the others, although immune to the infection; no births and deaths are assumed, as the interest is focused on single epidemics with a shorter time-scale than that of demographic changes; the population is closed, in particular there is no immigration or emigration. Except when stated otherwise, in the remainder of the chapter no particular model will be specified for the infection process, i.e. for the disease history of an infected individual and for the dynamics of the interaction with a susceptible one, since the results presented have a general form which does not depend on such level of detail. In particular, the presence or absence of a latent period (i.e. when individuals are infected but not yet infectious) will not affect the following discussion and, more generally, the concept of time will not be modelled explicitly.

However, two assumptions concerning the spread of the infection are required to be

satisfied:

- a) The disease history of an infective does not depend on the identity of the individual that caused the infection: in other words, it is assumed that there is no correlation in biological infectiousness between infector and infectee (Section 3.6 here, or and Diekmann & Heesterbeek, 2000, Sections 2.2 and 5.3).
- b) External conditions, like seasonal variations or other social changes, that would affect the mixing patterns in the population or the interaction between individuals and the pathogen, remain constant throughout the epidemic.

Such simplifying assumptions appear frequently (and often implicitly) in epidemiological modelling and are required here as sufficient conditions for the mathematical results to hold. More specifically, they represent exactly those assumptions that allow Ludwig's rank-based construction, as described in Chapter 3, to be used. Note however that they may be relaxed: assumption b) is mainly required since the quantities studied in the paper are treated as constant, although in a realistic context they can vary through time. A thorough comment on these assumptions is reported in Section 4.9 after the model has been discussed.

4.2.2 Assumptions concerning transmission

More detailed model assumptions are set as follows.

Let the population be partitioned into m_H groups, called households, each of size n_H ; the population is then partitioned again into m_W groups, called workplaces, each of size n_W . The concept of workplace is intended to be in a broad sense, referring to groups among which people spend most of their time while not at home (e.g. effective workplaces, schools, other types of social groups and so on). The two partitions overlap, such that each individual belongs to exactly one household and one workplace. The total population size is N . Of course, these parameters are constrained by the fact that N , m_H , m_W , n_H , n_W are positive integers and $N = m_H n_H = m_W n_W$ (see Figure 4.1).

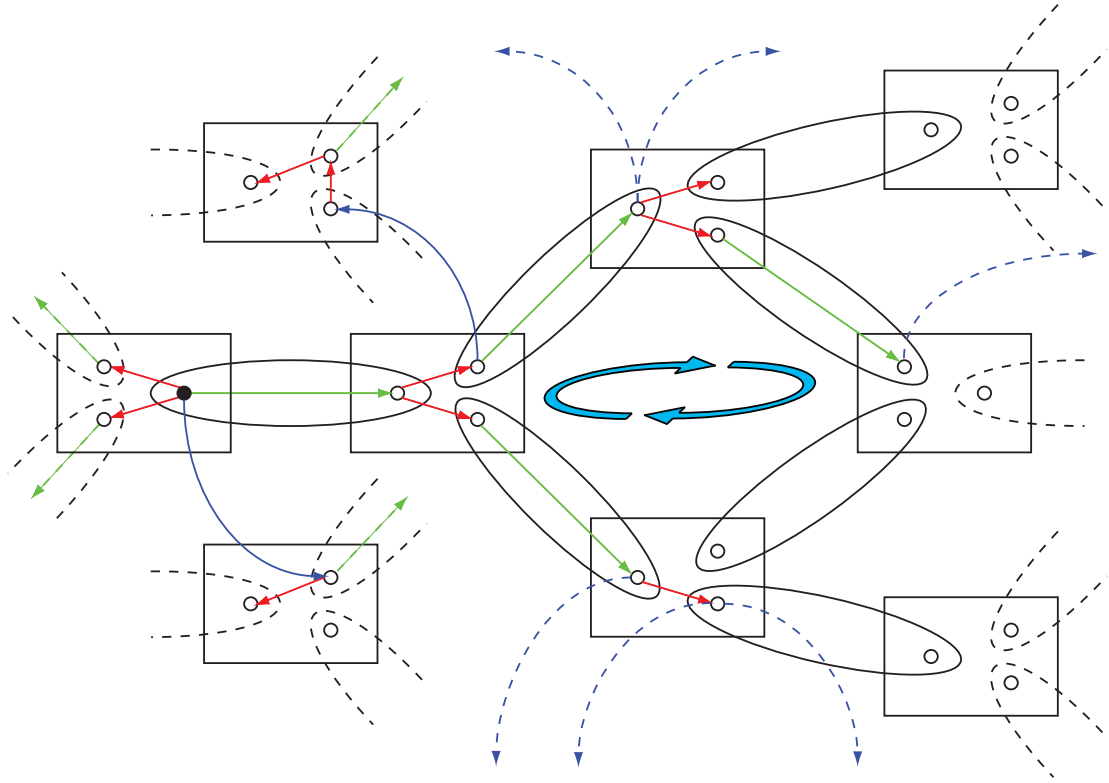


Figure 4.1: Schematic representation of an epidemic spreading in the households-workplaces model, with households (rectangles) of size $n_H = 3$ and workplaces (ellipses) of size $n_W = 2$. Arrows represent household (red), workplace (green) and global (blue) infections. The black dot represents the initial infective and the bold light blue arrow indicates a *cycle* (see text).

Each infectious individual has probabilities p_H and p_W to infect each other susceptible individual in the same household and workplace, respectively, and a probability p_G , usually much smaller, to infect each other susceptible individual in the population (including individuals from the same household or workplace). These are defined as *household*, *workplace* and *global infections*, respectively. Analogously, a *household*, *workplace* or *global infectious contact* (or simply *contact*) is defined as any event between two individuals which results in a household, workplace or global infection whenever the first individual is infectious and the second is susceptible. The adjective *local* refers to either a household or a workplace, and a *local structure* is either a household or a workplace. A local structure is said to be *susceptible* if all its members are susceptible and *infectious* when at least one of its

members is infectious. The epidemic process within an infected local structure, which is driven mainly (but not necessarily only) by local infections, is called a *local epidemic*. A local epidemic ends in the first moment in which there are no infectious members left in the local structure.

Following Ball *et al.* (1997) (see also Section 2.4.4), the first infected person in a local structure is defined to be the *primary case* and any other initial susceptible who is ultimately infected in the local epidemic is defined to be a *secondary case*. Therefore, in the remainder of this chapter, all the possible chains of transmission actually reaching each secondary case are thought of as direct infections from the primary case. As a consequence, secondary cases do not infect other individuals in the same local structure.

Note that this definition of secondary case is different from the usual one and therefore the average number of secondary cases according to this new definition will in general be different from R_0 , except in some extreme cases (e.g. no local infection at all, or instantaneous infection of everybody in the local structure, as assumed in Becker & Dietz, 1995).

Assume the population is wholly susceptible and the epidemic starts when one individual is infected. It is not investigated how this individual has been infected.

4.2.3 Assumptions concerning the social structure

Ball & Neal (2002) first noted how the underlying *social structure*, i.e. the configuration of the two partitions (in households and workplaces) can affect the course of an epidemic. It is therefore useful to formally describe this social structure, and a bipartite network (see Newman, 2002) seems to be a convenient way to represent it: there are m_H vertices of one type, representing the households (we call them *h-vertices*), and m_W vertices of another type, representing the workplaces (called *w-vertices*). Each h-vertex is reached by n_H (undirected) edges connecting it only with w-vertices and each w-vertex is reached by n_W edges connecting it only with h-vertices. No edges connecting vertices of the same type exist, but multiple edges between an h-vertex and a w-vertex are allowed. The total

number of edges is N , each of them representing an individual in the population.

Two individuals are said to be *acquaintances* if they share the same local structure and a *cycle* is defined as a chain of acquaintances of alternating types (i.e. one in a household, the next in a workplace, the next one again in a household and so on) that connects an individual to himself or herself (see Figure 4.1). Note that, in the bipartite network framework, individuals are represented as edges and two individuals are acquaintances if the edges that represent them depart from the same node; therefore, each cycle corresponds to a loop in the network (a closed path, in the terminology of graph theory). Since the network is bipartite, a cycle consists always of an even number of acquaintances. For a precise formalisation of this idea, see Ball & Neal (2002).

The only hypothesis Ball & Neal (2002) assume (for technical reasons) at the beginning of their study is that there are no cycles of size 2 in their social structure, i.e. the presence of pairs of individuals that share both the same household and workplace. However, they then mainly focus on a sequence of social structures with increasing population sizes in which the size of the smallest cycle tends to infinity as the total population size tends to infinity. For the sake of simplicity, a less general model is adopted here: apart from the size of the local structures, the social structure is assumed to be random, i.e. it is assumed that each member of a fixed household (workplace) also belongs to a randomly chosen workplace (household). In other words, the choice is to consider only bipartite *random* networks (Newman, 2002).

Consider now the limit when m_H , m_W and N tend to infinity while the sizes n_H and n_W are held fixed and focus the attention on the *early phase* of the epidemic, i.e. the initial phase when only a finite number of individuals is infected. Because of the random network assumption, the probability that a finite chain of acquaintances reaches a previously selected individual tends to 0 as $N \rightarrow \infty$, as does the probability of occurrence of cycles of finite size (see Newman, 2002). Therefore, local contacts (and finite chains of local contacts) reach susceptible individuals with probability tending to 1. Analogously, the probability that global contacts reach previously uninfected households and workplaces

tends to 1 too. A consequence of this fact is that, in the limit of an infinite population and in the early phase of the epidemic, local epidemics spread almost surely via local contacts only and that it is possible to define as *immune* a local structure in which a local epidemic has finished, since although some individuals may still be susceptible, the local structure is almost never infected again from outside (during the early phase).

Note that a bipartite random network representation is well defined only under these assumptions, because a node representing a local structure (and not a single individual), is infected only once. The number of infected edges exiting this node is randomly selected among all the outgoing edges following the final size distribution of the local epidemic just finished. A (almost impossible) re-infection of an already infected local structure would cause the same node to behave differently from the first time it was infected, with the new behaviour being dependent of how many susceptibles escaped the previous local epidemic.

4.3 Threshold parameters

4.3.1 Different types of infectives

Because of the definition of a secondary case considered above, initial susceptibles that are infected in a household epidemic cannot infect anybody else in their household and, since each of them belongs to a previously uninfected workplace with probability tending to 1 as N goes to infinity, it is convenient to call them *workplace primary cases* (w.p. cases). Analogously, primary cases in a household who were infected in their workplace are called *household primary cases* (h.p. cases). Individuals infected through global contacts are called *double primary cases* (d.p. cases), since they can trigger both household and workplace epidemics. Note that a w.p. case is also a household secondary case, and a h.p. case is also a workplace secondary case. Since finite loops appear with probability tending to 0 as $N \rightarrow \infty$, secondary cases both in a household and in a workplace can be neglected.

4.3.2 Clump reproduction number

Denote by μ_H (resp. μ_W) the average number of secondary cases in a household (resp. workplace), i.e. the average number of initial susceptibles who are ultimately infected in a local epidemic started by one single infective. R_g is defined as the average number of global infections made by a single infectious individual throughout his or her infectious period in a totally susceptible (infinite) population. The particular choice of avoiding any assumptions about the individual-to-individual infection process means that it is not investigated how to express (or compute) μ_H , μ_W and R_g in terms of more basic parameters (see Section 3.7 for more details about the average final size in a small group).

Ball & Neal (2002) gave a condition under which chains of local epidemics stop almost surely after a finite number of steps. They defined a *clump* as the set of people infected through one of these chains of local epidemics and gave a formula for its average size (Equation 3.10 in Ball & Neal, 2002). The average number of clumps infected by one of these clumps throughout its infectious life is defined as R_* and, as can be shown exploiting Wald's identity for epidemics (see Section 4.9 or Ball, 1986), it has the form

$$R_* = R_g \times \text{average size of a clump.}$$

In the present framework, the double type branching process used in Ball & Neal (2002) to describe the chain of local epidemics is driven by the next generation matrix (see Section 2.3.2 or Diekmann & Heesterbeek, 2000)

$$K = \begin{pmatrix} 0 & \mu_W \\ \mu_H & 0 \end{pmatrix}$$

since a workplace primary (w.p.) case generates on average only μ_W households primary (h.p.) cases and a h.p. case generates only μ_H w.p. cases. The dominant eigenvalue of K is $\sqrt{\mu_H \mu_W}$ and therefore a clump size is almost surely finite if and only if $\mu_H \mu_W < 1$.

Theorem 4.3.1. *In this notation, when $\mu_H \mu_W < 1$, the average size of a clump is*

$$\frac{(1 + \mu_H)(1 + \mu_W)}{1 - \mu_H \mu_W}. \quad (4.3.1)$$

Proof. Consider an individual infected through a global infection (i.e. a d.p. case). This individual starts an epidemic both in the household and in the workplace. Focus the attention on the household epidemic. The average number of infected individuals is given by the d.p. case plus μ_H secondary cases. Each of these μ_H cases infects on average μ_W cases in their workplaces, giving a total average of $\mu_H\mu_W$ new h.p. cases, which in turn start new household epidemics. Note that the average of the product of the number of cases generated in these two steps is equal to the product of the averages since, as $N \rightarrow \infty$, short loops occur with probability tending to 0 as does the probability that a household and a workplace share more than one individual tends to 0 too, thus making the epidemics in the household and the workplace develop independently in the infinite population limit.

The average total number of cases generated in this way is:

$$1 + \mu_H + \mu_H\mu_W + \mu_H^2\mu_W + \dots = (1 + \mu_H) \sum_{i=0}^{\infty} (\mu_H\mu_W)^i = \frac{1 + \mu_H}{1 - \mu_H\mu_W}, \quad (4.3.2)$$

provided that $\mu_H\mu_W < 1$.

The initial case also starts an epidemic in the workplace, giving rise to μ_W new household primary cases (h.p. cases), each of which generates a chain with total number of cases given again by (4.3.2). The total number of cases in the clump is then given by $(1 + \mu_W)$ of these chain, i.e. Equation (4.3.1). $\square$

Therefore R_* is expressed analytically by the formula

$$R_* = R_g \frac{(1 + \mu_H)(1 + \mu_W)}{1 - \mu_H\mu_W}. \quad (4.3.3)$$

Note that the average size of a clump tends to infinity as $\mu_H\mu_W$ tends to 1 from below and therefore R_* , although not formally defined, can be considered to be infinite when $\mu_H\mu_W \geq 1$. The reason is that in this case the epidemic is able to sustain itself based on chains of household and workplace transmission alone. Furthermore, note that the formula for R_* is symmetric in μ_H and μ_W , as it should be since households and workplaces play the same role in the process.

A comment about notation is due to avoid confusing R_* in this households-workplaces model with R_* in the households model (Section 2.4.5). In both models R_* should be

thought of as the reproduction number for the structure at the “highest possible level” that appears in the model, i.e. the structure that captures in itself all the complexity due to any form of local saturation. The remaining interaction between these “highest level” structures becomes then linear in the infinite population limit and such linearity is responsible for the factor R_g in both Equations (2.4.3) and (4.3.3) for R_* .

R_* satisfies the usual threshold property, i.e. large epidemics can occur only if $R_* > 1$. However, it has some limitations:

1. R_* can become infinite, because the average size of a clump can become infinite (Ball & Neal, 2002), and the condition for it to happen ($\mu_H \mu_W \geq 1$) is likely to occur in practice (it is sufficient to think about the large epidemics that may occur in schools);
2. The time required for a clump to form increases as $\mu_H \mu_W$ tends to 1 and becomes quickly comparable to the time of the whole epidemic;
3. A critical vaccination coverage obtained from this parameter will refer to clumps, but a clump is a random set of individuals and cannot be identified and vaccinated beforehand.

4.3.3 Household reproduction number

Other threshold parameters, referring to lower level structures, namely households, workplaces or individuals, may also be defined.

In particular, since a vaccine not only can, but is also likely to be delivered to entire households (Becker & Dietz, 1995; Becker & Utev, 1998; Wu *et al.*, 2006; Fraser, 2007), a reproduction number for households can be more relevant from a practical point of view. Analogously, it may also be more directly related to other forms of intervention targeted at households, like households quarantine and prophylaxis (Fraser, 2007). Therefore, define R_H as the average number of households infected by a “typical” household in a totally

susceptible population. Now, the infection can exit from a household following two routes: via a global infection or via a workplace infection, i.e. when an infected individual in the household triggers an epidemic in their workplace and all those ultimately infected are primary cases in new households.

Following Diekmann & Heesterbeek (2000), define R_{ij} , for $i, j \in L, G$, as the average number of households infected via a type- i infection by a household which has been infected through a type- j infection. The letters L and G refer to local workplace infections and global infections, respectively.

It has been chosen (in Section 4.2.2) to attribute all the final cases of a local epidemic to the individual that started it. In this way, when a household is infected locally ($j = L$), the household primary (h.p.) case does not play any active role in the workplace epidemic that caused his or her infection. Instead, this h.p. case starts a household epidemic and generates an average of μ_H secondary cases. Each of them belongs to a different workplace and starts a workplace epidemic with average final size μ_W , causing a total average of $\mu_H \mu_W$ new h.p. cases, each one in a different household. Therefore, $R_{LL} = \mu_H \mu_W$ (see Figure 4.2). Furthermore, in the household under consideration, each infective (including the primary case) also generates R_g new infectives outside, and each of these new infectives is a primary case in a new household (actually, is also a d.p. case). Therefore $R_{GL} = R_g (1 + \mu_H)$ (see Figure 4.2).

The same arguments apply when the household is infected globally ($j = G$), except from the fact that the h.p. case is a d.p. case and can now trigger an epidemic in his or her workplace as well as in the household. Therefore $R_{LG} = \mu_W (1 + \mu_H)$ and $R_{GG} = R_g (1 + \mu_H)$.

This construction can be formalized by exploiting generating functions, but this is not done here: see Ball & Neal (2002) for the same procedure in a slightly different context. The key idea is again that the average of the product of the random number of cases in subsequent local epidemics is equal to the product of the averages of each single epidemic because there is independence between them under the bipartite random network

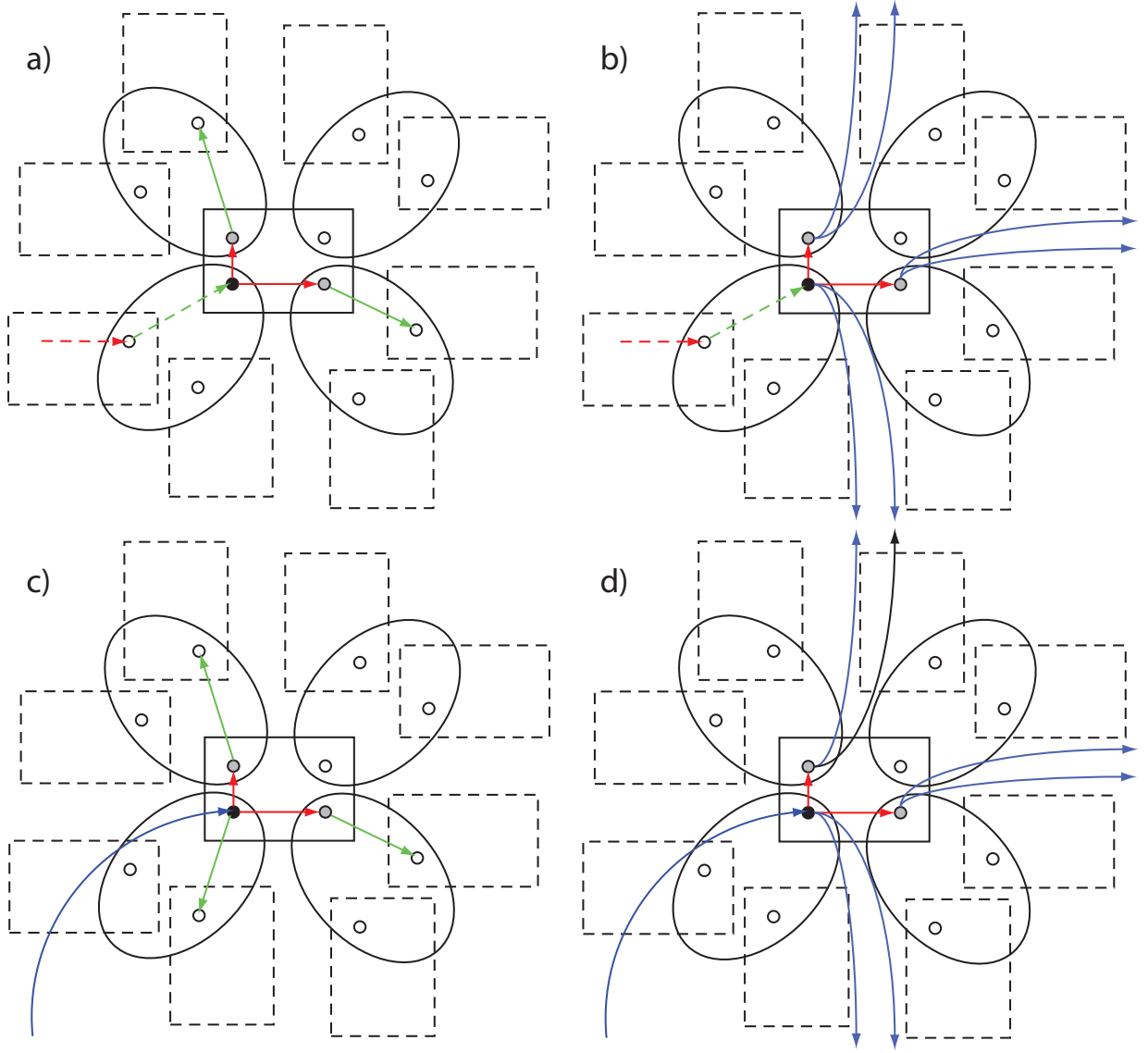


Figure 4.2: Graphical description of the quantities R_{GG} , R_{GL} , R_{LG} and R_{LL} (in order from a) to d)), used in the main text to construct the household-to-household next generation matrix, from which the threshold parameter R_H is then computed. In this case, $R_g = 2$, $\mu_H = 2$ and $\mu_W = 1$. Households and workplaces are represented by rectangles and ellipses, respectively, and arrows represent household (red), workplace (green) and global (blue) infections. The black dot represents the household primary case and the grey dots all the other infected cases in the household (attributed to, although not necessarily caused directly by, the primary case).

assumption and the infinite population limit.

The proper way of averaging the two types of household infections is to consider the next generation matrix (see Section 2.3.2 or Diekmann & Heesterbeek, 2000):

$$K_H = \begin{pmatrix} R_{LL} & R_{LG} \\ R_{GL} & R_{GG} \end{pmatrix}$$

A reproduction number for households is then given by its dominant eigenvalue,

$$R_H = \frac{R_g(1 + \mu_H) + \mu_H\mu_W}{2} + \frac{1}{2}\sqrt{[R_g(1 + \mu_H) + \mu_H\mu_W]^2 + 4R_g\mu_W(1 + \mu_H)}. \quad (4.3.4)$$

Note that, R_H is not linear in R_g , because of the non-linearity induced by the infection process in the workplaces. Furthermore, it is not symmetric in μ_H and μ_W : in fact, the attention has been focused on households, and workplaces have been used only as a means to reach other households.

4.3.4 Other reproduction numbers

Of course, the roles of households and workplaces could be swapped, thus providing a threshold parameter for workplaces, R_W , of the form

$$R_W = \frac{R_g(1 + \mu_W) + \mu_H\mu_W}{2} + \frac{1}{2}\sqrt{[R_g(1 + \mu_W) + \mu_H\mu_W]^2 + 4R_g\mu_H(1 + \mu_W)}. \quad (4.3.5)$$

For the sake of completeness, a threshold parameter R_I for individuals can also be defined with similar arguments: however, it is not worth going into details since it is more complex and less useful. The main difference is that now there are three types of individuals: d.p. cases, w.p. cases and h.p. cases and the next generation matrix is now:

$$K_I = \begin{pmatrix} R_g & R_g & R_g \\ \mu_H & 0 & \mu_H \\ \mu_W & \mu_W & 0 \end{pmatrix}.$$

4.3.5 Comparison between the threshold parameters

Among all these threshold parameters, the attention is focused primarily on R_* and R_H .

R_H satisfies the usual threshold property, i.e. large epidemics can occur only if $R_H > 1$, and is related to R_* according to the following

Theorem 4.3.2. *For all non-negative values of R_g, μ_H and μ_W ,*

- $R_* < 1 \Leftrightarrow R_H < 1$
- $R_* = 1 \Leftrightarrow R_H = 1$
- $R_* > 1 \Leftrightarrow R_H > 1$,

i.e. R_ and R_H share the same threshold at 1 and are consistent in their threshold property.*

Proof. If $R_g = 0$ then $R_H = \mu_H \mu_W$ and, when $R_H < 1$, $R_* = 0$, while when $R_H \geq 1$, R_* is ill-defined. Therefore it can be assumed that $R_g > 0$. The condition $R_* = 1$ is obtained when

$$R_g = \frac{1 - \mu_H \mu_W}{(1 + \mu_H)(1 + \mu_W)},$$

which, since $R_g > 0$, it implies $\mu_H \mu_W < 1$. Substituting R_g in the definition of R_H , the expression under square root becomes

$$\left[2 - \frac{1 + \mu_H^2 \mu_W}{1 + \mu_H} \right]^2$$

and from $\mu_H \mu_W < 1$ it follows $R_H = 1$.

When $R_* < 1$,

$$R_g < \frac{1 - \mu_H \mu_W}{(1 + \mu_H)(1 + \mu_W)},$$

which, since $R_g > 0$, again implies that $\mu_H \mu_W < 1$. The fact that $R_H < 1$ then follows from the fact that the square root is a monotonically increasing function. $\square$

Apart from sharing the same threshold at 1 (Theorem 4.3.2), R_* and R_H are non-trivially related. On the one hand, when R_* is infinite, no relationship with R_H can in

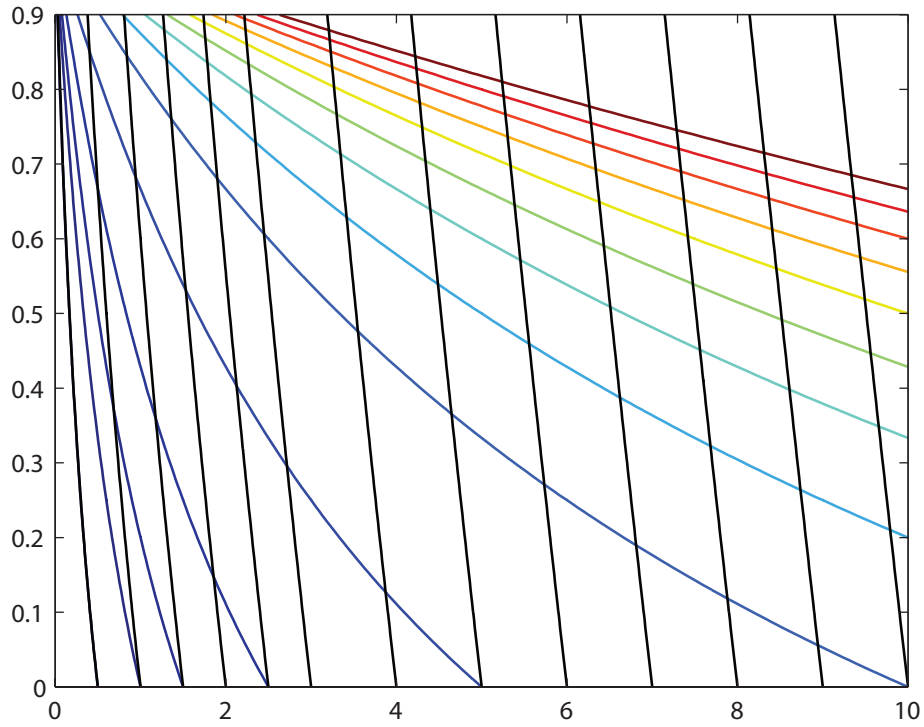


Figure 4.3: Level curves for the two threshold parameters R_* and R_H of Equations (4.3.3) and (4.3.4) (see text), plotted on the (R_g, μ_W) -plane and with $\mu_H = 1$ fixed. Black lines refer to R_H , coloured lines refer to R_* . Both parameters increase by moving from the lower-left region to the upper-right one and they share the leftmost level curve at 1.

fact be maintained; on the other hand, an increase (resp. decrease) in one or more basic parameters results in an increase (resp. decrease) of both R_* and R_H , but increasing one basic parameter while decreasing another may in general result in the two reproduction numbers changing in different directions. This is because R_* refers to clumps and R_H to households, and so it may well happen that fewer clumps are generated by a single clump, but their size is larger because more households are infected by a single household through workplace infections within the clump. As a numerical example (see Figure 4.3), by fixing $\mu_H = 0.9$ and changing (R_g, μ_W) from $(5, 0.4)$ to $(4.5, 0.5)$, R_* increases from 23.33 to 27 while R_H decreases from 10.77 to 9.95.

R_H overcomes some of the limitations of R_* presented above:

1. R_H can never become infinite;
2. The time a household requires to infect other households is limited and usually much shorter than the time of the whole epidemic;
3. R_H can be related to a critical vaccination coverage p_{crit}^H for households.

4.3.6 Similarities with other results

A construction similar to that leading to R_H has been adopted independently in Ball & Neal (2008), although with rather different motivation and derivation. In Ball & Neal (2008), a random contact network with specified degree distribution describes local infections occurring between individuals (instead of households), which are superimposed to global infections. As the population is large and the local contact network is random, local clustering is negligible and therefore there is no need to redefine the concept of secondary cases (as done here in Section 4.2.2): for this reason, the threshold parameter obtained by Ball & Neal (2008) is actually the basic reproduction number R_0 . In the households-workplaces model, instead, the network representing local contacts consists of fully connected sub-graphs (households and workplaces), joint by single nodes, and local infections between households spread through workplace epidemics, where the problem of the non-linearity induced by the local saturation is bypassed by referring all subsequent cases in a workplace epidemic to the primary case.

4.4 Average final size

Although unrelated to the various reproduction numbers that can be defined for the model, analytical results are available in the literature for the average final size of a large epidemic (Ball & Neal, 2002). As for the case of the households model of Section 2.4, they are based

on the concept of local susceptibility set. Recall from Section 2.4.6 that the *local susceptibility set* of an initially susceptible individual i is a random set consisting of individual i plus all other individuals that, if infected through a global contact, would infect individual i through a chain of local contacts (i.e. through household and workplace epidemics). The probability of i escaping infection is therefore equal to the probability of all individuals in i 's local susceptibility set escaping global infections.

Because individuals are all identical and the infection process is symmetric, the only relevant piece of information concerning the local susceptibility set of i is its size S_i and, $\forall i$, S_i is independently and identically distributed according to a random variable S .

With the same arguments used in Section 2.4.6 for the households model, the average final size of the households-workplaces model is given by the largest solution z of the equation:

$$1 - z = f_S(e^{-R_g z}),$$

where $f_S(s) = \mathbb{E}[s^S]$ represents the probability generating function of the size of the local susceptibility set (see also Ball & Neal, 2002).

As in the case of the households model, the distribution of S is the same distribution of the number of cases that an infected individual would infect through a chain of local infections, if the rest of the population were fully susceptible, and is therefore given by the size C of a clump. Its average is given in Theorem 4.3.1. Note, however, that the result holds only when the average clump size is finite, i.e. when $\mu_H \mu_W < 1$.

Results about the finer structure of the final size, describing how ultimately infected individual are distributed in households and workplaces, are more complex, and are not discussed here; refer to Ball & Neal (2002) for further details.

4.5 Vaccination

Unfortunately the relationship between R_H and p_{crit}^H is not as neat as the usual ones (see Equations (2.2.1) and (2.4.2)), because of the non-trivial dependence of the final size of a

local epidemic in a workplace on the initial number of susceptibles available. However, an important relationship can still be derived, although based on an inequality and therefore weaker than the usual one. It is based on the following result.

Theorem 4.5.1. *Consider an epidemic in a closed and homogeneously mixing population of size n , started by a single initial infective. Let μ be the average final size (excluding the initial case) of an epidemic spreading in it. Let a random number $m < n - 1$ of initial susceptibles be vaccinated (totally immune) and define $P = m/(n - 1)$. Let $p = \mathbb{E}[P]$ and $q = 1 - p$, and define μ_V to be the average final size of an epidemic after vaccination (again, excluding the initial case). Then $\mu_V \leq q\mu$.*

Remark 4.5.2. Two possible special cases of Theorem 4.5.1 are obtained when:

- i) a fixed number m out of the $n - 1$ susceptibles is vaccinated;
- ii) each susceptible is vaccinated independently with probability p : in this case, $m \sim \text{Bin}(n - 1, p)$.

Assume, for the time being, that all households have the same size. Then Theorem 4.5.1 can be applied to workplaces since, when a fraction p (or a random fraction P with average p) of the households is selected and all the members are vaccinated effectively, our assumption of random connections between households and workplaces implies that individuals in workplaces are vaccinated independently of each other and with probability p (case ii) above). Therefore:

$$\mu_W^V \leq q \mu_W. \quad (4.5.1)$$

Since after the vaccination μ_H is not modified and R_g is reduced to $R_g^V = q R_g$, a direct consequence is the following

Corollary 4.5.3. *The fraction p_{crit}^H of households (all of the same size) that one needs to vaccinate in order to control the epidemic satisfies*

$$p_{crit}^H \leq 1 - \frac{1}{R_H}. \quad (4.5.2)$$

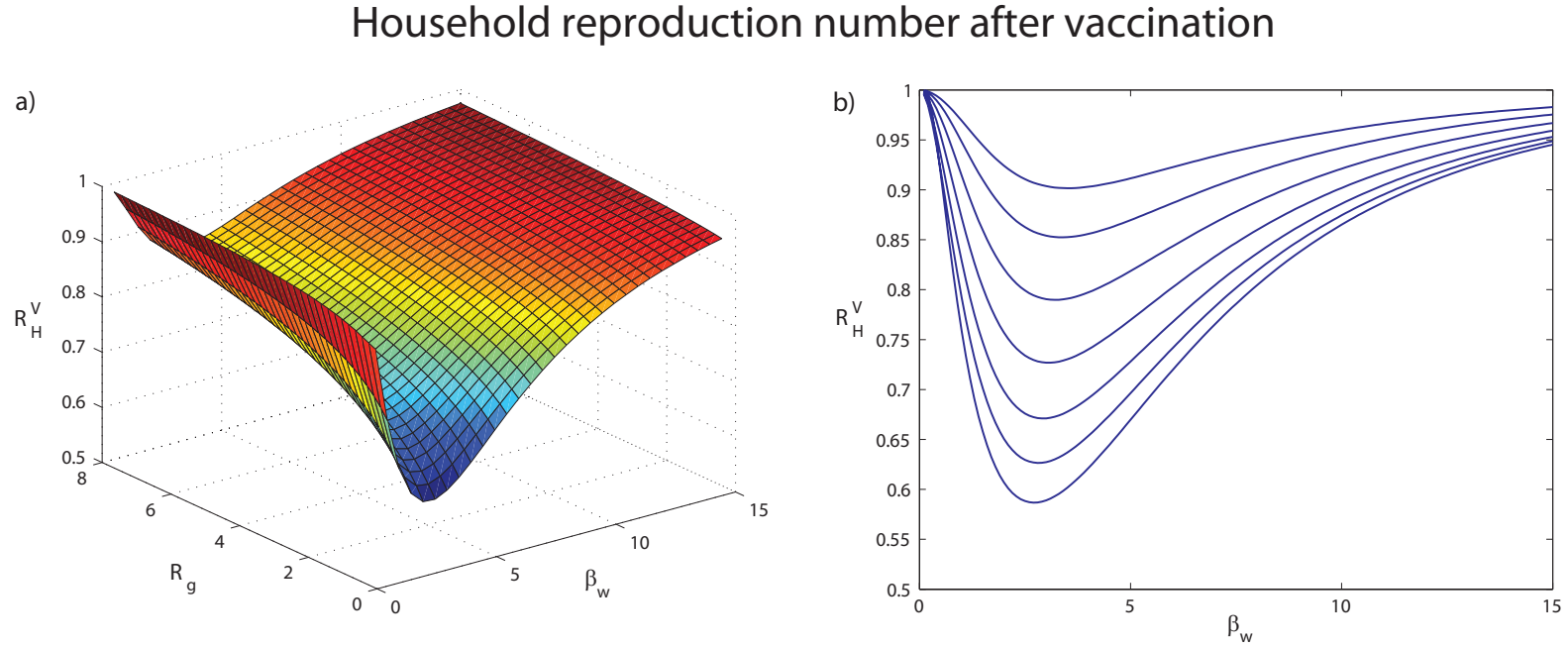


Figure 4.4: Values of R_H^V , i.e. the household-to-household reproduction number after the vaccination of a proportion $1 - 1/R_H$ of households (randomly selected). In *a*) R_H^V is plotted as a function of R_g and the within-workplace one-to-all infection rate β_w ; in *b*) R_H^V is plotted as a function of β_w for $R_g = 0.2, 0.5, 1, 2, 4, 8, 15$ (from bottom to top), in order to highlight that $R_H^V \leq 1$ always. Numerical values are obtained by fixing $\mu_H = 2$ (other values lead to qualitatively similar results) and assuming that all workplaces have size $n = 6$ and that the epidemic in a workplace spreads according to the standard *SIR* model defined in Section 2.1.4 (see also Ball, 1983b, 1986; Andersson & Britton, 2000) with infectious period of fixed length ι and individual-to-individual (one-to-one) infection rate $\lambda_w = \beta_w/(n - 1)$. The results are similar for other values of n , thus showing that it can be extended to a non-singular workplace size distribution (the value of μ_H already takes into account variable sizes for households). The model is equivalent to a Reed-Frost model with one-to-one escaping probability $q = \exp(-\lambda_w \iota)$ and to any other single-type model with time varying infectivity (see Section 3.2 or Ludwig, 1975; Pellis *et al.*, 2008). Other distributions for the length of the infectious periods in the standard *SIR* model lead to qualitatively similar results.

Corollary 4.5.3 states that the effect of vaccination on R_H is “more than linear”, i.e. vaccinating a proportion of households equal to $1 - 1/R_H$ will reduce the value of R_H at least to 1 if not below. This means that the value $1 - 1/R_H$ represents a *secure vaccination coverage* (in the terminology of Ball *et al.*, 2004). Figure 4.4 shows that vaccination of a proportion $1 - 1/R_H$ of households is just sufficient to reduce R_H below 1 when infectivity in workplaces is very low or very high. For intermediate values of such infectivity, less effort would be required, although an exact quantification can only be assessed numerically and depends on model assumptions about the individuals’ infectivity. Numerical values in Figure 4.4 are obtained using the standard *SIR* model defined in Section 2.1.4 (or in Ball, 1983b, 1986; Andersson & Britton, 2000).

4.6 School closure and other control policies

The other reason for adding the workplaces to the household structure, apart from making the model slightly more realistic and to investigate their contribution to the infection spread, is to study the effect of their closure on the spread of the epidemic. This can be simulated by setting $\mu_W = 0$ in both the formulae of R_* and R_H (Equations (4.3.3) and (4.3.4)). As expected, in both cases Equation (2.4.3) for R_* of the households model of Section 2.4.5 is recovered. Note however that the explicit presence of workplaces in the present model allows the effects of their closure to be disentangled from other forms of social distancing. In fact, should the individuals’ behaviour remain unchanged before and after the implementation of this control policy (as it is probably not the case, since individuals not at work spend their working time elsewhere), the value of the other parameters R_g and μ_H in the model would remain constant, while the parameter R_g in the households model would actually be reduced (in a non-trivial fashion) because it takes into account any form of between-household transmission, including that occurring at work.

Other forms of intervention are more difficult to study both qualitatively and quantitatively for various reasons.

On the one hand, from a qualitative point of view, the complex expression for R_H does not show a clear disentanglement between the effects on R_H of the change in any of the basic parameters. In this sense, the form of R_* might be more useful. However, a part from some particular measures of social distancing (e.g. wearing masks outside households and workplaces) which affects only R_g , most control policies, like households quarantine, isolation and also antiviral treatment of infected cases will reduce both R_g and μ_W (and sometimes also μ_H), with non-trivial consequences on both R_* and R_H . A particular case is that of prophylactic treatment of susceptible individuals in households with detected cases, since in the present setting it has the effect of reducing only μ_H : this is due to the assumptions about the large population and the particular connections between households and workplaces, that makes multiple introductions of the disease in the same household negligible in the initial phase of the epidemic.

On the other hand, quantitative estimation of the effects of most targeted interventions is also difficult in this setting mainly because effectiveness of these actions is highly dependent on the time delays between infection, detection of a case and following policy implementation. This requires a detailed specification of the dynamics of the individual-to-individual infection process, which has not been specifically modelled here.

4.7 Proof of Theorem 4.5.1

Intuitively, the factor q appearing in Theorem 4.5.1 derives from the fact that each initial susceptible, that would be infected in the absence of vaccination, is still susceptible after the vaccination with probability q . The “less than or equal to” comes from the additional (not easily quantifiable) contribution of stopping chains of infection, i.e. when a susceptible that would be infected in the absence of vaccination is still unvaccinated but escapes infection because his or her infector (or somebody else further back in the chain of infection) escaped infection because vaccinated.

Although this naive explanation is correct, confusion may arise when thinking that vaccination can be “wasted” on susceptibles that would have escaped infection anyway

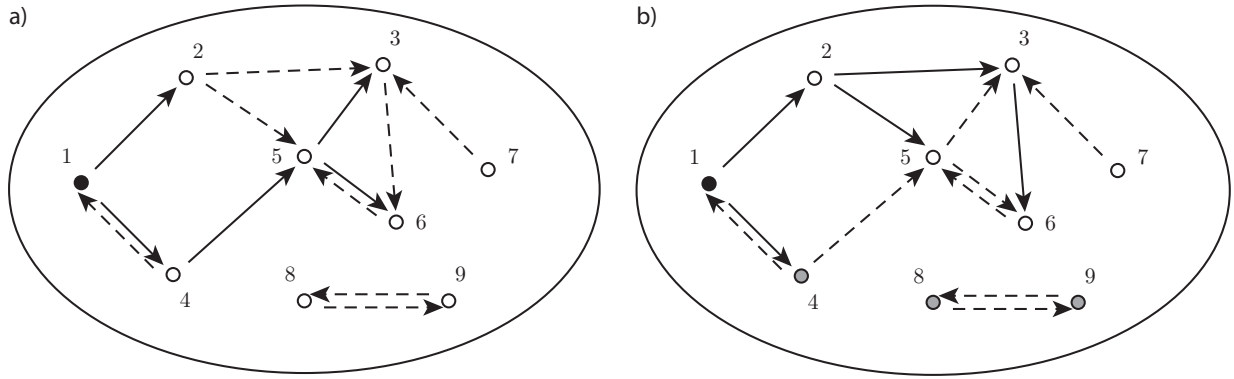


Figure 4.5: Schematic representation of an epidemic in a closed group, (a) before and (b) after vaccination. A dashed arrow between individuals i and j indicates that j would be infected during the infectious period of i , provided that i acquires infection and j is still susceptible when the potentially infectious contact occurs. In case of a real infection from i to j , the dashed arrow is substituted by a bold one. Infectious contacts from individuals that escape infection (or that are vaccinated) have no real meaning. The black dot represents the initial infective and the grey dots represent vaccinated individuals. The vaccination does not affect the graph of the infectious contacts, but it can interrupt branches in the tree of real infections (e.g. the case of individual 4). In this case, however, new branches may form (e.g. from individual 2) and individuals (e.g. 3, 5 or 6) that would have been protected via vaccination of their infectors can acquire infection from other individuals.

(as in the case of individuals 8 and 9 of Figure 4.5b) or from the fact that vaccination can stop some branches of real infections but may cause new others to appear (see Figure 4.5). Instead of working with trees of real infections, it is convenient to consider directed graphs representing infectious contacts between individuals (as done, for example, in Andersson & Britton, 2000, Chapter 7): the reason why this approach leads to the correct result is given in Ludwig (1975) or, more recently, in Pellis *et al.* (2008); see also Section 4.9 below. The final size is given by the number of nodes in the outgoing connected component of the initial case (see Figure 4.5). Consider a set of $n + 1$ nodes, labelled from 1 to $n + 1$, and assume without loss of generality that node 1 represents the initial infective. Consider now the set G of all the possible directed graphs g on the $n + 1$ nodes and the set V of all the possible vaccination schemes s on the n initial susceptibles $i = 2, \dots, n + 1$. Let Θ_i^g be the indicator function of individual i in an epidemic represented by the graph

g , defined as:

$$\Theta_i^g = \begin{cases} 1 & i \text{ is ultimately infected in } g \\ 0 & \text{otherwise} \end{cases}$$

and let $F^g = \sum_i \Theta_i^g$ be the final size of the epidemic. Moreover, let $F^{g,s}$ and $\Theta_i^{g,s}$ be the final size and the indicator function of i , respectively, after vaccination scheme s is applied. Note that $\mu_V = \mathbb{E}_{(G,V)}[F^{g,s}]$, where the expectation is taken over the joint distribution of the vaccination scheme $s \in V$ and the graph $g \in G$, but that, since we consider infectious contacts and not real infections, the graph d is independent from the vaccination scheme s and therefore $\mathbb{E}_{(G,V)}[F^{g,s}] = \mathbb{E}_G[\mathbb{E}_V[F^{g,s}]]$. Moreover, $\mu = \mathbb{E}_G[F^g]$. The aim is therefore to prove that

$$\mathbb{E}_G[\mathbb{E}_V[F^{g,v}]] \leq q\mathbb{E}_G[F^g].$$

Since there are no assumptions concerning the individual-to-individual infection process, it is not possible to specify the probability of occurrence of each $g \in G$. However, the result holds anyway because it can be proved that, for each *specific* $\bar{g} \in G$,

$$\mathbb{E}_V[F^{\bar{g},v}] \leq qF^{\bar{g}}. \quad (4.7.1)$$

To prove that Equation (4.7.1) holds, consider a specific graph $\bar{g}$. Then, the indicator function of the particular individual $\bar{i}$ after a specific vaccination scheme $\bar{v}$ is related to that in the absence of vaccination by the relationship:

$$\Theta_{\bar{i}}^{\bar{g},\bar{v}} = C_{\bar{i}}^{\bar{g},\bar{v}} \Theta_{\bar{i}}^{\bar{g}},$$

where $C_{\bar{i}}^{\bar{g},\bar{v}}$ is a factor defined as

$$C_{\bar{i}}^{\bar{g},\bar{v}} = \begin{cases} 1 & \text{if } \bar{i} \text{ is not vaccinated and not all chains reaching } \bar{i} \text{ are stopped} \\ 0 & \text{otherwise.} \end{cases}$$

Therefore:

$$\begin{aligned} \mathbb{P}(C_{\bar{i}}^{\bar{g},\bar{v}} = 1) &= \mathbb{P}(\bar{i} \text{ is not vaccinated in } \bar{s}) \\ &\quad \times \mathbb{P}(\text{not all the chains reaching } \bar{i} \text{ are stopped} \mid \bar{i} \text{ is not vaccinated in } \bar{s}) \\ &\leq 1 - P, \end{aligned}$$

where P is the proportion vaccinated in $\bar{s}$.

Finally,

$$F^{\bar{g}, \bar{s}} = \sum_i \Theta_i^{\bar{g}, \bar{s}} = \sum_i C_i^{\bar{g}, \bar{s}} \Theta_i^{\bar{g}}$$

and therefore

$$\mathbb{E}_V[F^{\bar{g}, \bar{s}}] = \sum_i \mathbb{E}_V[\Theta_i^{\bar{g}, \bar{s}}] = \sum_i \mathbb{E}_V[C_i^{\bar{g}, \bar{s}}] \Theta_i^{\bar{g}} \leq \sum_i \mathbb{E}_V[1 - P] \Theta_i^{\bar{g}} = \sum_i q \Theta_i^{\bar{g}} = q F^{\bar{g}}.$$

4.8 Variable household and workplace sizes

The model can be easily extended to variable household and workplace sizes. The only difference from the fixed size case is that the parameters μ_H and μ_W , the average final sizes of local epidemics, take into account these distributions. In particular, let h_n be the probability that a randomly selected household has size n . Then the probability that a randomly chosen individual belongs to a household of size n is

$$\pi_n = \frac{nh_n}{\sum_n nh_n}.$$

Note that the probabilities that the selected individual belongs to households of different sizes are weighted according to the number of individuals they contain and this takes into account the fact that larger structures are more likely to be infected than smaller ones; distributions like this are sometimes referred to as *size-biased distributions* (see Becker & Dietz, 1995; Ball *et al.*, 1997). Let $\mu_H^{(n)}$ be the average final size of an epidemic in a household of size n , starting with a single initial case (not counted) and $n-1$ susceptibles. Then μ_H is defined as

$$\mu_H = \sum_n \mu_H^{(n)} \pi_n \tag{4.8.1}$$

An analogous definition holds for μ_W .

By interpreting μ_H and μ_W as the average sizes of epidemics in the household and workplace of a randomly selected individual, all the previous results apply. In particular, the result for the critical vaccination coverage for households (Equation (4.5.2)) still holds

when households are selected independently of their size, i.e. according to the distribution h_n (see proof of Theorem 4.8.1 below).

Because larger households are more likely to be infected and spread the infection more than smaller ones, vaccination strategies targeted at larger households are more appealing. In particular one would expect that a strategy where individuals are selected at random and their households are vaccinated (households are selected according to π_n) performs better than the random selection of households; in turn, selecting only the largest households should achieve an even better result. This intuition is supported by the following result, which can be proved under the assumption of *density-dependent* contact rate (i.e. the within-household individual-to-individual infectivity β_h does not depend on the household size and therefore the average number of contacts increases with the household size), but which I conjecture holding also under the assumption of *frequency-dependent* contact rate (the average number of within-household contacts remains constant as the household size increases).

Theorem 4.8.1. *Assume a number of doses of vaccine sufficient to vaccinate a proportion P of the population is available. Assume that the vaccine is 100% effective and that it is delivered to entire households. Then the greatest reduction in R_H is achieved by vaccinating the largest households first.*

Proof. Denote by P_n the random fraction of households of size n that are vaccinated and let $i \in H_n$ mean that individual i belongs to a household of size n .

If the constraint is that households are vaccinated until a fraction P of the population of individuals is vaccinated, then the P_n 's are related to P in the following fashion:

$$\begin{aligned}
 P &= \text{fraction of the population vaccinated} \\
 &= \mathbb{P}(\text{a randomly selected individual } i \text{ is vaccinated}) \\
 &= \sum_n \mathbb{P}(i \text{ is vaccinated} \mid i \in H_n) \mathbb{P}(i \in H_n) \\
 &= \sum_n P_n \pi_n,
 \end{aligned}$$

The effect on R_H depends on the vaccination proportions $(P_1, P_2, \dots)$ through their effects on R_g , μ_W and μ_H . Vaccination of a proportion P of the population reduces R_g to $R_g^V = (1 - P)R_g$.

Focus the attention on a workplace that has been infected. Because of the bipartite random network assumption, each initial susceptible j has a probability of being vaccinated given by

$$\begin{aligned}\mathbb{P}(j \text{ is vaccinated}) &= \sum_n \mathbb{P}(j \text{ is vaccinated} \mid j \in H_n) \mathbb{P}(j \in H_n) \\ &= \sum_n P_n \pi_n = P,\end{aligned}$$

independently of the values of the P_n 's. Whatever the vaccination strategy, the conditions are those of Theorem 4.5.1 with the same $p = \mathbb{E}[P]$ and therefore the result achieved is the same.

Finally, let μ_H^V be the size-biased average final size (excluding the single primary case) of a household epidemic after vaccination. Then:

$$\mu_H^V = \sum_{n=1}^{+\infty} \mu_H^{(n)} \frac{(1 - P_n) \pi_n}{1 - P},$$

where the denominator $1 - P$ is necessary to normalise the distribution $\{(1 - P_n) \pi_n\}_{n \in \mathbb{N}}$ so that the average is taken across households conditioning on an epidemic taking place. The reduction due to the fact that vaccinated households escape infection altogether is included in the effects of vaccination on R_g and μ_W before. The particular case when households are selected for vaccination independently of their size, i.e. $P_n = P$ for all n , gives $\mu_H^V = \mu_H$ (Equation (4.8.1)).

Now, consider two household sizes k and n , with $k < n$. Denote by $\mu_H^{V_k}$ and $\mu_H^{V_n}$ the size-biased average final sizes of a household epidemic when vaccination is targeted only to households of size k or n , respectively. When only households of size k are vaccinated, $P_n = 0$ and therefore $P = \pi_k P_k$

$$\begin{aligned}\mu_H^{V_k} &= \frac{1}{1 - P} \left[\mu_H^{(k)} \pi_k \left(1 - \frac{P}{\pi_k}\right) + \mu_H^{(n)} \pi_n \right] \\ &= \frac{1}{1 - P} \left[\mu_H^{(k)} \pi_k + \mu_H^{(n)} \pi_n - P \mu_H^{(k)} \right].\end{aligned}$$

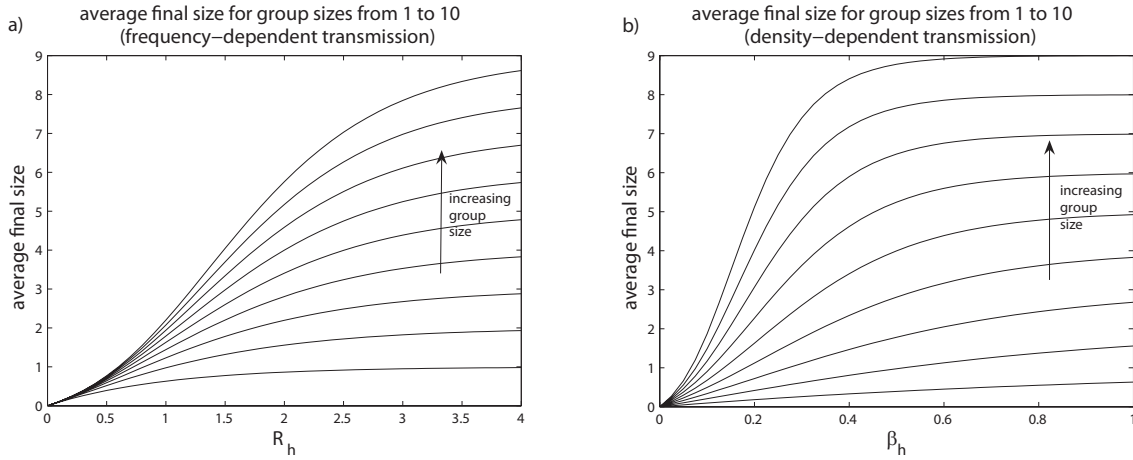


Figure 4.6: Average final sizes (excluding the primary case) of epidemics in households of sizes n from 1 to 10, with 1 initial infective, under the assumptions of a) frequency-dependent and b) density-dependent contact rates. Given a fixed duration $d = 1$ of the infectious period, the horizontal axis in a) represents the average number of contacts $R_h = (n - 1)\beta_h$ generated by a single infective, which remains constant as the household size increases (β_h is inversely proportional to $n - 1$). The horizontal axis in b) represents the individual-to-individuals infection rate β_h , which remains constant as the household size increases (R_h increases).

When only households of size n are vaccinated, the result is the same but with indexes k and n swapped. The only asymmetry is in the last term of the last line and therefore

$$\mu_H^{V_n} \leq \mu_H^{V_k} \iff \mu_H^{(k)} \leq \mu_H^{(n)}, \quad (4.8.2)$$

which I conjecture being always true when $k < n$ (figure 4.6).

If density-dependent contact rate is assumed, a formal proof of Conjecture (4.8.2) relies on Theorem 4.5.1. In fact, given $k < n$, the assumption of density-dependent contact rate guarantees that the average final size $\mu_H^{(k)}$ of an epidemic with a single initial case and k initial susceptibles is the same as an epidemic with a single initial case, k initial susceptibles and $n - k$ initial immunes; therefore, in a household of size n , Theorem 4.5.1 states that, after the vaccination of $n - k$ individuals, $\mu_H^{(k)} \leq (k/n)\mu_H^{(n)}$ and therefore that $\mu_H^{(k)} \leq \mu_H^{(n)}$.

Figure 4.6a suggests that the result is true even under the assumption of frequency-dependent contact rate, which empirical evidence shows to fit epidemic better than when density-dependent contact rate is assumed: for example, de Jong *et al.* (1995) and Hethcote

(2000) concluded that the frequency-dependent assumption is never rejected from data fitting, as opposed to the density-dependent assumption; furthermore, even in the case of small mixing groups like households, data fitting in (Cauchemez *et al.*, 2004) of the individual-to-individual infection rate $\lambda = \beta/n^\eta$ in households of size n lead to a maximum likelihood estimate of $\eta = 0.97$, very close to 1.

I have not been able to prove the result of Theorem 4.8.1 in general, but I conjecture its validity, when frequency-dependent contact rate is assumed, based on graphical arguments. Figures similar to Figure 4.8.1a are obtained for various distributions of total infectivity across individuals and also when the individual-to-individual infection rate scales with the total group size n (as in Cauchemez *et al.*, 2004), instead of $n - 1$. $\square$

Theorem 4.8.1 suggests to vaccinate the largest households first: an administratively convenient proxy for this strategy could be that of targeting the households with children first, as large households typically consist in families with children (House & Keeling, 2008). Note that this also means that the benefits of this strategy shown in Theorem 4.8.1 are further enhanced by the fact that children are known to mix more intensely than adults and therefore to contribute more to the infection spread.

However, Theorem 4.8.1 refers to the case when the constraint is on the total fraction of the population vaccinated, as for example when the cost of vaccine production is high or when a limited number of doses is available. Vaccinating households with children is not necessarily subject to the same constraint.

If the constraint is placed on the total number of households that can be vaccinated (for example, when the cost of producing doses of vaccine is negligible compared to the cost of selecting the household and delivering the vaccine), the result of Theorem 4.8.1 is even further enhanced by the fact that the cost of vaccinating the same number of large or small households is the same, while the positive effects of vaccinating larger households are still as in the previous case.

Finally, it is worth recalling that all these strategies, although convenient from an administrative point of view, are sub-optimal, since they “waste” vaccine doses in the

attempt to vaccinate all individuals in a household and therefore do not exploit the herd immunity effect (Ball *et al.*, 1997; Becker & Starczak, 1998; Ball & Lyne, 2002; Ball *et al.*, 2004) and they do not target those categories of individuals (e.g. children) that are known to mix more intensely among themselves or with others (see House & Keeling, 2008).

4.9 Sensitivity of results to simplifying assumptions

Here, the assumptions made in Section 4.2.1 are discussed, namely that the behaviour of an infected individual is independent of the behaviour of their infector and that external conditions that may affect individuals' behaviour (e.g. seasonal variations in contact patterns) remain constant while the epidemic progresses.

The first assumption is commonly considered in mathematical epidemiology: in fact, little analytical work has focused on epidemics where this assumption is violated. Notable exceptions are the papers by Kendall & Saunders (1983) and Svensson & Scalia-Tomba (2001), where a model of two competing strains of a pathogen spreading in a population and conferring cross-immunity to infected individuals is studied, and the paper by Ball & Britton (2007), where individuals can develop a more or less severe form of infection and infected individuals are more likely to generate secondary cases with the same form of infection as their own.

The present chapter mainly deals with reproduction numbers, which are by definition average values, and are therefore well defined without specifying particular assumptions concerning the spread of the infection. However, the nice feature of some reproduction numbers, such as those described in Equations (2.4.3), (4.3.3) and (4.3.4), is the possibility of expressing them as functions of simpler terms, that mainly consist of other reproduction numbers (i.e. R_g) and average sizes of local epidemics. These results are not straightforward since they involve quantities that are dependent on one another (Ball, 1986; Ball & Neal, 2008).

The simplest example to discuss is given by the factorisation of the household repro-

duction number for the households model:

$$R_* = R_g(1 + \mu_H) \quad (4.9.1)$$

(i.e. Equation (2.4.3): R_* is the household reproduction number, R_g is the average number of global contacts an infective makes and μ_H is the average final size of a “typical” household epidemic, excluding the initial infective). In this case, for example, the average number of global infections generated by individuals involved in a household epidemic depends on their total infectivity (e.g. through the length of their infectious period), which is also responsible for the final size of the local epidemic itself. The factorization of Equation (4.9.1) has been proved for the standard *SIR* epidemic model (Ball, 1986), in which case it is referred to as Wald’s identity for epidemics (a particular case of a general identity appearing in Wald, 1947), but it may not hold in other cases.

The original proof of the Wald’s identity for epidemics, when the standard *SIR* model of Section 2.1.4 is assumed, is reported in Ball (1986) (see also Andersson & Britton, 2000). An alternative proof is the following.

Consider a household of size n and an epidemic starting with individual $i = 1$ infected. Denote by Z the final number of individuals ultimately infected in the household epidemic (excluding the initial case) and by Υ the sum of the global “infection pressure” on all individuals in the population due to all the individuals involved in the household epidemic, i.e. the area under the curve representing the infection prevalence over time, multiplied by the one-to-all global infection rate β :

$$\Upsilon = \beta I_1 + \sum_{i=2}^{Z+1} \beta I_i$$

(the sum contains Z terms), where the I_i ’s represent the random lengths of infectious period (in case of infection) and are independent and identically distributed according to the random variable I .

Wald’s identity for epidemics states that $\mathbb{E}[\Upsilon] = \beta \mathbb{E}[I](1 + \mathbb{E}[Z])$, which corresponds to Equation (4.9.1) since, by definition, $R_* = \mathbb{E}[\Upsilon]$, $\mu_H = \mathbb{E}[Z]$ and $R_g = \beta \mathbb{E}[I]$. Note that,

in the random sum Υ , Z and the I_i 's are not independent of one another. The “trick” that allows the factorization of Equation (4.9.1) is that Z is independent of the I_i 's of all the remaining individuals that are not infected and therefore play no role in the household epidemic. Let Ξ be the sum of the (hypothetical) infectiousness of the individuals that escape infection:

$$\Xi = \sum_{i=Z+2}^n \beta I_i.$$

Now the factorization is allowed (see any probability textbook, e.g. Grimmett & Welsh, 1986, Section 4.4) and therefore (n is constant):

$$\mathbb{E}[\Xi] = \mathbb{E}[n - 1 - Z] \beta \mathbb{E}[I] = (n - 1 - \mathbb{E}[Z]) \beta \mathbb{E}[I].$$

Finally, since $\mathbb{E}[\Xi + \Upsilon] = n \beta \mathbb{E}[I]$,

$$\mathbb{E}[\Upsilon] = n \beta \mathbb{E}[I] - \mathbb{E}[\Xi] = \beta \mathbb{E}[I] (1 + \mathbb{E}[Z]).$$

Note how the result depends on the I_i 's of the individuals that escape infection and on their independence in relation to the household epidemic. This may not be true under more general assumptions: for example, when mild and severe infectives generate secondary cases of their same type the I_i 's of the individuals who escaped infection are not well defined, because they depend on who would have been more likely to infect them and this in turn depends on how the epidemic progresses and on the number of cases of each type occurring.

Another alternative proof, suggested by Prof. F. G. Ball (personal communication), is obtained by noting that

$$\Upsilon = \beta I_1 + \sum_{i=2}^n \beta I_i \Theta_i,$$

where Θ_i is the indicator function of the event “individual i is ultimately infected” (i.e. with value 1 if the event is verified and 0 otherwise). Since the characteristics of the disease history of individual i are independent of whether i is infected or not,

$$\begin{aligned} E[\Upsilon] &= \beta E[I_1] + \sum_{i=2}^n \beta E[I_i \Theta_i] \\ &= \beta E[I_1] + \sum_{i=2}^n \beta E[I_i] E[\Theta_i] \end{aligned}$$

and, since all initial susceptibles are identical,

$$E[\Theta_i] = \frac{E[Z]}{n-1}.$$

Therefore

$$E[\Upsilon] = \beta E[I](1 + E[Z]),$$

i.e. Equation (4.9.1).

To sum up, when the standard *SIR* model is considered, Wald's identity for epidemics can be formally proved. Equations (4.3.3) and (4.3.4) can in turn be proved by repeatedly invoking Wald's identity for epidemics and by observing that, in the infinite population limit, when households and workplaces are randomly linked and during the early phase of the infection spread, local epidemics develop independently of one another (see Ball & Neal, 2008).

Note, however, that the key property of the standard *SIR* model used in all there factorisation is that it is possible to attribute an “infectious life” to each individual before the epidemic starts, under the caveat that individuals that will escape infection will not “use” their infectious life.

As already shown in Section 3.6, this is true only when:

1. the disease history of infected individuals does not dependent on the individuals that caused their infection;
2. the disease history of infected individuals does not depend on the absolute time of their infection.

In the attempt to extend the results of this paper to more general models than the standard *SIR* one, it is necessary to understand which of the assumptions of the standard *SIR* model can be relaxed.

Firstly, the description of the individual-to-individual contact patterns as Poisson processes, independent of one another, seems to be an unavoidable assumption: apparently,

no mathematical description of a stochastic epidemic exists for which this assumption is not (explicitly or implicitly) considered. This study is no exception in this sense.

The second assumption is that of homogeneous mixing. Again, this cannot be relaxed, at least locally, and is therefore assumed within each household or workplace.

The third assumption concerns the particular dynamics of the standard *SIR* model, and this can be relaxed: the argument in Section 3.2 shows that the final size distribution of the standard *SIR* epidemic model is invariant with respect to the actual infectiousness profile of individuals and depends only on their total infectiousness (see also Ludwig, 1975, and the final size formula in Section 3.7 or in Ball, 1986). This argument can therefore be invoked to extend the results of this chapter to all other models with time varying infectivity and including latent periods of any form. However, as pointed out in Section 3.6 or in Pellis *et al.* (2008), this property again relies on conditions 1. and 2. stated above.

Even if they are not both satisfied, the average size of a local epidemic is well defined, although its dependence on more basic parameters may in general be different from that of the standard *SIR* model. However, a factorisation like that of Wald's identity for epidemics may not hold in this case, although I am not aware of either proofs of its validity or counterexamples. Therefore, as a precautionary measure, assumptions **a)** and **b)** have been required in Section 4.2.1.

A further comment concerns the possibility of relaxing assumption **b)**. Typical external conditions that would affect individuals' behaviour and infectiousness, like seasonality or other social factors, usually change slowly and, if the epidemic is short enough, they can be considered approximately constant. If the epidemic is long or their change is not negligible, the results of this chapter may still be considered valid if their change is sufficiently slow that they can be considered constant at least for the duration of the infectious period of an infection unit in the model (an individual, a household or a clump, depending on the context); in this case, the values of the parameters will slowly change as the epidemic progresses. Note that such an argument is reasonable for reproduction numbers like R_0 for the homogeneous mixing model, R_* for the households model and R_H for the households-

workplaces model, since the infectiousness spread by a single unit (individual or household) in all these cases is limited in time. As far as the clump-to-clump reproduction number R_* is considered, it is necessary to recall that the time needed for a clump to spread its infectivity may be comparable to that of the whole epidemic (see Section 4.3.2), therefore causing the effects of the external changes not to be limited in time.

There are, however, some important examples of external condition which occur suddenly and for which the above comment does not apply: the most important of them is the implementation of a control policy. In this case, condition 2) may not be satisfied, since an individuals' infectiousness during a local epidemic would now depend on the absolute time of infection. For this reason, rigorous mathematical papers often explore the effect of different control policies by assuming that they are implemented before the infection spreads, so that external conditions remain unaltered throughout the epidemic (e.g. Roberts, 2004; Aldis & Roberts, 2005; Roberts *et al.*, 2007).

Strictly speaking, the real-time implementation of a control policy may invalidate the results presented in the paper. However, it is arguable that, except during the local epidemics that occur when the control policy starts to be implemented, about which nothing formal can be said in general, the population conditions can be roughly assumed to remain constant in time up to the time of the implementation and from that moment onwards.

To sum up, in most practical contexts, assumption b) can be dropped, but it may be necessary to keep track of how the reproduction numbers change in time and to accept that some results may not hold at and around times of sudden changes in social behaviour.

4.10 Summary, comments, limitations and future research

In this chapter, a model with households and workplaces has been defined and studied. Homogeneous mixing within each household and each workplace is superimposed to a background homogeneous mixing (at a small rate). As the interest has been placed on

reproduction numbers, the attention is focused on the early phase of the epidemic in the limit of an infinite population. In order to allow for analytical results, households and workplaces are assumed to be connected at random, i.e. forming a bipartite random network.

A new reproduction number for households R_H has been defined and studied, based on the idea that the infection can exit from a households both via a global infection or via a workplace epidemic. R_H overcomes some of the limitations from which the clump-to-clump reproduction number R_* suffers (Ball & Neal, 2002). In particular, it can never become infinite and it is therefore always informative about how much effort should be put into controlling the epidemic. As a numerical example, let $(R_g, \mu_H, \mu_W) = (1/3, 1, 1)$: while $R_* = \infty$, $R_H = 2$. In order to control the epidemic it is therefore necessary to reduce the household-to-household transmission by a factor 2. By keeping fixed both μ_H (i.e. acting only on the between-households transmission) and R_g (for example because one might think it cannot be easily measured), it is possible to find (now R_H changes monotonically with μ_W) that μ_W needs to be reduced from 1 to 0.2 to achieve control. Of course the same result can be obtained from R_* , since it shares the same threshold with R_H , but requiring a reduction of R_* from ∞ to 1 is less informative from a practical point of view.

Another remark in support of R_H is that a selected household requires a limited time to infect other households. Therefore, in a reasonably large but finite population, it is often possible to ignore the depletion of susceptibles for long enough to be able to observe a number of household-to-household infections which is sufficiently large for the law of large numbers to be invoked, i.e. for having the epidemic behaviour mainly driven by the household-to-household reproduction number (an average value). On the contrary, already when $\mu_H\mu_W$ is smaller than but close to 1, the time required for a clump to form may be so long that any realistic population size could be too small to make the exponential phase long enough to observe clumps reaching the end of their infectious period before the depletion of susceptibles in the population starts having a major role in the epidemic dynamics.

An inequality (Equation (4.5.2)) for the critical vaccination coverage for households has been formally proved, which provides an upper bound for the effort needed to safely contain the epidemic when vaccination is targeted at randomly selected households instead of randomly selected individuals: such a vaccination policy, although not optimal (Becker & Starczak, 1998; Ball & Lyne, 2002; Ball *et al.*, 2004; House & Keeling, 2008), is likely to be the one adopted in most practical cases (Becker & Utev, 1998; Fraser, 2007). Note that, if the assumption of random connections between households and workplaces is not met, then Inequality (4.5.2), although not strict, remains valid and therefore the vaccination of a proportion of randomly selected households equal to $1 - 1/R_H$ still guarantees infection extinction. Other household vaccination strategies have been considered, with the conclusion that targeting the largest household first generally achieves the best result.

It has also been commented on the fact that the same argument that leads to Equation (4.3.4) for R_H can be used to derive a similar reproduction number R_W for workplaces (Equation (4.3.5)). All the results valid for R_H are also valid for R_W when the roles of households and workplaces are exchanged. The attention has been mainly focused on R_H , since the household is often a more natural target for control policies than a workplace, because it is more clearly identifiable, because of the often higher internal transmission coefficient or because most of the population lives in households. In some contexts, however, R_W may be more relevant than R_H , for example when vaccination is delivered to children in schools or other forms of control policy targeted at workplaces are studied.

The presence of workplaces in addition to households allows to study the effect of social distancing achieved when people stop going to work. Such action is surely a draconian measure and is not likely to be implemented because of its economical implications. However, school closure is not only feasible but also more likely to occur under conditions of particular stress for the health care system and, as a side effect, it may force many workers to remain at home with their children.

It is reasonable to assume that school or workplace closure has an effect on other routes

of transmission, since individuals not at work spend their time elsewhere. In this case, it may well happen that a reduction of some parameters (e.g. R_g and μ_W) is accompanied by an increase in others (e.g. μ_H if, for example, more time is spent at home). These types of change in the parameters set can result in different reproduction numbers changing in different directions (see Section 4.3.5). Therefore, their effect may be read differently by the two reproduction numbers R_* and R_H and it can be argued that the information provided by both parameters should be taken into account and compared.

Other forms of intervention, especially those based on real-time reaction to detected or suspected cases, are more difficult to study, since they are highly dependent on the time delays between infection, detection of a case and policy implementation. This requires a detailed description of the dynamics of the individual-to-individual infection process, which has not been given in this model.

The choice of not modelling the dynamics of the infectious-to-susceptible interaction has the clear advantage that the results of the model hold quite generally. As a drawback, it makes the quantities in the formulae for the reproduction numbers less easily related to more fundamental ingredients (like infectivity, duration of infectiousness and so on) and therefore more difficult to measure. In particular, while μ_H and μ_W could be in principle estimated from data on where infected cases live and work, estimation of R_g may be more difficult. However, the method illustrated in Section 7.1 for the estimation of R_H from the real-time growth rate, can be used also to quantify R_g .

This approach presents other limitations. First of all, it has been assumed that each individual selects a workplace at random, thus giving rise to a social structure which can be described as a bipartite random network. Such a condition is clearly not met in a real society, because there is a spatial component that cannot be ignored (e.g. children going to school close to home, adults trying to find a job not too distant from where they live) and some forms of social aggregation (e.g. work colleagues deciding to register their children to the same school; cultural, religious or political affinities and so on). Relaxing this hypothesis can probably be achieved only by using individual based simulations and

is briefly investigated in Section 7.2. However, this random network assumption provides the worst possible scenario, since it minimizes the chance that the infection re-enters in the same infected household or workplace, an event that would slow down the epidemic and reduce its final size.

A second limitation is the idealisation of the concept of workplace, assumed to be a place where individuals spend most of their time while out of home: in practice, it is often unclear who should be considered as sharing a workplace with whom. However, the same definition problems arise in quantifying the effects of most of the social distancing control measures implemented during an infection outbreak. A more thorough discussion is reported in Section 6.2.1.

Many other interesting issues should be further explored: first, gaining a clear understanding of the differences between a households-workplaces model, as described in this chapter, and other structured population models where the local saturation of susceptibles is not taken into account (e.g. in Roberts, 2004; Aldis & Roberts, 2005; Roberts *et al.*, 2007) or is taken into account only for the within household transmission (e.g. as in Becker *et al.*, 2005); second, investigating the benefits of a further distinction between schools and workplaces, motivated by a difference in size distributions, transmission parameters, contact rates and other characteristics of the individuals (adult/child), the economical implications of their closure, as well as realistic feasibility of parameter estimation in the two environments (Cauchemez *et al.*, 2008); third, exploring how important the presence of workplaces is, in addition to the household structure, in correctly estimating the target of the critical vaccination coverage, as in Becker & Utev (1998) for the comparison between a households model and a model with homogeneous mixing.

Furthermore, in order to assess the impact that school and workplace closure might have in an emerging outbreak situation, extensive effort needs to be focused on quick and robust methods for the estimation of the changes in global and household transmission that are likely to occur when the control policy is implemented. Some *a priori* analysis could be carried out by assuming that the social behaviour in such conditions is similar to what is

obtained during school holidays, as in Cauchemez *et al.* (2008). However, human behaviour under considerable stressful conditions, like during an emerging outbreak, is difficult to predict and may turn out to be substantially different from what can be observed during normal holiday time. For this reason, real-time estimation methods remain invaluable tools and much effort should be spent in identifying which data need to be collected and how.

Chapter 5

Real-time growth rate

5.1 Introduction

As already mentioned in Section 2.2.3, reproduction numbers, probabilities of a large outbreak, final size distributions and final size averages and distributions all represent important quantities to summarise much about the spread of the infection. Under the conditions that allow Ludwig's construction to be performed (see Section 3.6) all this information is independent of the dynamics of both the transmission process at the individual level and of the whole epidemic. However, the actual dynamics are important and useful in themselves, as they convey information about the time available to react to the spread and therefore of the effectiveness of all the control policies occurring in real-time and the impact of delays in their implementation.

More precisely, information about the epidemic dynamics would allow gaining understanding, for example, of the following issues:

1. how long the epidemic is going to take to sweep through the population, and therefore for how long the health care system is going to be under stress;
2. how quick the reaction should be at the beginning of the epidemic to control it at the source;

3. how effective “responsive” control interventions (e.g. prophylactic treatment of detected cases to reduce their infectivity, prophylactic treatment of known contacts of a case, isolation or quarantine of cases or entire households, pre-emptive culling intervention for animals, various forms of social distancing, ...) are likely to be;
4. how large the incidence at the peak of the epidemic is expected to be (shorter and more “explosive” epidemics have a larger incidence at the peak than longer ones with the same final size) and therefore how much stress the health care system is expected to experience in the most critical moment.

The simplest piece of information concerning the epidemic dynamics is the real-time growth rate. The review of the literature given in Sections 2.2.3 and 2.3.4 presented some standard results, which are briefly recalled in the next section for convenience. The following Sections 5.3 and 5.4 extend the results to the households and the households-workplaces model respectively. Section 5.5 focuses on the dynamics in small groups when the Markovian case of the standard *SIR* model with constant recovery rate is assumed, while Section 5.6 describes an approximation to the dynamics in a small group when a non-random time-varying infectivity profile is considered.

5.2 Models with no local saturation

Consider the case of a large homogeneously mixing population in which each infective makes infectious contact with all other individuals (one-to-all) according to a non-random time-since-infection infectivity profile given by a function (or, more in general, a distribution) $\beta(\tau)$. In Section 2.2.3 it has been shown that the real-time growth rate r , conditional on a large epidemic taking place, is given by the implicit solution of the Lotka-Euler equation

$$\int_0^{+\infty} \beta(\tau) e^{-r\tau} d\tau = 1, \quad (5.2.1)$$

where the left-hand side is the Laplace transform of β calculated in r , $L_\beta(r)$.

Sections 2.3.4 and 2.3.5 extended the result to more general cases, among which that of the single-type model with random disease history: all individuals have the same susceptibility and a random infectivity drawn from the same set of disease histories. In general, let $B(\tau)$ represent a random disease history, drawn from a suitable set $\mathcal{B}$ with probability distribution denoted by the probability measure $\zeta(X)$ (i.e. $\int_{\mathcal{B}} d\zeta(X) = 1$). Thanks to the fact that the Laplace transform L_B is linear in B , the equation for the real-time growth rate r is:

$$\begin{aligned} 1 &= \int_{\mathcal{B}} \left(\int_0^{+\infty} B(\tau) e^{-r\tau} d\tau \right) \zeta(dB) \\ &= \int_0^{+\infty} \left(\int_{\mathcal{B}} B(\tau) \zeta(dB) \right) e^{-r\tau} d\tau \\ &= \int_0^{+\infty} \beta(\tau) e^{-r\tau} d\tau \end{aligned}$$

where $\beta(\tau)$ represents the average infectiousness profile of an individual.

The classical example is given by the standard *SIR* model. In this case, $B(\tau) = \beta \Theta_{[0,I]}(\tau)$, where β represents the constant one-to-all infection rate, $\Theta_{[l_1,l_2]}$ represents the Heaviside function, defined as

$$\Theta_{[l_1,l_2]}(x) = \begin{cases} 1 & \text{when } l_1 \leq x \leq l_2, \\ 0 & \text{otherwise,} \end{cases}$$

and the length of the infectious period I is drawn from an exponential distribution with parameter ν . In this case, $\mathcal{B} = \{B(\tau) = B_I(\tau) = \beta \Theta_{[0,I]}(\tau), I \in \mathbb{R}^+\}$ and $\zeta(dB_I) = \nu e^{-\nu I} dI$. Then

$$\begin{aligned}
1 &= \int_0^{+\infty} \left(\int_0^{+\infty} \beta \Theta_{[0,I]}(\tau) e^{-r\tau} d\tau \right) \nu e^{-\nu I} dI \\
&= \int_0^{+\infty} \left(\int_0^{+\infty} \beta \Theta_{[0,I]}(\tau) \nu e^{-\nu I} dI \right) e^{-r\tau} d\tau \\
&= \int_0^{+\infty} \left(\int_0^{+\infty} \beta \Theta_{[\tau,+\infty)}(I) \nu e^{-\nu I} dI \right) e^{-r\tau} d\tau \\
&= \int_0^{+\infty} \left(\int_\tau^{+\infty} \beta \nu e^{-\nu I} dI \right) e^{-r\tau} d\tau \\
&= \int_0^{+\infty} \beta e^{-\nu\tau} e^{-r\tau} d\tau \\
&= \frac{\beta}{r + \nu},
\end{aligned}$$

where it has been used the fact that $\Theta_{[0,I]}(\tau) = \Theta_{[\tau,+\infty)}(I)$. Solving for r gives the usual $r = \beta - \nu$.

When a random susceptibility is considered in addition to the random infectivity, care needs to be taken in how the average infectivity of an individual is defined, because more susceptible individuals are more likely to be infected. If susceptibility and infectivity are represented by the random quantities A (a scalar) and $B(\tau)$ (a function), respectively, then the average infectivity profile at each time τ is given by the weighted average $\beta(\tau) = \mathbb{E}[AB(\tau)]$, which is in general different from the product of the averages when the two random quantities are correlated.

When a multitype model is considered, with susceptibility and disease history assumed to be non-random for each type, it is necessary to define the operator

$$K_r = \left(\int_0^{+\infty} \beta_{ij}(\tau) e^{-r\tau} d\tau \right) = (L_{\beta_{ij}}(r)),$$

where $\beta_{ij}(\tau)$ represents the average infectivity profile from a single infective of type j towards all individuals of type i , which takes into account, as well as the particular mixing patterns, also the infectivity of type j , the susceptibility of type i , and the proportion of individuals of type i in the population. The real-time growth rate r is therefore obtained as the solution of the implicit equation

$$\rho(K_r) = 1. \quad (5.2.2)$$

The left-hand side $\rho(M)$ is the dominant eigenvalue of the matrix M (see Section 2.3.2) and the solution exists and is unique when $R_0 > 1$, i.e. when a large epidemic is possible (see Section 2.3.4 or Diekmann & Heesterbeek, 2000)

The particular case of separable mixing (Section 2.3.5) allows for great simplification and makes the computation of the infectivity profile of an average individual in the multitype model very similar to that of the model with random susceptibility and infectivity: a weighted average of the infectivity of different types, with weights proportional to susceptibility-availability $\alpha_i = a_i h_i$ of each type, i.e. the product of the susceptibility a_i and the proportion h_i of individuals of type i in the population.

The analysis becomes more complex when randomness is superimposed on the multitype structure (*multitype randomised model*). However, at least in the case of separable mixing, it is still sufficient to average across the distribution of disease histories for each type j , with weights given by the distribution of susceptibilities, to recover the same condition above.

If separable mixing cannot be assumed for the multitype model, then the analysis is even more complex. It is still reasonable to expect that the system can reduce to that of a multitype model, by averaging across each type, although a general proof would be needed. A particular case of this type is considered in Section 5.4.2.

5.3 Households model

Since a random disease history is simply averaged thanks to the linearity of the Laplace transform (average weighted by the susceptibility, if present), in the remainder of the chapter individuals are assumed to have susceptibility 1 and non-random disease histories with time varying infectivity $\beta(\tau)$. Furthermore, both in the households model (Section 2.4) and in the households-workplaces model (Chapter 4), only a single type of individual is

considered, who however mixes with others at different rates in households, workplaces and in the community. Therefore, $\beta(\tau)$ will denote the time-varying infectivity of an infective in the community and $\beta_h(\tau)$ and $\beta_w(\tau)$, both proportional to $\beta(\tau)$, will denote the time varying infectivity of an infective towards other household and workplace members.

As a passing nod, it is worth mentioning at this stage that requiring β, β_h and β_w to be proportional to one another, despite being mathematically convenient, represents a rather strong assumption. Although it could be reasonable in the case of mild infections, it may also be argued that an infected case is likely to spend more time outside the household at the beginning of the infectious period and much less later on, when the advent of symptoms typically forces the individual to stay at home, thus increasing the within-household infectivity and reducing the infectivity in other environments. Extensions of the model to infectivity profiles β, β_h and β_w not proportional to one another are possible, but will not be considered here.

5.3.1 Real-time growth rate for households

Consider initially a households model with all households with the same size n . When a household is infected, a stochastic epidemic spreads through it. Let $U \in \mathcal{U}$ be a parameter vector that fully describes the household epidemic and let $\zeta(U)$ be the measure over $\mathcal{U}$ associated with the probability density of occurrence of each possible household epidemic. Despite the disease history of individuals being non-random, the disease history of a household is random and is denoted by $B_H^U(\tau_H)$, where τ_H represents the time since the infection of the household.

However, using the same arguments as before, the real-time growth rate for households, r_H , is determined by the average household infectivity $\beta_H(\tau_H) = \int_{\mathcal{U}} B_H^U(\tau_H) \zeta(dU)$ as the solution of

$$L_{\beta_H}(r_H) = 1. \quad (5.3.1)$$

In the case of households of different sizes, the model can be seen as a multitype randomised model with separable mixing, where the size of each household determines its

type. In fact, if h_n represents the fraction of households of size n , a household of size n has probability $\pi_n = nh_n / \sum_n nh_n$ (see Section 4.8) of being infected and, after infection, exerts a random time-varying infectivity $\beta_H^{U_n}(\tau_H)$ with average, say, $\beta_H^{(n)}(\tau_H)$. The overall real-time growth rate is given again by Equation (5.3.1), where now $\beta_H(\tau_H) = \sum_n \pi_n \beta_H^{(n)}(\tau_H)$.

5.3.2 Real-time growth rate for individuals

The real-time growth rate r_H refers to households. However, the interest usually lies on the real-time growth rate that describes the exponential growth of the number of individuals. The following result reassures that they always coincide.

Theorem 5.3.1. *Let r be the real-time growth rate for the exponential growth of the number of infected individuals and let r_H be the real-time growth rate for the exponential growth of the number of infected households. Then $r = r_H$.*

Remark 5.3.2. The result is reasonable: if $H(t)$ is the prevalence of households at time t , then the number of infected individuals $Y(t)$ is roughly given by $(1 + \mu_H)H(t)$, and if $H(t) \propto e^{rt}$, then $Y(t)$ shares the same growth rate.

Proof. For the sake of simplicity, consider again households all of the same size n . The result can be readily extended to households of variable sizes. As done in Section 4.2.2, it is convenient to refer to all the non-primary cases as secondary cases, independently of whether they have been infected directly or indirectly by the primary case.

Denoting by $\gamma_H(\tau_H)$ the average rate at which new cases appear in the household, it is shown in Section 5.8 below that the total infectivity of the household can be written as

$$\beta_H(\tau_H) = \beta(\tau_H) + \beta * \gamma_H(\tau_H),$$

where $f * g$ denotes the convolution of the functions f and g . The factorisation is not obvious and relies on the same assumptions discussed in Sections 3.6 and 4.9.

The real-time growth rate r_H for households is given by the implicit solution of

$$L_{\beta_H}(r_H) = 1.$$

The real-time growth rate r for individuals is obtained when the two-type model with primary and secondary cases, introduced in Section 2.4.5, is considered. In this framework, a primary case generates other primary cases following $\beta(\tau) = \beta(\tau_H)$ (the household is infected when the primary case is infected) and secondary cases following $\gamma_H(\tau_H)$; a secondary case generates only primary cases following $\beta(\tau)$. For r to be the real-time growth rate for individuals, the linear operator

$$K_r = \begin{pmatrix} \int_0^\infty \beta(\tau_H) e^{-r\tau_H} d\tau_H & \int_0^\infty \beta(\tau) e^{-r\tau} d\tau \\ \int_0^\infty \gamma_H(\tau_H) e^{-r\tau_H} d\tau_H & 0 \end{pmatrix} \quad (5.3.2)$$

must have dominant eigenvalue 1. A quick look at the characteristic equation (e.g. using Descartes' sign rule) reveals that the two eigenvalues have opposite signs. Imposing 1 to be an eigenvalue (which means it can only be the dominant one) implies that

$$1 - \int_0^\infty \beta(\tau_H) e^{-r\tau_H} d\tau_H = \left(\int_0^\infty \gamma_H(\tau_H) e^{-r\tau_H} d\tau_H \right) \left(\int_0^\infty \beta(\tau) e^{-r\tau} d\tau \right)$$

which, by properties of the Laplace transform (the linearity and the fact that the Laplace transform of a convolution is the product of the Laplace transforms), can be written as

$$L[\beta + (\gamma_H * \beta)](r) = 1. \quad (5.3.3)$$

Since r and r_H are both the unique solution (unique because the left-hand side is monotonic; see Section 2.2.3) of the same equation, it follows $r = r_H$. $\square$

5.4 Households-workplaces model

5.4.1 Fixed household and workplace sizes

Theorem 5.3.1 suggests that, during the exponentially growing phase, the real-time growth rate of all structures (with finite average size) is the same. Therefore, it is sufficient to calculate the easiest one to obtain: in this case, the one for households.

Note that, when clumps have finite average size, the same real-time growth rate can also be thought as referring to clumps. However, the usefulness of this interpretation is limited, given that the time a clump requires to form can in general be comparable to that of the entire epidemic. Therefore, in realistic settings, any exponentially growing phase could be too short to observe enough clumps to motivate a deterministic approximation.

As done in Section 4.3.3, households are distinguished according to how their primary case is infected. Again, assume initially that all households have the same size n_H and all workplaces have size n_W . This can be seen as a two-type randomised model.

In particular, denote as before by τ_H the time since the infection of a household and by $\gamma_H(\tau_H)$ the average rate at which new cases are generated in a household. Analogously, consider τ_W and $\gamma_W(\tau_W)$ defined in the same fashion for workplaces.

The average rate at which a household infected locally infects another household locally is given by $\beta_{LL}(\tau_H) = \gamma_W * \gamma_H(\tau_H)$ (recall that the primary case is not counted actively when the local structure is infected locally; see Section 4.3.3). Analogously the average rate at which a household infected globally infects other households locally is given by $\beta_{LG}(\tau_H) = \gamma_W(\tau_H) + \gamma_W * \gamma_H(\tau_H)$. The other two cases are: $\beta_{GG}(\tau_H) = \beta_{GL}(\tau_H) = \beta(\tau_H) + \beta * \gamma_H(\tau_H)$.

The real-time growth rate r is obtained by imposing that the operator

$$K_r = \begin{pmatrix} \int_0^{+\infty} \beta_{LL}(\tau_H) e^{-r\tau_H} d\tau_H & \int_0^{+\infty} \beta_{LG}(\tau_H) e^{-r\tau_H} d\tau_H \\ \int_0^{+\infty} \beta_{GL}(\tau_H) e^{-r\tau_H} d\tau_H & \int_0^{+\infty} \beta_{GG}(\tau_H) e^{-r\tau_H} d\tau_H \end{pmatrix} \quad (5.4.1)$$

has dominant eigenvalue equal to 1.

Recall from Section 4.3.3 that

$$\rho(K_0) = \rho \begin{pmatrix} R_{LL} & R_{LG} \\ R_{GG} & R_{GG} \end{pmatrix} = R_H.$$

Since the interest here is on studying the real-time growth rate in the case of a large epidemic $R_H = \rho(K_0)$ is required to be greater than 1. In addition, since $\lim_{r \rightarrow \infty} \rho(K_r) = 0$

and $\rho(K_r)$ is a monotonic function of r (see Section 2.3.4), a unique solution to the equation

$$\rho(K_r) = 1 \quad (5.4.2)$$

exists and represents the real-time growth rate.

When households and workplaces have variable sizes, the computation of r reduces again to that of a two-type model, although a formal argument for it requires a more cumbersome notation. The key points, as usual, are that the epidemics in a household and a workplace evolve independently of one another, and that the size of one structure is independent of the size of the other because of the bipartite random network assumption.

The next section focuses on supporting this last statement by sketching a proof for it, and can be skipped if the reader is not particularly interested in the details.

5.4.2 Variable household and workplace sizes

First, the rate at which a household infected either locally or globally infects another household globally does not involve workplaces and is therefore obtained by the same argument used for the households model in the previous section.

In the case of β_{LL} and β_{GL} , the additional presence of a workplace epidemic has to be taken into account. The notation is the following: $\pi_{n_H}^H$ and $\pi_{n_W}^W$ represents respectively the probability that a randomly selected individual belongs to a household of size n_H and a workplace of size n_W ; $\gamma_H^{(n_H)}(\tau_H)$ and $\gamma_W^{(n_W)}(\tau_W)$ represent respectively the rate of appearance of new cases in a household of size n_H at time τ_H after the infection of the household and in a workplace of size n_W at time τ_W after the infection of the workplace.

Consider first the case of β_{LL} . A household of size n_H infects a workplace of size n_W at a rate $\gamma_H^{(n_H)}(\tau_H)\pi_{n_W}^W$, because the size of the workplaces where individuals living in that household work is independent of the size of the household. Then, for the same reason, the workplace infects households of size n'_H at a rate given by $\gamma_W^{(n_W)}(\tau_W)\pi_{n'_H}^H$. By considering the contribution of each possible workplace size, a household of size n_H infected locally

infects locally a household of size n'_H at an average rate of

$$\begin{aligned}\pi_{n'_H}^H \left(\sum_{n_W} \pi_{n_W}^W \gamma_H^{(n_H)} * \gamma_W^{(n_W)}(\tau_H) \right) &= \pi_{n'_H}^H \gamma_H^{(n_H)} * \left(\sum_{n_W} \pi_{n_W}^W \gamma_W^{(n_W)} \right) (\tau_H) \\ &= \pi_{n'_H}^H \gamma_H^{(n_H)} * \gamma_W(\tau_H),\end{aligned}$$

where $\gamma_W = \sum_{n_W} \pi_{n_W}^W \gamma_W^{(n_W)}$ is the average rate at which secondary cases are infected in the workplace of a randomly selected individual.

In the case of β_{LG} , consider a household of size n_H infected globally. The primary case is in a workplace of size n_W with probability $\pi_{n_W}^W$ and therefore a household of size n'_H is infected locally through this primary case at a rate

$$\pi_{n'_H}^H \sum_{n_W} \pi_{n_W}^W \gamma_W^{(n_W)}(\tau_W) = \pi_{n'_H}^H \gamma_W(\tau_H),$$

where γ_W is defined as before and $\tau_W = \tau_H$ because the two local structures are infected at the same time, i.e. the time of the infection through a global contact of the common primary case.

Any other household case is in a workplace of size n'_W with probability $\pi_{n'_W}^W$ and therefore the rate of infection of households of size n'_H through secondary cases is given, like before, by

$$\pi_{n'_H}^H \gamma_H^{(n_H)} * \left(\sum_{n_W} \pi_{n_W}^W \gamma_W^{(n_W)} \right) (\tau_H) = \pi_{n'_H}^H \gamma_H^{(n_H)} * \gamma_W(\tau_H).$$

The total rate obtained by adding the two contributions together is given by

$$\pi_{n'_H}^H \left(\gamma_W(\tau_H) + \gamma_H^{(n_H)} * \gamma_W(\tau_H) \right).$$

The real-time growth rate r is obtained by imposing a dominant eigenvalue 1 to the block operator (the index H has been dropped from the time τ)

$$K_r = \begin{pmatrix} \pi_{n'_H}^H \int_0^{+\infty} \gamma_H^{(n_H)} * \gamma_W(\tau) e^{-r\tau} d\tau & \pi_{n'_H}^H \int_0^{+\infty} \left(\gamma_W(\tau) + \gamma_H^{(n_H)} * \gamma_W(\tau) \right) e^{-r\tau} d\tau \\ \pi_{n'_H}^H \int_0^{+\infty} \left(\beta(\tau) + \gamma_H^{(n_H)} * \beta(\tau) \right) e^{-r\tau} d\tau & \pi_{n'_H}^H \int_0^{+\infty} \left(\beta(\tau) + \gamma_H^{(n_H)} * \beta(\tau) \right) e^{-r\tau} d\tau \end{pmatrix}$$

Note that each of the four blocks has the same dimension $N_H \times N_H$, where N_H represents the largest household size considered (assumed to be finite) and consists of a matrix with rank 1. Therefore K_r has rank 2. However, thanks to the particular structure of K_r , the computation of the dominant eigenvalue can be simplified according to the following result:

Theorem 5.4.1. *Consider m -dimensional real non-zero column vectors u, v and a, b, c, d and consider the $m \times m$ matrices*

$$A = ua^T, B = ub^T, C = vc^T, D = vd^T,$$

where T denotes the transposition. Finally consider the $2m \times 2m$ block matrix

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

Then the matrix M has only two non-zero eigenvalues λ_1 and λ_2 , which are the same eigenvalues of the 2×2 matrix

$$m = \begin{pmatrix} a^T u & b^T u \\ c^T v & d^T v \end{pmatrix},$$

whose elements are the only non-zero eigenvalues of the matrices A, B, C and D . In particular, the dominant eigenvalue of M and m is the same.

Proof. The proof proceeds by direct computation, based on the fact that, if λ is an eigenvalue of m with corresponding eigenvector $(x, y)^T, x, y \in \mathbb{R}$, then $(ux, vy)^T$ is an eigenvector of M corresponding to the same eigenvalue λ . $\square$

Theorem 5.4.1 implies that K_r can be reduced to the two dimensional case of fixed household and workplace sizes with

$$\begin{aligned}
\beta_{LL}(\tau_H) &= \sum_{n_H} \pi_{n_H}^H \gamma_H^{(n_H)} * \gamma_W(\tau_H) \\
&= \gamma_H * \gamma_W(\tau_H) \\
\beta_{LG}(\tau_H) &= \sum_{n_H} \pi_{n_H}^H \left(\gamma_W(\tau_H) + \gamma_H^{(n_H)} * \gamma_W(\tau_H) \right) \\
&= \sum_{n_H} \pi_{n_H}^H \gamma_W(\tau_H) + \sum_{n_H} \pi_{n_H}^H \gamma_H^{(n_H)} * \gamma_W(\tau_H) \\
&= \gamma_W(\tau_H) + \gamma_H * \gamma_W(\tau_H),
\end{aligned}$$

where $\gamma_H = \sum_{n_H} \pi_{n_H}^H \gamma_H^{(n_H)}$ represents the average rate of occurrence of new cases in the household of a randomly chosen individual and the last equality holds because $\sum_{n_H} \pi_{n_H}^H = 1$.

5.5 Exact dynamics in a small population

The important results concerning the epidemic final size heavily rely on the possibility of neglecting the temporal dynamics of the infection process (see Section 4.9). The full dynamics are in fact analytically intractable in general. However, the analytical framework of continuous-time Markov chains can be used in the particular case of a constant recovery rate, as described in this section. The following one presents instead a method suggested by Fraser (2007) to approximate the otherwise intractable dynamics when a model with non-random time-varying infectivity is used.

5.5.1 Continuous time Markov chains

In this section the definition and the basic properties of a continuous-time Markov chain are recalled (they can be found on any standard probability textbook; see, for example, Grimmett & Stirzaker, 2001). A direct application to epidemics is then considered.

Given a set Ω , finite or countable, of possible states of a system, a *continuous-time Markov chain* (CTMC) on Ω is a family $\mathcal{X} = \{X(t) : t \geq 0\}$ of random variables indexed

with $t \in [0, +\infty)$ which satisfies the following *Markov property*:

$$\mathbb{P}(X(t_n) = j \mid X(t_1) = i_1, X(t_2) = i_2, \dots, X(t_{n-1}) = i_{n-1}) = \mathbb{P}(X(t_n) = j \mid X(t_{n-1}) = i_{n-1}),$$

for all $i_1, i_2, \dots, i_{n-1}, j \in \Omega$ and any sequence $t_1 < t_2 < \dots < t_n$ of times.

The Markov property is a memoryless property, as it states that the probability that the system is in one state at a certain point in time, knowing where it was at a previous point in time, is not affected by the history preceding that time.

For the sake of simplicity, consider Ω finite. A continuous time Markov chain can be described through its *generator matrix*, or *Q-matrix*, a $|\Omega| \times |\Omega|$ matrix

$$Q = (q_{ij}),$$

where $q_{ij}, i \neq j$ represents the transition rate from i to j and $q_{ii} = -\sum_{j, j \neq i} q_{ij}$ (assumed to be finite $\forall i$) so that each row sums to 0.

The stochastic dynamics of the system are described by the family $\{P(t), t \geq 0\}$ defined through the matrix exponential

$$P(t) = e^{tQ}.$$

Each element $p_{ij}(t)$ represents the probability that the system is in state j , given that t units of time before it was in state i . Note that the Q -matrix has non-negative off-diagonal elements (i.e. $q_{ij} \geq 0, i \neq j$) and therefore $P(t) = e^{tQ}$ is non-negative (i.e. $p_{ij}(t) \geq 0, \forall i, j, t$)

5.5.2 Application to epidemics in a closed group

In the case of an epidemic, given that individuals are all assumed identical, that the total population is constant and that an individual can be susceptible (X), infected (Y) or recovered (Z), the state space Ω of the system is given by all the pairs (X, Y) where $X \geq 0, Y \geq 0$ and $X + Y \leq n$ (except for the state $(n, 0)$, in which all individuals are susceptible and there is no epidemic): Z is automatically given by $Z = n - X - Y$.

It is convenient to represent the state space on a lattice and highlight the possible transitions by adding arrows, labelled with the transition rates, as done in Figure 5.2. The only transitions allowed are summarised in Figure 5.1, where $\lambda = \beta/(n-1)$ represents the infective-to-susceptible infection rate and ν represents the constant recovery rate.

Event	State transition	Transition rate
Infection	$(i, j) \rightarrow (i-1, j+1)$	$\frac{\beta ij}{n-1} = \lambda ij$
Recovery	$(i, j) \rightarrow (i, j-1)$	νj

Figure 5.1: Table describing all the possible state transitions in the continuous-time Markov chain description of an epidemic, together with their rates of occurrence.

In order to write explicitly the generator matrix Q , it is necessary to specify a map that associates the pair (i, j) to a single index k , such that the state $\omega_k \in \Omega$ represents the epidemic state $(X, Y) = (i, j)$. The required map and its inverse are cumbersome to describe analytically, but they are nothing more than a way of counting all the states. The counting order is here chosen as in Figure 5.2 and the conversion is formally expressed by the maps:

$$k = \kappa_n((i, j))$$

$$(i, j) = \kappa_n^{-1}(k).$$

The element q_{hk} of the generator matrix represents the rate of progression from state ω_h to ω_k , i.e. from state $(i_h, j_h) = \kappa_n^{-1}(h)$ to state $(i_k, j_k) = \kappa_n^{-1}(k)$. The correspondent element $p_{hk}(t)$ of $P(t) = e^{tQ}$ represents the probability that the system is in state (i_k, j_k) , given that it was in state (i_h, j_h) t units of time before or, equivalently, that the system is in state (i_k, j_k) at time t if it started from state (i_h, j_h) .

Another quantity of interest is the average time spent by the system in each state by time t or during the entire epidemic. This information is contained in the matrix

$$M(t) = \int_0^t e^{\sigma Q} d\sigma,$$

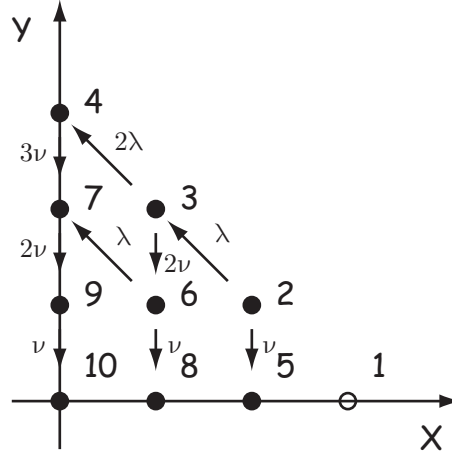


Figure 5.2: Graphical representation of all possible states that the system describing an *SIR* epidemic in a population of size $n = 3$ can assume. Each point (x, y) represents the state with x susceptibles, y infectives and $z = n - x - y$ immunes. States are also labelled sequentially with index $k = 1, \dots, 10$, to exemplify one possible definition of the conversion maps κ_3 and κ_3^{-1} (see text). The arrows represent all possible state transitions and are labelled with the correspondent rate of occurrence: diagonal arrows refer to infections while vertical ones refer to recoveries. λ denotes the one-to-one infection rate and ν the recovery rate. The states on the X -axes have no infectives and therefore are absorbing states. State 1, corresponding to all susceptible individuals, cannot occur during an epidemic.

where the integral is taken componentwise. In particular, each element $m_{hk}(t)$ of $M(t)$ represents the average time spent in state k by time t if the system started from state h . If an absorbing state is present in the state space Ω , then the average time spent in that state by time t will diverge as $t \rightarrow +\infty$.

5.5.3 Application to the households model

When the infectious profile of a household is considered, it is convenient to include among the possible states of the household only those ones where there is at least one infective, as shown in Figure 5.3. By doing so, the absorbing states (i.e. the states on the x -axis) are excluded from the computation and therefore it is also informative to consider the average time spent in each possible state throughout the household epidemic. In this case it is assumed improperly that the matrix Q contains also the rates at which the household fully

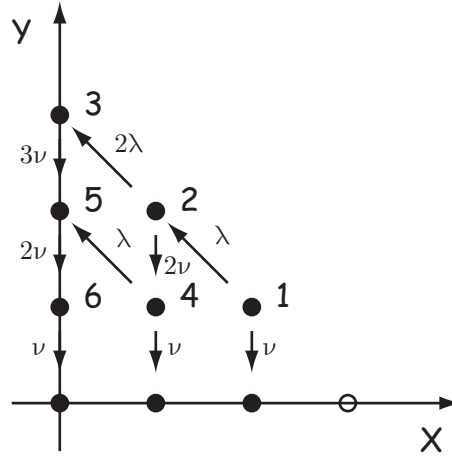


Figure 5.3: Graphical representation of all possible states that the system describing an *SIR* epidemic in a population of size $n = 3$ can assume. The description is the same as that of Figure 5.2, but here only the “active” states, i.e. the states with at least one infective, are considered. If the group represents a household, these states are the only ones during which the household is infectious. As these states are all transient (i.e. none of them is an absorbing state), the average time spent in each of these states is finite.

recovers. More rigorously, Q should be seen as $Q = \Sigma - \Delta$, where Σ is the actual generator matrix of the Markov chain, and Δ is a diagonal “death” matrix (“death” in the sense of ending the household infectious life).

The limit of $M(\tau_H)$ as $\tau_H \rightarrow +\infty$ exists finite if and only if the integral converges, which is equivalent (see Diekmann & Heesterbeek, 2000) to the spectral bound of Q being negative or, equivalently, to Q being invertible and $-Q^{-1}$ being non-negative (i.e. with all non-negative elements). (Recall that the spectral bound of Q is defined as the $\sup\{\text{Re}(\vartheta) : \vartheta \text{ is an eigenvalue of } Q\}$). If this is the case, then

$$M := \lim_{\tau_H \rightarrow +\infty} M(\tau_H) = -Q^{-1}.$$

The element m_{hk} of M therefore describes the average time spent in state k if the system started from state h . Since the total rate of global infections when the system is in state k is given by βi_k , the total infectivity of the household can be easily computed by summing across all states:

$$R_* = \sum_k \beta i_k (-Q^{-1})_{1k},$$

where the row index 1 is fixed because the state $k = 1$, corresponding to $(X, Y) = (n-1, 1)$, is the one from which a household starts when infected from outside.

A similar approach can be used when the real-time growth rate r is studied. In fact, the infectious profile of the household at time τ_H after its infection is given by

$$\beta_H(\tau_H) = \sum_k \beta i_k P_{1k}(\tau_H) = \sum_k \beta i_k (e^{Q\tau_H})_{1k}.$$

The real-time growth rate r is such that $L_{\beta_H}(r) = 1$, where now

$$\begin{aligned} L_{\beta_H}(r) &= \int_0^{+\infty} \beta_H(\tau_H) e^{-r\tau_H} d\tau_H \\ &= \int_0^{+\infty} \left(\sum_k \beta i_k P_{1k}(\tau_H) \right) e^{-r\tau_H} d\tau_H \\ &= \sum_k \beta i_k \int_0^{+\infty} (e^{Q\tau_H})_{1k} e^{-r\tau_H} d\tau_H \\ &= \sum_k \beta i_k \int_0^{+\infty} (e^{Q\tau_H} e^{-rJ\tau_H})_{1k} d\tau_H \\ &= \sum_k \beta i_k \left(\int_0^{+\infty} e^{(Q-rJ)\tau_H} d\tau_H \right)_{1k} \\ &= \sum_k \beta i_k \left(-(Q-rJ)^{-1} \right)_{1k}, \end{aligned}$$

where J represents the identity matrix and it has been used for $e^{-rJ\tau}$ the fact that for a diagonal matrix

$$\Delta = \begin{pmatrix} \delta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \delta_n \end{pmatrix},$$

the exponential is the diagonal matrix given by

$$e^\Delta = \begin{pmatrix} e^{\delta_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{\delta_n} \end{pmatrix}.$$

An iterative method can now be used to solve Equation (5.3.1) for r (the solution exists and is unique, see Section 2.2.3). Each step requires only that a matrix is inverted, instead of the inefficient computation of integrals and matrix exponentials.

5.5.4 Application to the households-workplaces model

The computation for the real time growth rate can be extended to the model with households and workplaces: however, in this case, not only the rate of global infections is required, but also the rates γ_H and γ_W at which new cases appear in households and workplaces.

Note that, if the system is in state (i_k, j_k) at time t , the jump into that state has occurred before time t . In order to calculate the rate at which a new case occurs, it is necessary to consider the rate of reaching each state that has more than one infective. The rate at time t at which the system reaches state (i_k, j_k) because of an infection event is given by

$$\lambda(i_k + 1)(j_k - 1)p_{\kappa_n(i_k+1, j_k-1)}(t),$$

where $p_{\kappa_n(i_k+1, j_k-1)}(t)$, despite the cumbersome notation, simply represents the probability that at time t there is one susceptible more and one infective less than what required and this quantity is multiplied by the rate of occurrence of an infection. Analogously, although not needed in this case, the rate of reaching the same state (i_k, j_k) at time t because of a recovery, is given by

$$\nu(j_k + 1)p_{\kappa_n(i_k, j_k+1)}(t).$$

Refer again to Figure 5.3 for the within-group transition rates.

Following this argument, γ_H is constructed as

$$\gamma_H(\tau_H) = \sum_k \lambda_h s_k i_k p_{1k}(\tau_H),$$

where λ_h denotes the infectious-to-susceptible infection rate within the household. An analogous definition holds for γ_W .

Each element of the operator K_r defined in (5.4.1) needs to be computed. As before, the elements containing β_{GG} and β_{GL} do not need particular attention as their construction is the same of that for the households model. The term containing β_{LL} is elaborated as

$$\begin{aligned}
\int_0^{+\infty} \beta_{LL}(\tau_H) e^{-r\tau_H} d\tau_H &= \int_0^{+\infty} \gamma_H * \gamma_W(\tau_H) e^{-r\tau_H} d\tau_H \\
&= \left[\int_0^{+\infty} \gamma_H(\tau_H) e^{-r\tau_H} d\tau_H \right] \left[\int_0^{+\infty} \gamma_W(\tau_H) e^{-r\tau_H} d\tau_H \right] \\
&= \left[\int_0^{+\infty} \left(\sum_h \lambda_h s_h i_h P_{1h}^H(\tau_H) \right) e^{-r\tau_H} d\tau_H \right] \left[\int_0^{+\infty} \left(\sum_k \lambda_w s_k i_k P_{1k}^W(\tau_H) \right) e^{-r\tau_H} d\tau_H \right] \\
&= \left[\sum_h \lambda_h s_h i_h \left(\int_0^{+\infty} e^{(Q_H - rJ)\tau_H} d\tau_H \right)_{1h} \right] \left[\sum_k \lambda_w s_k i_k \left(\int_0^{+\infty} e^{(Q_W - rJ)\tau_H} d\tau_H \right)_{1k} \right] \\
&= \left[\sum_h \lambda_h s_h i_h ((rJ - Q_H)^{-1})_{1h} \right] \left[\sum_k \lambda_w s_k i_k ((rJ - Q_W)^{-1})_{1k} \right],
\end{aligned}$$

where $\lambda_h, \lambda_w, P^H, P^W, Q_H$ and Q_W are, respectively, the infective-to-susceptible infection rates, the probabilities of being in each state of the local epidemic and the Q -matrices for households and workplaces. The second line follows from the first because of the well-known property that the Laplace transform of a convolution is equal to the product of the Laplace transforms of the factors.

The term containing β_{GL} needs just the addition of the contribution of another workplace epidemic and is therefore given by

$$\int_0^{+\infty} \beta_{LG}(\tau_H) e^{-r\tau_H} d\tau_H = \left[1 + \sum_h \lambda_h s_h i_h ((rJ - Q_H)^{-1})_{1h} \right] \left[\sum_k \lambda_w s_k i_k ((rJ - Q_W)^{-1})_{1k} \right].$$

Finally, Equation (5.4.2) is solved iteratively for r .

This computational method provides the exact numerical value for the real-time growth rate r because the local saturation of susceptibles is correctly taken into account. However, the method is limited to the Markovian case of constant recovery rate. In the next section, an approximate method is described for the case of a non-random time-varying disease history. This method has been first suggested by Fraser (2007).

5.6 Approximate dynamics in a small population

The exact dynamics for the spread of an infection in a homogeneously mixing population described in the previous section are limited to the Markovian case. However, for models characterised by a non-random time varying infectivity profile, an approximated method for dealing with the otherwise intractable dynamics has been proposed in Fraser (2007). The author applies the method directly to a households model with the primary concern of estimating the reproduction number R_* for households, given the real-time growth rate r , the generation time distribution $\omega(\tau)$ and an estimation of the infectivity within households. I will show in Section 7.1 how such a method can be extended to the households-workplaces model and applied to the case of influenza to estimate R_H . Instead, here the same ideas are presented as a method for the estimation of the real-time growth rate under the assumption that all other parameters are known.

It is assumed that the generation time distribution $\omega(\tau)$ is the same for every individual and that the infectivity profiles within households, within workplaces or in the community are all proportional to $\omega(\tau)$. In other words, at time τ after the infection, an infective makes infectious contact with all other individuals (one-to-all) globally at rate $\beta(\tau) = R_g\omega(\tau)$, in the household at rate $\beta_h(\tau) = R_h\omega(\tau)$ and in the workplace at rate $\beta_w(\tau) = R_w\omega(\tau)$, where R_g, R_h and R_w therefore represent the average number of infections contacts and infective makes (i.e. the average number of infections an infective would cause if it was not for the depletion of susceptibles) in the community, household and workplace respectively.

The method should be easily extendable to the case of random total infectivities with averages R_g, R_h and R_w , of random generation-time distribution with average $\omega(\tau)$ and of different generation time distributions in different environments. These extensions, however, are not presented here and will be the focus of future research.

5.6.1 Epidemic in a closed group

Consider a group of size 2 and assume one individual is initially infectious. Then the time distribution of infectious contacts towards the other case is given by $\omega(\tau)$. However,

because repeated contacts occur towards the same individual and only the first one results in an infection, the rate at which the second case is infected should take into account the probability of the initial susceptible avoiding the infection over time (*surviving probability*). The idea proposed by Fraser (2007) is that the overall rate of infection can be approximated by the rate of infectious contacts $\omega(\tau)$, thus ignoring the contribution of the surviving probability, i.e. ignoring the presence of repeated contacts between the same individuals.

Such an approximation may appear as quite crude when only two individuals are considered. But when the population size n is larger than 2, the probability that multiple contacts occur between the same pair of individuals is smaller. Furthermore, a lower infection rate implies fewer infectious contacts and a lower probability of repeated contacts within the same pair. Therefore, the approximation is expected to improve as n increases and the infectivity parameter decreases (see Section 5.7).

However, when $n > 2$, various chains of infection can occur: Fraser (2007) proposes to use the Reed-Frost model to compute the average number of cases in each generation and to assume that each generation infects the cases in the following generation according to a time distribution given by $\omega(\tau)$. Therefore, given that the initial infective constitutes generation 1, generation 2 contains an average of μ_2 cases, which occur at times distributed as $\omega(\tau)$, generation 3 contains an average of μ_3 cases, which occur at times distributed as $\omega_{[2]}(\tau) = \omega * \omega(\tau)$, and so on. Here, $\omega_{[i]}(\tau)$ represents the i -th autoconvolution of ω , with $\omega_{[1]} = \omega$, and μ_i represents the average number of cases in generation i , computed using the Reed-Frost model. The maximum number of generations is n .

Note that, apart from ignoring repeated contacts between the same individuals, the method ignores the fact that generations can overlap, when an individual counted in generation i in the Reed-Frost model is actually infected by other cases belonging to later generations. In the terminology of Ludwig's construction of Chapter 3, this means that the real-time generation of each individual is approximated by his or her rank, and therefore that real cases tend to occur somewhat earlier than what is described by the approximation (e.g. see Figure 2 of Fraser, 2007).

In summary, the overall average rate $\gamma(\tau)$ at which new cases appear in the population

is approximated by

$$\gamma_{max}(\tau) = \sum_{i=2}^n \mu_i \omega_{[i-1]}(\tau). \quad (5.6.1)$$

Both elements involved in the approximation, namely neglecting overlapping generations and repeated contacts between the same individuals, cause the approximation to overestimate the time of occurrence of new cases.

In order to provide also an underestimate of those times, Fraser (2007) suggested to lump all cases, except from the initial one, in the second generation. This gives rise to the other approximation:

$$\gamma_{min}(\tau) = \omega(\tau) + \mu \omega * \omega(\tau), \quad (5.6.2)$$

where $\mu = \sum_{i=2}^n \mu_i$ represents the average number of initial susceptibles who are ultimately infected. The rate γ_{min} does not represent an actual lower bound approximation in general, as it still ignores the effect of overlapping generations and repeated contacts. For example, in the extreme case of two individuals, $\gamma_{min} = \gamma_{max}$ and they both overestimates the times of infection of new cases. Most of the time, however, the rate γ_{min} underestimates those times, thus giving rise to a faster epidemic but with the same final size $1 + \mu$. The rate γ_{min} represents a good approximation to the true rate γ when only a few generations occur, i.e. when the group size is small (but not too small) and the infectivity is high (e.g. as in the case of application to measles in Fraser, 2007).

5.6.2 Application to the households model

In the households model, given that each individual infects outside at a rate $\beta(\tau) = R_g \omega(\tau)$, the approximated infectiousness profile of a household of size n is given by

$$\begin{aligned} \beta_H(\tau_H) &= \beta(\tau_H) + \beta * \gamma_H(\tau_H) \\ &\approx R_g \omega(\tau_H) + R_g \omega * \left(\sum_{i=2}^n \mu_i \omega_{[i-1]} \right) (\tau_H) \\ &= R_g \sum_{i=1}^n \mu_i \omega_{[i]}(\tau_H). \end{aligned}$$

The last equality follows from the fact that $\mu_1 = 1$ and $\omega_{[1]} = \omega$.

The real-time growth rate r is then given by the implicit solution of

$$\begin{aligned} 1 &= \int_0^{+\infty} \beta_H(\tau_H) e^{-r\tau_H} d\tau_H \\ &\approx \int_0^{+\infty} \left(R_g \sum_{i=1}^n \mu_i \omega_{[i]}(\tau_H) \right) e^{-r\tau_H} d\tau_H \\ &= R_g \sum_{i=1}^n \mu_i \left(\int_0^{+\infty} \omega_{[i]}(\tau_H) e^{-r\tau_H} d\tau_H \right), \end{aligned}$$

where the third line follows from the linearity of the integral. A simplification is obtained by following Fraser (2007) and noting that, defining

$$\int_0^{+\infty} \omega(\tau) e^{-r\tau} d\tau = \frac{1}{R},$$

where R represents the basic reproduction number that would be related to r if the households structure were ignored, the properties of the Laplace transform imply that

$$\begin{aligned} \int_0^{+\infty} \omega_{[i]}(\tau_H) e^{-r\tau_H} d\tau &= \left(\int_0^{+\infty} \omega(\tau) e^{-r\tau} d\tau \right)^i \\ &= \frac{1}{R^i}. \end{aligned}$$

Therefore, the approximated real-time growth rate for the households model is given by the equation

$$R_g \sum_{i=1}^n \frac{\mu_i}{R^i} = 1,$$

where, as defined above, $R = \left(\int_0^{+\infty} \omega(\tau) e^{-r\tau} d\tau \right)^{-1}$.

The extension to a distribution of household sizes is straightforward.

5.6.3 Application to the households-workplaces model

Consider the households-workplaces model and assume, for the time being, that all households and all workplaces have the same sizes n_H and n_W , respectively.

The real-time growth rate is then obtained by requiring that the operator

$$K_r = \begin{pmatrix} \int_0^{+\infty} \beta_{LL}(\tau_H) e^{-r\tau_H} d\tau_H & \int_0^{+\infty} \beta_{LG}(\tau_H) e^{-r\tau_H} d\tau_H \\ \int_0^{+\infty} \beta_{GL}(\tau_H) e^{-r\tau_H} d\tau_H & \int_0^{+\infty} \beta_{GG}(\tau_H) e^{-r\tau_H} d\tau_H \end{pmatrix}$$

has dominant eigenvalue 1. Note that, for $r = 0$, $\rho(K_0) = R_H$. A large epidemic is considered and it is therefore assumed $R_H > 1$.

With the same construction as before, the following approximations are considered:

$$\begin{aligned} \gamma_H(\tau_H) &\approx \sum_{i=2}^{n_H} \mu_i^H \omega_{[i-1]}(\tau_H) \\ \gamma_W(\tau_W) &\approx \sum_{j=2}^{n_W} \mu_j^W \omega_{[j-1]}(\tau_W), \end{aligned}$$

where γ_H and γ_W represent the rates at which new cases appear in households and workplaces, τ_H and τ_W represent the time since infection of the household or the workplace and μ_i^H and μ_j^W represents the average number of cases in generation i in a Reed-Frost household epidemic and in generation j in a Reed-Frost workplace epidemic, respectively.

Consider the “LL” term of K_r :

$$\begin{aligned} \int_0^{+\infty} \beta_{LL}(\tau_H) e^{-r\tau_H} d\tau_H &= \int_0^{+\infty} \gamma_W * \gamma_H(\tau_H) e^{-r\tau_H} d\tau_H \\ &= \left(\sum_{i=2}^{n_H} \mu_i^H \int_0^{+\infty} \omega_{[i-1]}(\tau_H) e^{-r\tau_H} d\tau_H \right) \left(\sum_{j=2}^{n_W} \mu_j^W \int_0^{+\infty} \omega_{[j-1]}(\tau_W) e^{-r\tau_W} d\tau_W \right). \end{aligned}$$

Defining

$$R = \left(\int_0^{+\infty} \omega(\tau) e^{-r\tau} d\tau \right)^{-1}$$

as before, the “LL” term simplifies in

$$\int_0^{+\infty} \beta_{LL}(\tau_H) e^{-r\tau_H} d\tau_H \approx \left(\sum_{i=2}^{n_H} \frac{\mu_i^H}{R^{i-1}} \right) \left(\sum_{j=2}^{n_W} \frac{\mu_j^W}{R^{j-1}} \right).$$

With similar arguments, and recalling that, when a case infected globally, the epidemics in the household and the workplace start simultaneously and therefore $\tau_H = \tau_W$ in this case,

$$\begin{aligned}
\int_0^{+\infty} \beta_{LG}(\tau_H) e^{-r\tau_H} d\tau_H &= \int_0^{+\infty} (\gamma_W(\tau_H) + \gamma_W * \gamma_H(\tau_H)) e^{-r\tau_H} d\tau_H \\
&\approx \left(1 + \sum_{i=2}^{n_H} \frac{\mu_i^H}{R^{i-1}} \right) \left(\sum_{j=2}^{n_W} \frac{\mu_j^W}{R^{j-1}} \right), \\
\int_0^{+\infty} \beta_{GL}(\tau_H) e^{-r\tau_H} d\tau_H &= \int_0^{+\infty} \beta_{GG}(\tau_H) e^{-r\tau_H} d\tau_H \\
&= \int_0^{+\infty} (\beta(\tau_H) + \beta * \gamma_H(\tau_H)) e^{-r\tau_H} d\tau_H \\
&\approx \frac{R_g}{R} + \frac{R_g}{R} \left(\sum_{i=2}^{n_H} \frac{\mu_i^H}{R^{i-1}} \right) \\
&= R_g \left(\sum_{i=1}^{n_H} \frac{\mu_i^H}{R^i} \right).
\end{aligned}$$

The last equality follows from the fact that $\mu_1^H = 1$.

Therefore

$$K_r = \begin{pmatrix} \left(\sum_{i=2}^{n_H} \frac{\mu_i^H}{R^{i-1}} \right) \left(\sum_{j=2}^{n_W} \frac{\mu_j^W}{R^{j-1}} \right) & \left(1 + \sum_{i=2}^{n_H} \frac{\mu_i^H}{R^{i-1}} \right) \left(\sum_{j=2}^{n_W} \frac{\mu_j^W}{R^{j-1}} \right) \\ R_g \left(\sum_{i=1}^{n_H} \frac{\mu_i^H}{R^i} \right) & R_g \left(\sum_{i=1}^{n_H} \frac{\mu_i^H}{R^i} \right) \end{pmatrix}$$

from which r can be computed numerically as the solution of $\rho(K_r) = 1$.

5.7 Numerical investigation

The main ingredient used in dealing with the real-time growth rate when local saturation effects cannot be neglected is the function γ , defined as the average rate at which a new infection occurs in a small group. It is worth analysing this function in detail.

It follows directly from its definition that, if $X(t)$ denotes the average number of susceptibles and $I_C(t)$ the average cumulative incidence of an *SIR* epidemic in a homogeneously mixing population at time t ,

$$\gamma(t) = \frac{d}{dt}I_C(t) = -\frac{d}{dt}X(t). \quad (5.7.1)$$

The first equality states that the rate at which new cases appear is the same rate at which the cumulative incidence increases, which is a trivial statement (although formal details involve the exchange between the derivative and the integral defining the average across possible outcomes of the epidemic process). The second equation follows directly from the fact that the total population is constant.

This characterisation, however, is not particularly explicit. Therefore it is worth exploring numerically this complex model component γ , and in particular its analytical approximations presented in Section 5.6.1.

First, consider a group of size 2. It is worth starting from the Markovian case obtained when there is a constant recovery rate ν , as in this case the exact dynamics can be expressed analytically. Strictly speaking, in this case, the approximate method does not apply, as it is defined for a non-random disease history. The method, however, can be easily extended to random disease histories, by computing the average number of cases in each generation using the randomised Reed-Frost model, instead of the simple Reed-Frost model. In this particularly simple case, if λ denotes the one-to-one infection rate and ν the recovery rate, the average number of cases in generation 2 is given by $\lambda/(\lambda + \nu)$, i.e. the probability that the second case is infected. The rate of infection of the secondary case is given by $\gamma(\tau) = \lambda e^{-(\lambda + \nu)\tau}$ (the first of the two events, infection or recovery, occurs at a rate $\lambda + \nu$), while the rate at which the primary case makes infectious contacts with the other case is given by $\omega(\tau) = \nu e^{-\nu\tau}$.

Therefore, the approximation of Fraser (2007) to the rate $\gamma(\tau)$ is

$$\gamma_{max}(\tau) = \frac{\lambda\nu}{\lambda + \nu} e^{-\nu\tau}. \quad (5.7.2)$$

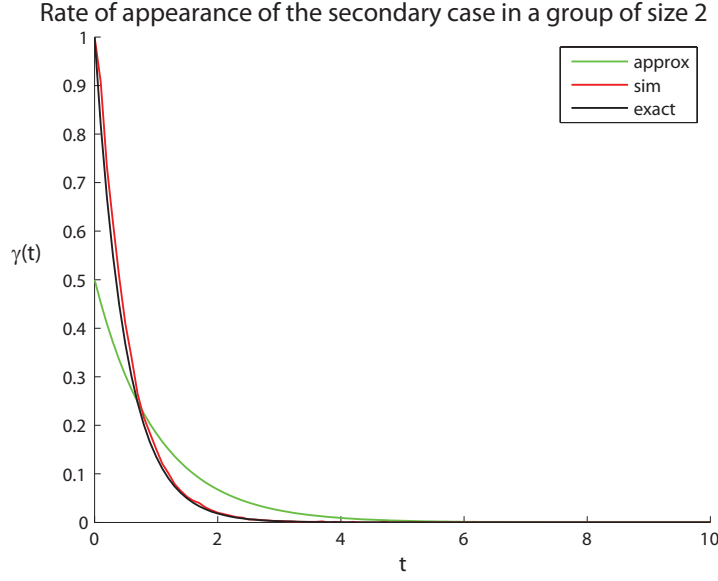


Figure 5.4: Plot of the rate at which the initial susceptible in a group of size $n = 2$ acquires infection from the other case. The exact rate $\gamma(\tau)$ (black) is compared with the rate obtained by averaging across 100 000 stochastic simulations (red) and the analytical approximation (green) described in Equation (5.7.2). The disease history corresponds to the Markovian case (exponentially distributed length of infectious period) with constant infectivity $\lambda = 1$ and recovery rate $\nu = 1$.

As it should be, it is easily verified that

$$\int_0^{+\infty} \gamma(\tau) d\tau = \int_0^{+\infty} \gamma_{max}(\tau) d\tau = \frac{\lambda}{\lambda + \nu} = \mu_2 = \mu.$$

Figure 5.4 shows the true rate γ in black and γ_{max} in green, together with the computation of $\gamma(t) = \frac{d}{dt} I_C(t)$ in red, where $I_C(t)$ is obtained by averaging across 100,000 individual-based stochastic simulations.

Note how Fraser's approximation performs poorly in this case, due to the fact that it neglects repeated contacts between the individuals.

When a non-random time-varying infectivity is considered, no exact analytical result is available and therefore the comparison of the approximate rate γ_{max} is made with the average of 100 000 individual-based stochastic simulations (see Section 6.1).

Figures 5.5 and 5.6 show the analytical approximation (green) and the numerical simulation (red) for groups of size $n = 2, 5$ and 10, an average one-to-all total infectivity of

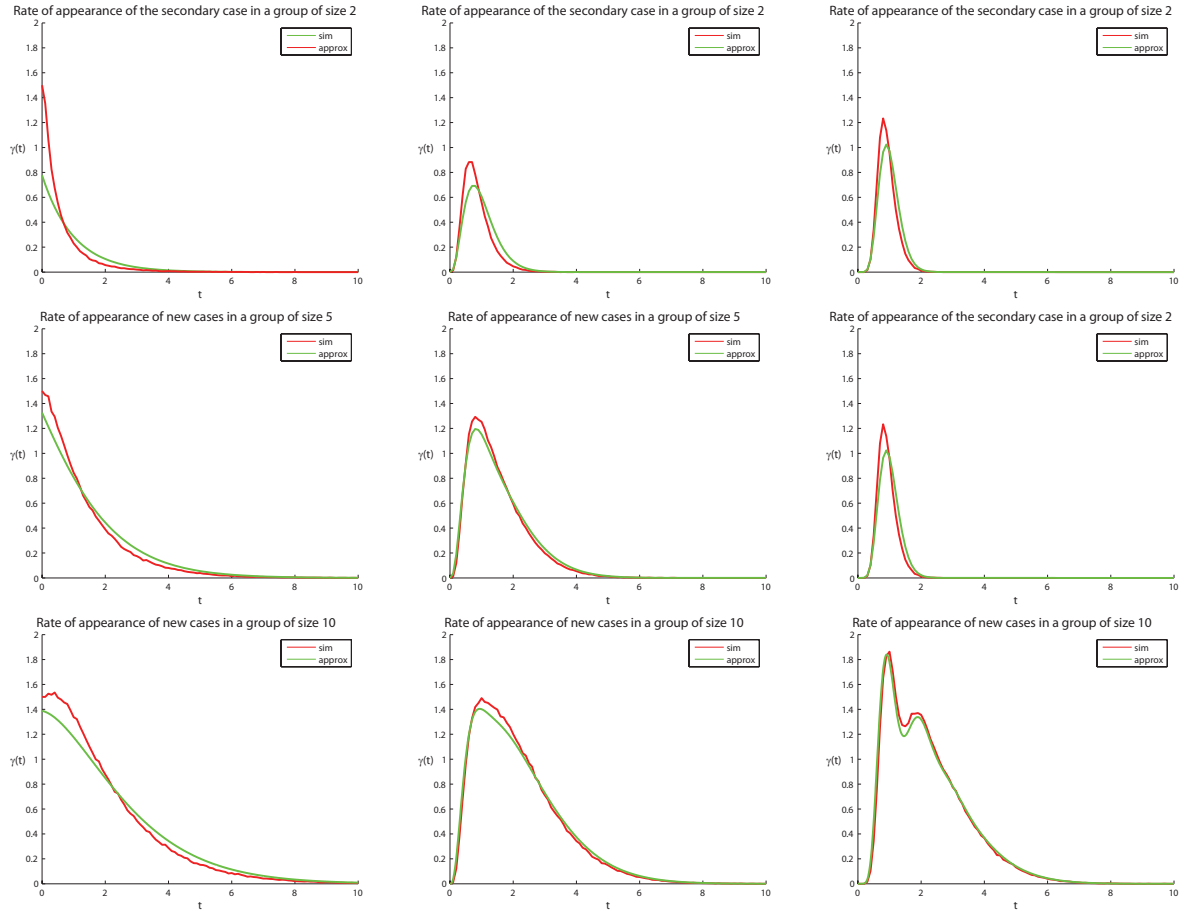


Figure 5.5: Plots of the rate at which new cases occur in a household epidemic started from a single initial case. In this case, $R_0 = 1.5$ and the generation time distribution is given by $\omega(\tau) \sim \Gamma(\alpha, \delta)$ with generation time $T_g = 1$. The analytical approximation (green) defined in Equation (5.6.1) is compared with the average over 100 000 stochastic simulation (red). The group size changes by rows: from top to bottom, $n = 2, 5$ and 10 . The shape of the infectivity profile changes by columns: from left to right, $\alpha = 1$ (exponential), 4 and 10 . The scale parameter δ changes accordingly.

$R_0 = 1.5$ and a $\Gamma(\alpha, \delta)$ -distributed generation time distribution $\omega(\tau)$, with shape parameter $\alpha = 1$ (exponential), 4 and 10 and fixed generation time $T_g = \alpha/\delta = 1$. As can be seen, the approximation does not appear to improve significantly as the group size grows, but it improves considerably with increasing α , i.e. when secondary cases of an infective are more peaked around the generation time T_g .

When R_0 within the group has value 1.5 , then each individual on average makes 1.5

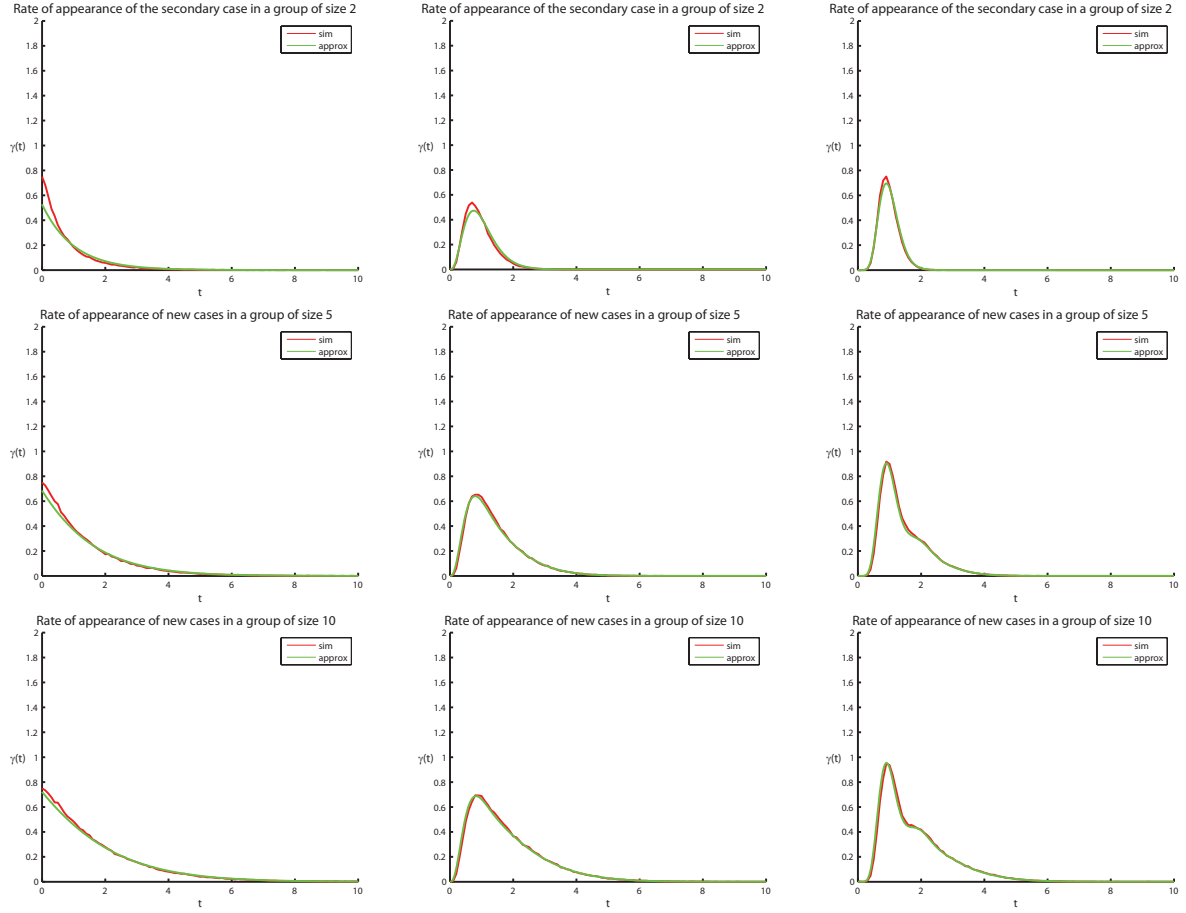


Figure 5.6: Plots of the rate at which new cases occur in a household epidemic started from a single initial case. All details are the same as in Figure 5.5, except that now $R_0 = 0.75$.

infectious contacts with other cases. However, when R_0 decreases, so does the probability of repeated contacts towards the same individuals. In fact the approximation dramatically improves as R_0 decreases: the same situations described above are reported in Figures 5.6 for $R_0 = 0.75$.

Note however that, even when the approximation appears not to be particularly accurate, as in the case of $R_0 = 1.5$ above, when the method is used to estimate the real-time growth rate in a structured model, the average infectious profile of the group (e.g. a household) is multiplied by e^{-rt} and therefore, for positive r , the discrepancies between the approximation and the true rate $\gamma(t)$ that occur for large t are weighted much less

than those occurring for small t . Estimates of real-time growth rates are therefore more accurate when the generation time is Γ -distributed with shape parameter $\alpha > 1$, i.e. when the rate of appearance of new cases at the very start of the household epidemic is negligible and thus the discrepancy between the real and the approximate rate of appearance of new cases is minimal for small t (as opposed to the exponentially decaying infectivity obtained for $\alpha = 1$, see Figures 5.5 and 5.6). Such a consideration offers an explanation for the surprisingly strong predictive power of this method highlighted in Fraser (2007).

5.8 Factorisation of the global infectivity profile of a group

Throughout the chapter, a factorisation of the global infectivity profile of a group in terms of the internal dynamics, described by the rate γ , and the global infectivity that each individual exerts outside the group, has been used. This section offers a formal proof of such a factorisation.

It has been noted already that, under suitable conditions, the total infectivity of a household R_* can be factorised in the form

$$R_* = R_g(1 + \mu_H),$$

where R_g represents the average global infectivity of an infected individual and μ_H represents the average number of initial susceptibles that are ultimately infected in a household epidemic started by a single initial case.

The result is known as the Wald's identity for epidemics (after the general identity appearing in Wald, 1947, of which it is a particular case; see also Section 4.9, Ball, 1986, or Andersson & Britton, 2000) and heavily relies on the same assumptions that allow Ludwig's construction of Chapter 3 to be performed. For a proof of it, see Section 4.9, Ball (1986) or Andersson & Britton (2000).

Wald's identity for epidemics concerns the total global infectivity of a household. However, for the argument proposed in this chapter, a similar result is needed for the infection

rates at each point in time. This is provided by the following

Theorem 5.8.1. *Consider a household of n individuals that mix homogeneously and consider a stochastic epidemic spreading from a single initial infective. Assume that infective i has a global time-varying infectivity profile $B_i(\tau)$ distributed according to a random trajectory $B(\tau)$, drawn at the time of infection independently of the time at which the infection occurs and of the identity of the infector, and then evolving deterministically in time-since-infection τ . Assume that, at time τ after the infection, individual i makes global infectious contact with any other individual according to a Poisson process with rate $B_i(\tau)$ and that the within-group infectious contacts occur at the points of a Poisson process with a rate proportional to $B_i(\tau)$. The infectivity profiles of different individuals are mutually independent. Finally, assume that $\beta(\tau)$ represents the average of $B(\tau)$ and that the average rate at which new household cases occur is given by $\gamma_H(\tau)$. Then, at each point in time, the overall average infectivity profile of the entire household $\beta_H(\tau)$ can be factorised as*

$$\beta_H(\tau) = \beta(\tau) + \beta * \gamma_H(\tau).$$

Proof. The contribution of the first individual does not represent a problem and can be added at the last stage. Therefore, the focus will be on the convolution term, which will be denoted by the random variable $G_H(\tau)$. Now, let U be a random vector, containing all information describing the outcome of the epidemic. The interest is then focused, $\forall \tau$, on the average $\mathbb{E}_U[G_H(\tau)]$.

First, consider the case where $B(\tau)$ is non-random and equal to $\beta(\tau)$ for every infective. If there are only two individuals in the household, then $U = (T_2)$, where T_2 represents the random time at which individual 2 is infected. Therefore:

$$\mathbb{E}_{T_2}[G_H(\tau)] = \int_0^{+\infty} \beta(\tau - s_2) f_{T_2}(s_2) ds_2 = \beta * f_{T_2}(\tau),$$

where $f_{T_2}(s_2)$ is the probability density of individual 2 getting infected at time s_2 .

Consider now the case of three individuals. As the two initial susceptibles are indistinguishable, instead of working with the times T_2 and T_3 of infection of individuals 2 and

3, it is convenient to work with the times $T_{(2)}$ and $T_{(3)}$ at which the second and the third case occur. Then $U = (T_{(2)}, T_{(3)})$ and, given that the joint density can be expressed as

$$f_U(s_{(2)}, s_{(3)}) = f_{T_{(3)}|T_{(2)}}(s_{(3)}|s_{(2)})f_{T_{(2)}}(s_{(2)}),$$

at each time τ ,

$$\begin{aligned} \mathbb{E}_U[G_H(\tau)] &= \int_0^{+\infty} \int_0^{+\infty} [\beta(\tau - s_{(2)}) + \beta(\tau - s_{(3)})] f_U(s_{(2)}, s_{(3)}) ds_{(2)} ds_{(3)} \\ &= \int_0^{+\infty} \beta(\tau - s_{(2)}) f_{T_{(2)}}(s_{(2)}) ds_{(2)} + \int_0^{+\infty} \int_0^{+\infty} \beta(\tau - s_{(3)}) f_{T_{(3)}|T_{(2)}}(s_{(3)}|s_{(2)}) f_{T_{(2)}}(s_{(2)}) ds_{(3)} ds_{(2)} \\ &= \int_0^{+\infty} \beta(\tau - s_{(2)}) f_{T_{(2)}}(s_{(2)}) ds_{(2)} + \int_0^{+\infty} \beta(\tau - s_{(3)}) f_{T_{(3)}}(s_{(3)}) ds_{(3)} \\ &= \int_0^{+\infty} \beta(\tau - s) \left(f_{T_{(2)}}(s) + f_{T_{(3)}}(s) \right) ds \\ &= \beta * \gamma_H(\tau), \end{aligned}$$

where $f_{T_{(3)}}$ represents the marginal distribution of the third individual getting infected during the epidemic, and $\gamma_H = f_{T_{(2)}} + f_{T_{(3)}}$.

Analogously, in the case of n household members, the result holds with

$$\gamma_H = \sum_{i=2}^n f_{T_{(i)}}.$$

Consider now the extension to the case of random infectivity profile.

Let $b_i(\tau)$ be a realisation of the random variable $B_i(\tau)$, describing the infectivity profile of individual i and distributed, independently for each individual, according to the random trajectory $B(\tau)$ defined of a suitable space $\mathcal{B}$ with measure ζ and with time-point average

$$\beta(\tau) = \mathbb{E}_B[B(\tau)] = \int_{\mathcal{B}} b(\tau) \zeta(db).$$

Assume, for notational convenience, that the random variable $B(\tau)$ can be expressed by a density f_B . This is not necessary in general.

In the case of two individuals, $U = (B_1, T_2, B_2)$. Note that, in the joint distribution

$$f_U(b_1, s_2, b_2) = f_{B_1}(b_1)f_{T_2|B_1}(s_2|b_1)f_{B_2|B_1, T_2}(b_2|b_1, s_2),$$

the factor $f_{B_2|B_1, T_2}(b_2|b_1, s_2)$ is in fact simply $f_{B_2}(b_2)$, because of the key assumption that the particular infectivity profile can be drawn independently of the time of the infection and the identity (i.e. of the characteristics) of the infector.

Therefore, $\forall \tau$,

$$\begin{aligned} \mathbb{E}_U[G_H(\tau)] &= \int_0^{+\infty} \int_{\mathcal{B}} \int_{\mathcal{B}} b_2(\tau - s_2) f_{B_1, T_2, B_2}(b_1, s_2, b_2) db_1 db_2 ds_2 \\ &= \int_0^{+\infty} \int_{\mathcal{B}} \int_{\mathcal{B}} b_2(\tau - s_2) f_{B_1}(b_1) f_{T_2|B_1}(s_2|b_1) f_{B_2}(b_2) db_1 db_2 ds_2 \\ &= \int_0^{+\infty} \int_{\mathcal{B}} b_2(\tau - s_2) \left(\int_{\mathcal{B}} f_{T_2|B_1}(s_2|b_1) f_{B_1}(b_1) db_1 \right) f_{B_2}(b_2) db_2 ds_2 \\ &= \int_0^{+\infty} \int_{\mathcal{B}} b_2(\tau - s_2) f_{T_2}(s_2) f_{B_2}(b_2) db_2 ds_2 \\ &= \int_0^{+\infty} \left(\int_{\mathcal{B}} b_2(\tau - s_2) f_{B_2}(b_2) db_2 \right) f_{T_2}(s_2) ds_2 \\ &= \int_0^{+\infty} \beta(\tau - s_2) f_{T_2}(s_2) ds_2 \\ &= \beta * \gamma_H(\tau). \end{aligned}$$

The extension to more than two individual is omitted, but follows the same argument.

Again,

$$\gamma_H = \sum_{i=2}^n f_{T_{(i)}}.$$

The proof is concluded by trivially adding the contribution of the initial infective. $\square$

It is worth stressing again how such a factorisation is heavily based on the same assumptions behind Wald's identity for epidemics and Ludwig's rank-based construction of Chapter 3.

5.9 Summary and conclusions

In this chapter, a thorough methodological discussion about how to compute or approximate the real-time growth rate for a stochastic epidemic has been presented.

In summary, the Lotka-Euler equation

$$\int_0^{+\infty} \beta(\tau) e^{-r\tau} d\tau = 1$$

that provides an implicit condition for the real-time growth rate when a single-type model with non-random time-varying infectivity profile $\beta(\tau)$ is considered can be extended to:

1. a random infectivity profile, with average $\beta(\tau)$;
2. a random susceptibility, in addition to a random infectivity: the only difference is that $\beta(\tau)$ is now a weighted average, where the weights are given by the distribution of susceptibilities;
3. a multitype model: in this case, the Lotka-Euler equation generalises into a condition that the dominant eigenvalue of the matrix

$$K_r = \left(\int_0^{+\infty} \beta_{ij}(\tau) e^{-r\tau} d\tau \right)$$

has dominant eigenvalue 1;

4. a multitype randomised model, where $\beta_{ij}(\tau)$ are now the suitable averages of the infectivity profiles between types i and j .

The assumption of separable mixing drastically simplifies the results.

When small groups are considered, as in the households or the households-workplaces models, local saturation increases dramatically the complexity of the results but, under suitable simplifying conditions, the global infectivity profile of a households (and similarly for workplaces) can be factorised in the form

$$\beta_H(\tau) = \beta(\tau) + \beta * \gamma_H(\tau),$$

where $\beta(\tau)$ represents the average infectivity profile of an infective outside the households and $\gamma_H(\tau)$ represents the average rate at which new cases occur in the household epidemic, i.e. the average rate at which the household cumulative incidence increases.

When both households and workplaces are considered, averaging across all random disease histories and across all possible households and workplaces epidemics, the model reduces to a two-type model (where the two types are households infected through a local workplace contact and a global contact, denoted by L and G) and therefore r can be obtained imposing that

$$K_r = \begin{pmatrix} \int_0^{+\infty} \beta_{LL}(\tau) e^{-r\tau} d\tau & \int_0^{+\infty} \beta_{LG}(\tau) e^{-r\tau} d\tau \\ \int_0^{+\infty} \beta_{GL}(\tau) e^{-r\tau} d\tau & \int_0^{+\infty} \beta_{GG}(\tau) e^{-r\tau} d\tau \end{pmatrix}$$

has dominant eigenvalue 1.

Numerical computation of the rate $\gamma_H(\tau)$ can be performed exactly only in the Markovian case of exponentially distributed length of infectious period. However, a method proposed in Fraser (2007) to approximate the rate $\gamma_H(\tau)$ is described. The method is based on the idea of computing the number of cases in each generation using a Reed-Frost model and describing their times of occurrence by autoconvoluting the generation time distribution, which approximates the distribution of times between cases in subsequent generations.

Fraser's method is applied to the households-workplaces model in Section 7.1 to estimate the household reproduction number R_H from the real-time growth rate r .

Numerical exploration of the quality of the approximation is also performed, and reveals that the approximation improves with decreasing within-household infectivity and generation time distributions that are more peaked around their average.

Future research should extend the method of Fraser to random disease histories, which could be achieved by substituting the simple Reed-Frost model with a randomised one.

Chapter 6

Simulation of stochastic epidemics

In this chapter, the construction of a numerical code to simulate stochastic epidemics is described. Such a computational tool is then used to perform a theoretical exploration of the contribution of the presence of workplaces in shaping the infection spread and to investigate the relationships between important epidemiological quantities, like R_0 , the real-time growth rate and the average final size of the epidemic, which are usually complex and often analytically intractable functions of more fundamental model parameters and are therefore represent more naturally a model output than a model input.

6.1 Epidemic simulator

6.1.1 Introduction

The households-workplaces model considered in Chapter 4 is mathematically challenging and analytical results are limited. In particular, the analysis is mainly confined to the household reproduction number R_H . Other results, mainly concerning the probability of a large outbreak and the final size are available (see Ball & Neal, 2002) but are rather technical and require restrictive conditions to be satisfied.

Furthermore, results about the system dynamics, the daily incidence at the peak and the time of the peak are very limited, despite the practical relevance that these quantities

have in informing the health care system about the stress it is likely to experience and the time available to respond to the epidemic spread. For all these reasons, and as an instrument to check analytical and semi-analytical results against, a computer code has been written to simulate epidemics *in silico*.

The simulation is stochastic, which means that each run provides a different realisation. The fundamental computer operation on which the whole simulation is based is the generation of (pseudo-)random numbers, through which the occurrence or non-occurrence of events is decided. This allows the exploration of the probabilistic outcome which characterises stochastic models, but at the price of increased computational intensity and a more complex sensitivity analysis.

The simulation is individual-based, so that it is sufficiently flexible to allow the estimation of reproduction numbers (by counting infectious contacts or real infections and averaging across individuals) and to keep track, if needed, of who acquires infection from whom, and when.

In general, the simulations of a time process can be *synchronous* or *asynchronous*. In a synchronous simulation, the time changes in fixed time steps and the system state at a certain time, together with the generation of random numbers, is used to update the system state at the next time-step. The computation is usually efficient, but the discreteness in time introduces approximations in the process and, in general, distributions of time intervals between events (e.g. the generation time distribution) may be significantly altered.

On the contrary, an asynchronous simulation is exact in time (up to machine precision, of course). An excellent example is given by the so-called *Gillespie algorithm* (see Gillespie, 1976, 1977, or Keeling & Rohani, 2008, pag. 201-205), in which, after an event has occurred, the time interval to the next event is first drawn, then the type of event occurring at that time is decided and the system is updated. For this reason, asynchronous simulations are also called *event-based* simulations. The Gillespie algorithm is computationally efficient, but is limited to Markovian processes: drawing the time to the next event, based

only on the current state, requires an assumption of memoryless in the process, which is usually satisfied only when the process is a superposition of Poisson processes with constant rates (e.g. constant recovery rate, in the case of an epidemic).

In order to allow for exact results, the simulation coded is asynchronous, but the limitation to Markovian processes only is avoided by storing events in a time-ordered list which is followed as the epidemic progresses and is continuously enriched by the addition of new events. To maintain the time order, every time a new event needs to be stored, the list is searched for the right place where to place it.

The presence of an event list to store, scan and update is expensive both in terms of computing time and memory usage, and therefore limits the population size to the order of a few million individuals on a common PC (processor: 1.86 GHz; RAM: 1 GB; operating system: Windows XP). However, it has the strong benefit of allowing the use of very general forms of disease histories which, except in the case of constant infectivity and exponentially distributed length of infection, result in non-Markovian processes. Furthermore, it is sufficiently flexible to be readily extended to more complex models, like the households-workplaces one, while maintaining exactness in time.

The programming language used to code the simulation is C, which has been chosen for its flexibility and its high level of efficiency.

6.1.2 Homogeneously mixing population

The structure of the simulation is here described in the case of a single-type and homogeneously mixing population. Households and workplaces are added later on.

Individual-based simulation

The simulation is individual-based, i.e. a memory slot is reserved to store the characteristics of each individual: an identification number, an infectious state (susceptible, infective, ...), a time of infection, a time of passage from latent to infectious stage (if there is a latent period), a time of recovery, the number of infectious contacts the individual generates and

the number of real infections. Some of the characteristics of individuals, like the time of infection, are stored once and for all; others, like the infectious state, are updated as new events occur. If needed, other properties could be added, like the generation index or the time elapsed since the infection of the infector and the infection of the individual under consideration.

Event-based simulation

The simulation is event-based, i.e. each event is treated separately. In particular, information for each event occurring (infection, recovery, maybe passage from latent to infectious stage) is stored in the time-ordered event list.

Whenever an individual is infected, his or her infectious life is constructed according to the model adopted for disease history. Two main model types have been considered: a model where all individuals experience the same disease history characterised by a time-varying infectiousness (hereafter denoted as *TVI*, see Section 2.1.3) and a model where the infectious period is characterised by constant infectiousness of a length randomly drawn from an arbitrary distribution (hereafter denoted by *CI*, constant infectivity; see Section 2.1.4). In this *CI* model, an latent period with random length can be absent, giving the standard *SIR* model of Section 2.1.4 or present (*SEIR* model). The only practical constraint on the distributions used in both models is that it must be possible to numerically draw random numbers from them.

The epidemic runs as follows:

1. For each initial infective, his or her infectious life is constructed and all the infectious contacts towards other individuals are stored in the main event list.
2. The time is updated to the time of the next event.
3. If the next event is an infectious attempt from i to j and j is susceptible at that time, then j becomes infectious and his or her infectious life is constructed.
4. If the next event is the recovery of i , then i 's status changes from I to R .

Each new event is stored in the main event list in time order, i.e. the list is searched for the right place. This operation is quite inefficient, but the efficiency can be improved using a *hash table* (see below).

Since what happens at time t has influence only on the future, the epidemic can be efficiently followed (by keeping track of the changes in the states of the individuals and storing the relevant quantities) while constructing and adding to the event list the next infections and recoveries, i.e. there is no need to construct the full event list first and then follow it through again in order to analyse the epidemic.

Assume now that individual i is infected. In the *TVI* model, the times of the infections from i are drawn from the specified distribution and the recovery of i is then placed at a fixed time after the last infection event: this allows the use of distributions, like the gamma distribution, for which there is no upper limit for the duration of the infectious period (non-compact support). For the *CI* models, in the *SEIR* case a length for the latent period is first drawn and the event “transition from E to I ” is placed in the main event list. Then, for both the *CI* models (*SIR* and *SEIR*), the length of the infectious period is first drawn and the recovery event is stored in the main event list. All the infectious contacts are then drawn from a uniform distribution and placed in the main event list. This guarantees that different infectious contacts from an infective are not independent in the *CI* model, but are independent *conditional* on the length of the infectious period (e.g. see Diekmann & Heesterbeek, 2000).

Efficiency can be improved with some “tricks”:

1. For each infectious contact, it can be first selected the identity of a proposed infectee j . If j is already infected or recovered at the present time (time of infection of i), then the event is discarded (auto-infections are allowed, although they are not stored). This avoids drawing a random number for the time of infection when not needed (as it may be costly, like in the case of a gamma distribution).
2. If j is susceptible, then the infection event is stored, even if it may happen that,

before that time, j will be infected by someone else and that this infectious contact will not result in an infection. No way to avoid such extra storage was found.

3. For small groups and large infectiousness (e.g. relevant for households) it may happen that j experiences multiple infectious contacts from j . When a susceptible is contacted, then he or she is considered contacted (will eventually end up infected) and the time of the contact is drawn. If contacted again, the time of this new contact will replace the old one only if smaller; otherwise it is discarded. All the infectious contacts just created are stored and all contacted individuals are then reclassified as susceptibles, so that each of them is still eligible for infections from other individuals between the current time and the time of the infectious contact just stored.

Other strategies can be used to improve the speed of the simulation. It has been mentioned before that searching along the linked list of events looking for the right place of a new event is particularly inefficient, with a complexity that was estimated via empirical trials to be roughly cubic in the population size N . The main reason for this is that the average time interval between the infection of i and the infection of j remains constant as the epidemic progresses, but the density of events in time increases towards the peak of the epidemic (roughly proportionally to the number of cases at the peak).

The complexity can be reduced by creating an array of time points ($h_0 = 0, h_1, h_2, \dots$) with fixed time-step dh and breaking the event list (Figure 6.1a) in sublists (Figure 6.1b). When a new event needs to be placed, the right time point h_n can be immediately found and only the small sublist needs to be scanned through.

Selecting the right time-step is not trivial, as it is optimal when inversely proportional to the number of events at the peak, which is unknown and depends not only on the population size, but also on other parameters like the infectivity and the recovery rate.

However, keeping fixed all the other parameters, the peak is proportional to N , so it is sufficient to set the time-step inversely proportional to N . Therefore, from a practical perspective, it is convenient to run various epidemics on a small population searching manually for the best time-step and then scale it down accordingly as the population size

increases and the runs become computationally more intensive.

The improvement is remarkable as the complexity seems to decrease from cubic to linear. In computer science language, the chain of sublists of Figure 6.1b is usually referred to as *direct chained hash table*.

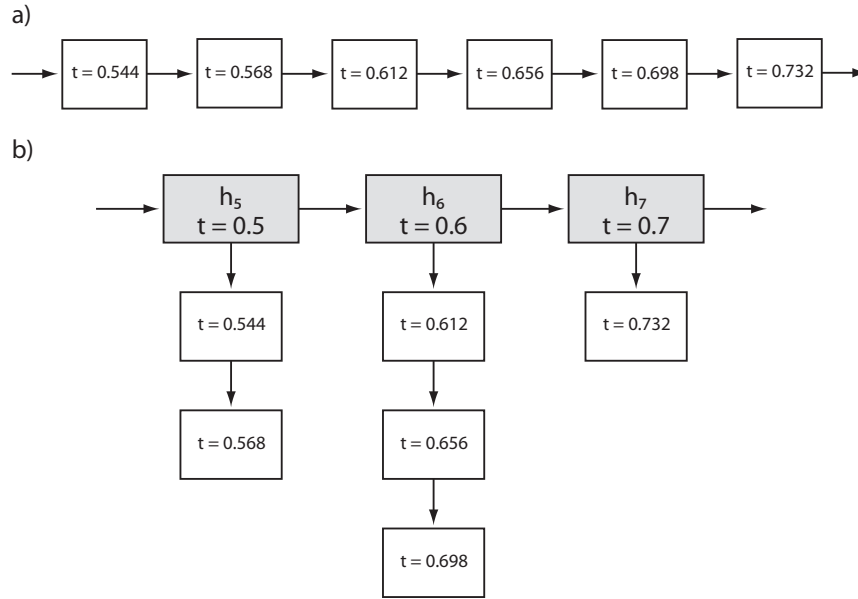


Figure 6.1: A portion of the structure used to store the events on which the stochastic simulation is based. Only the time of occurrence of each event is represented. In *a*) the events are all placed on a linked list: searching the right point of the list where to store an event is computationally inefficient, as the density of events increase considerably at the peak of the epidemic. In *b*) the list of *a*) is broken into small sublists: the efficiency in searching the right point where to store an event is much improved as it is sufficient to consider directly the time-step h_n of interest and search only through a small sublist.

As the size of the population is N and the only events occurring are infections and recoveries (and passages from latent to infectious stage, when *SEIR* is used), the number of total events is always less than $2N$ (or $3N$ in *SEIR*). However, some events need to be stored even if they are not useful, namely, the supposed infections of an individual at time t that does not occur since the same individual is infected by someone else before t . For this reason there is no fixed upper bound on the total number of events, and memory for storing the events would need to be allocated as the epidemic progresses. This is inefficient,

if done for each single event.

One possible improvement is obtained by using a *memory pool*, i.e. by allocating memory for a number of events which is thought to be sufficiently large all at once before running the simulation and use that pool every time an event needs to be stored. If then the storage room is still insufficient, other memory can be allocated *ad hoc*.

6.1.3 The simulation output

Multiple realisations of the epidemic are performed. After each epidemic has run, the relevant information is stored on a vector at time points $(t_0 = 0, t_1, t_2, \dots)$ with fixed time-step dt . The epidemic process is therefore exact in time but is then post-processed in a convenient discrete-time framework.

A major outbreak is defined as an outbreak infecting more than a fixed proportion of the total population size (the cut-off is taken to be 1% of the total population). Major epidemics are then shifted in time and synchronised at the peak, in order to superimpose properly the exponentially growing phases despite the random time the each epidemic requires to become established.

The real-time growth rate r is computed at each time point t_k by considering the time window $[t_k, t_k + \Delta t]$ and using the formula:

$$r(t_k) = \frac{\ln(I(t_k + \Delta t)) - \ln(I(t_k))}{\Delta t}.$$

The window size Δt is taken to be equal to the generation time T_g . In order to improve precision, $I(t_k)$ is taken to be the sum of the prevalence values across all epidemics (properly shifted in time).

An example of the outcome for the real-time growth rate is reported in Figure 6.2a.

The computation of the basic reproduction number R_0 and of the effective reproduction number R_e is less trivial to perform. The effective reproduction R_e is defined as the average number of new infections generated by a “typical” infective. Therefore, R_0 and R_e coincide when the population is fully susceptible but, when global saturation effects start playing

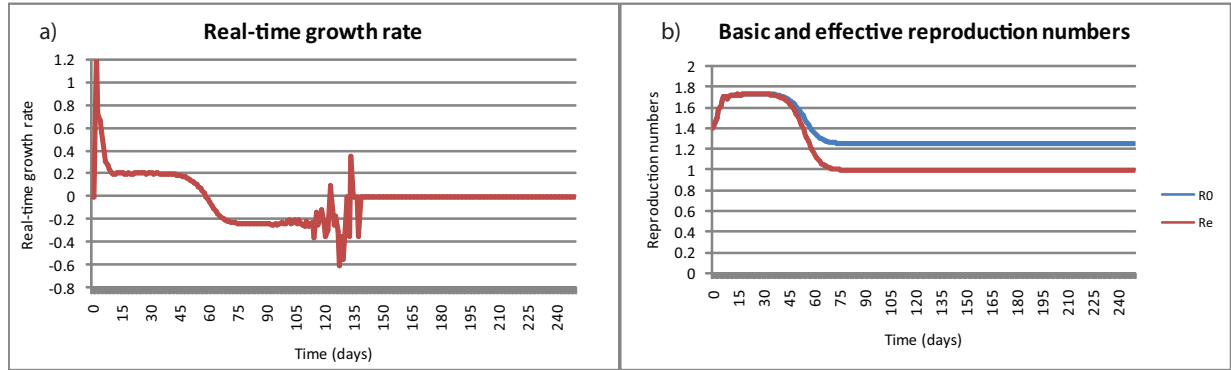


Figure 6.2: An example of a possible simulation outcome. In *a*) the real-time growth rate r is plotted as a function of time: note the presence of an initial phase driven by random events, as the number of infectives is small, followed by an exponentially growing phase during which r stabilises on the value of interest. Then the peak of the epidemic is reached and the number of infectives starts declining. The small number of cases occurring at the end of the epidemic is again responsible for the random oscillations seen around day 120. In *b*) both the basic and the effective reproduction are plotted as functions of time: again, after an initial phase driven by random events, they stabilise on a constant value, which has to be the same for both parameters since almost all infectious contacts result in real infections. The values on which they stabilise towards the end of the epidemic depend on how they are computed (see text).

a role, not all infectious contacts result in infectious and therefore R_e is in general smaller than R_0 .

The two reproduction numbers are computed by averaging the number of infectious contacts and of real infections across all individuals infected in the time window $[0, t_k]$ so that, as time increases, the number of terms in the computation of the average grows. Note that, since the computation is conditioned on a large epidemic having occurred, the reproduction numbers tend to be overestimated because the few cases at the beginning tend to generate more infections than the average: in fact, if this is not the case, the epidemic is likely to go extinct and to be discarded from the computation of R_0 . However, as time progresses, the overestimation becomes negligible.

The details about the computation of R_0 and R_e are non-trivial. The two parameters cannot be computed at time t_k by dividing the number of infections (or contacts) up to time t_k by the number of individuals that got infected up to t_k (cumulative incidence), as

some of these individuals have not finished spreading their infection.

However, they also cannot be computed at time t_k by averaging the number of infections only across those individuals that recovered by time t_k , as this would exclude from the average all those individuals that are still infectious at time t_k : this is not a problem if the length of the infectious period is constant but, if there is variability in the length of the infectious periods, those individuals that are infectious for a long time (and that are usually responsible for many infections) have to “wait” for long before being included in the average; instead, those with short infectious periods (and that generate fewer cases than expected) are included immediately and cause an underestimation of the reproduction numbers.

For all these reasons, the computation of the reproduction numbers at time t_k is performed by considering all the individuals that have been infected by time t_k , but counting *all* their contacts or real infections, including those occurring after time t_k . The contacts are then summed altogether and divided by the number of individuals considered, i.e. by the cumulative incidence at t_k .

Finally, when dealing with multiple epidemics, R_0 and R_e can be calculated for each large epidemic and then averaged across epidemics, or the numbers of contacts (or infections) and the number of cases can be summed across epidemics and then the ratio is computed. The two methods provide the same asymptotic estimates for R_0 and R_e as their differences become negligible when the average is taken across a large number of individuals. The method used in the simulation is the second one.

The output of the simulation representing the values of the reproduction numbers is given in Figure 6.2*b*. Note that the values coincide during the exponentially growing phase.

At the end of the epidemic, the value of R_e converges to 1, as it should since the number of total infections is equal to the final size minus the initial infectives and in a large population the fraction of initial infectives is negligible.

The value of R_0 should remain constant, but this is not the case because of the details concerning how contacts are stored in the simulation: in particular those infectious contacts

hitting individuals that have already been infected by the time of the contact are not stored (see the strategies to improve efficiency described above) and therefore not counted; the addition of these infections gives a constant value of R_0 also after the exponentially growing phase.

6.1.4 Households-workplaces model

When households and workplaces are considered, it is necessary to construct the social structure beforehand and then run the epidemic.

Given the total population size N and the size distribution for households, each household is constructed by drawing a random size for it from the distribution and filling it with individuals. The places available in the household, as many as the size of the household, are here referred to as *h-slots*. Households are constructed until the number of h-slots exceeds N . The size of the last household is then reduced so that the total number of h-slots matches the total population size; the influence on the generated household size distribution is negligible when the population is large.

Each household is characterised by an identity number, a size and the identity numbers of the individuals it contains. Each individual is endowed with the identity number of the household he or she belongs to.

Individuals are placed in households in order as the households are constructed (this is computationally efficient and does not compromise the independence between two or more simulation runs, as the initial infectives for each epidemic are selected at random from the population). Workplaces are similar to households, except that the individuals are selected at random to fill the *w-slots*.

The network obtained with such a construction is a bipartite random one with specified degree distribution for both node types (see Section 4.2.3 or Newman, 2002).

The epidemic spreads similarly to the case without households and workplaces. The main differences are:

1. There are three parameters R_h , R_w and R_g describing the average number of contacts that each infective makes in each of the three environments (household, workplace and community).
2. When drawing the infectious life of an individual, contacts in each environment are considered separately, except for the fact that, if a susceptible is contacted more than once in different environments or in the same by the same infective, only the first contact is stored as a possible infection.
3. When an infectious event is stored, the environment where it occurred is also recorded.

Since most analytical results refer to households, it is useful to track the epidemic spread also at the household level. However, analytical results mainly refer to the exponentially growing phase, during which simplifying approximations, like that of assuming that re-infections of the same household from outside do not occur, can be exploited. In the simulation, however, such approximations break down when the exponentially growing phase finishes. The strategy adopted to estimate the quantity of interest is that of defining some other quantities (usually a lower bound and an upper bound) that all coincide with the desired one during the exponentially growing phase: apart from allowing the numerical estimation of the quantity of interest, such a method provides also information on the start, the length and the end of the exponentially growing phase.

An example is given by the average final size of a household epidemic $1 + \mu_H$. For each infected household, the cases in the household epidemic are counted until the time of the first reintroduction from outside while the household epidemic is still running: the average of this quantity across all households infected by time t provides a lower bound for the estimate of $1 + \mu_H$ at time t . The cases in the household epidemic are still counted until the first moment at which there are no infectives left in the household: the average of this quantity across all households infected by time t provides an upper bound for the estimate of $1 + \mu_H$ at time t , as a reintroduction can occur while the household epidemic is ongoing

and this would increase the final size. Note that the two quantities should be the same during the exponentially growing phase, as households epidemic have a high probability of terminating before the occurrence of a reintroduction. After the exponentially growing phase, reintroductions become more likely and cases in each household are still counted to provide information about the ultimate final size, i.e. the one recorded after the full epidemic has terminated. A example output is represented in Figure 6.3a.

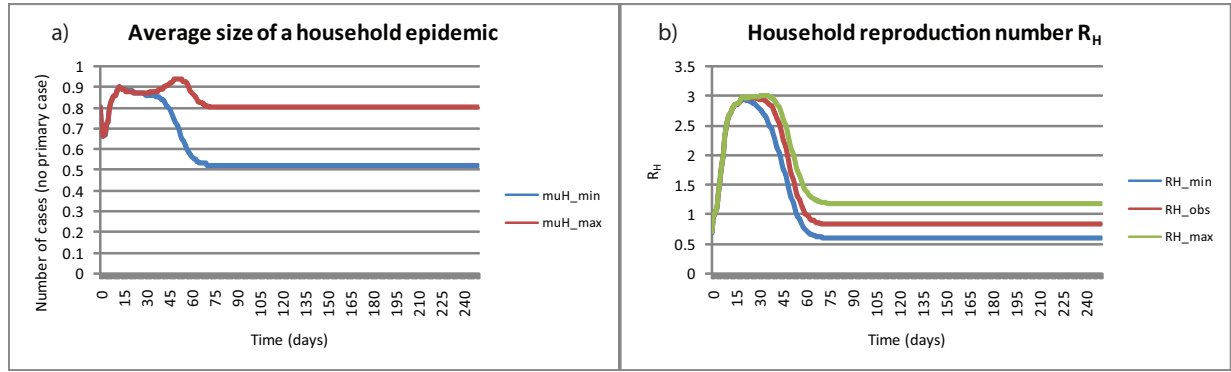


Figure 6.3: An example of a possible simulation outcome for the spread among households. In *a*) the average size of a household epidemic (excluding the primary case) is plotted as a function of time: an upper and a lower bound are considered (see text), which should coincide during the exponentially growing phase. In *b*) the household reproduction number R_H is plotted as a function of time: three different values are computed, which should coincide during the exponentially growing phase. Note that the larger and more complex the unit considered is (the household, in this case), the shorter the suitable window for the exponentially growing phase becomes: therefore, the estimation of R_H becomes less accurate as the severity of the epidemic increases.

The computation of the household reproduction number is based on the same kind of ideas, although details are more intricate and are not described. When a household is infected, the workplaces of all the cases infected in the household epidemics keep track of the identity of the “infector” household, so that, when each workplace epidemic has finished, its final size (representing the number of new households infected) is added to a counter referred to the “infector” household. Of all possible quantities that can be computed, only three are stored, and a possible output is represented in Figure 6.3b.

6.1.5 A more realistic social structure

Although mathematically convenient, the social structure obtained when households and workplaces are connected at random is unrealistic. A more realistic social structure is then constructed from the random one, by rewiring the network connecting households and workplaces according to a Metropolis-Hastings algorithm in order to match a specified distribution of distances travelled by individuals between their household and their workplace. Such an approach is mainly motivated by the availability of data concerning commuting patterns in a population. The algorithm is similar to that used in Riley & Ferguson (2006).

In order to rewire the network according to some spatial patterns, it is first necessary to position households and workplaces in space. The “city” is considered to be a square of side 1 and households and workplaces are positioned at random in it. Denote by $k(d)$ the probability density that a randomly selected individual travels a distance d from home to work. Denote by $\mathcal{N}$ the set of all individuals, with cardinality N , and let i represent the generic individual; let j be the generic household index and let H_j be the corresponding household; analogously, let k be the generic workplace index referring to workplace W_k . Then the probability of each particular network configuration $\mathcal{Z}$ to occur, given that household H_j is located at (x_j, y_j) , $\forall j$, and workplace k is located at (x_k, y_k) , $\forall k$, is:

$$\begin{aligned}\mathbb{P}(\mathcal{Z}) &= C \times \prod_{i \in \mathcal{N}} [\mathbb{P}(i \text{ is in } H_j) \mathbb{P}(i \text{ is in } W_k) \mid \text{dist}(H_j, W_k)] \\ &= C \times \prod_{i \in \mathcal{N}} \left[\frac{|H_j|}{N} k(\text{dist}(H_j, W_k)) \frac{|W_k|}{f_{H_j}(\text{dist}(H_j, W_k))} \right]\end{aligned}$$

where C is the normalising constant, $|\cdot|$ indicates the number of individuals in $\cdot$, dist is the Euclidean distance and $f_{H_j}(d)$ represents the density of the number of w-slots (i.e. number of places available within workplaces) at a distance d from household H_j .

In practice, since households and workplaces are discrete units and cannot be properly described using densities, the kernel $k(d)$ is discretised in $k(d_n)$, $n = 1, 2, \dots$ and $f_{H_j}(d_n)$ counts the number of workplace slots in the area between the circle of radius d_{n-1} that of radius d_n (annulus), centred at the location of H_j .

The Metropolis-Hastings algorithm works as follows:

1. two individuals a and b , with distances d_a and d_b between their households and workplaces, are selected at random from the whole population;
2. their workplaces are swapped;
3. the new distances travelled by a and b , d_a^{new} and d_b^{new} , are calculated;
4. the likelihood of the new network $\mathbb{P}(\mathcal{Z}^{new})$ is calculated;
5. the new network configuration is accepted with probability $\Pi = \mathbb{P}(\mathcal{Z}^{new})/\mathbb{P}(\mathcal{Z})$.

Note that, by swapping the workplaces of a and b , all the conditions that constrained the initial random network (position of households and workplaces, size of households and workplaces, total population size) are still satisfied.

6.1.6 Efficient approximations

Computing, for each household, how many workplace slots are available at every (discretised) distance, has complexity $O(N^2)$. For this reason, two approximate ways of speeding up the algorithm are described here.

The first one consists in modifying the kernel $k(d)$ by assuming that, for each selected household j , the number of workplaces slots at distance $d \in [d_1, d_2]$ from H_j is proportional to the area of the annulus between the circles of radii d_1 and d_2 . It is therefore necessary to divide, point by point, the kernel by the area of the annulus, which gives a pair-wise kernel $\kappa_1(d) = k(d)/2\pi d$. The term *pair-wise* (see Riley & Ferguson, 2006) refers to the fact that an individual is $\kappa_1(d_1)/\kappa_1(d_2)$ times more likely to select a workplace at distance d_1 from his or her household with respect to one at distance d_2 (if the ratio is smaller than one, then the workplace at distance d_2 is preferred).

Note that, in the definition of κ_1 , errors are introduced at long distances, when the circles exceed the boundaries of the square. However, error quantification is difficult since it depends on the position of H_j .

Another approximation consists in treating the position of H_j as random and dividing the kernel, point by point, by the probability that two randomly scattered points on the square have distance d from each other. Such a probability density, denoted by $h(d)$, can be constructed numerically beforehand or, if the discretisation of the kernel $k(d)$ is coarse enough compared to the number of households and workplaces in the square (*overcrowding*), can be approximated by the initial distribution of distances travelled by individuals when the network connecting households and workplaces is random.

More formally, this second strategy corresponds to ignoring the detailed structure of the network configuration and use a Metropolis-Hastings method to sample from a distribution on the set D of all the N -dimensional vectors of distances travelled by each individual:

$$D = (d_i), \quad i = 1, \dots, N$$

During the acceptance-rejection step, two distances d_a and d_b are selected randomly, and are proposed to be substituted by two new distances d_a^{new} and d_b^{new} . Since the formal selection of the two initial distances is thought as selecting two individuals at random, which have both households and workplaces placed randomly, the ideal operation of swapping their workplaces causes the two new distances to be drawn from the proposal distribution $h(d)$, independently of the actual state of the system. Therefore, the new acceptance probability Π must include this proposal distribution:

$$\begin{aligned} \Pi(d_a, d_b, d_a^{new}, d_b^{new}) &= \min \left(1, \frac{k(d_a^{new})k(d_b^{new})h(d_a)h(d_b)}{k(d_a)k(d_b)h(d_a^{new})h(d_b^{new})} \right) \\ &= \min \left(1, \frac{\kappa_2(d_a^{new})\kappa_2(d_b^{new})}{\kappa_2(d_a)\kappa_2(d_b)} \right), \end{aligned}$$

where $\kappa_2(d) = k(d)/h(d)$ is the new pair-wise kernel used in this approximation.

In all the algorithms discussed above, the main inefficiency remains the fact that very small and very large distances are unlikely to be proposed and therefore that the Markov Chain mixes at a very slow rate. Further research could focus on a way to force small distances to be proposed in a way similar to that of Riley & Ferguson (2006). Large distances are less relevant for the situation considered here, since the interest is focused on

investigating the presence of short loops, which are more likely to form with more localised connections between households and workplaces.

6.2 The role of workplaces in the epidemic spread

6.2.1 Motivation

This section focuses on a thorough exploration of how the presence of workplaces shapes the epidemic dynamics. The exploration presented here is rather theoretical, more than being specifically aimed at obtaining quantitatively realistic results, since there is almost no data available about workplace sizes or transmission intensity within workplaces.

Such data are also not expected to be readily available, because there is not even a clear definition for a workplace. While most individuals live in a household and most children go to school, the idea of workplace is much less precise. Some issues arising when trying to identify a workplace for modelling purposes are the following:

1. there is a great variety in the possible types of jobs, much more than the different types of schools and household structures;
2. the modes of interaction between individuals at work are extremely variable across different types of jobs: for example, a teacher spends probably most of the time among other individuals, while those working in a call centre or delivering post or milk have arguably much less social interaction during working hours;
3. some individuals might not have a job at all;
4. even in the same working environment, it is very difficult to decide who should be modelled as sharing the workplace with whom: should an entire university be modelled as a workplace? A college? A department? An office?

Any working structure could in principle be used, so long as the transmission is then modelled in the appropriate fashion. For example, if a university is considered as a work-

place, then it is probably necessary to model the interaction between individuals using a network description, since homogeneous mixing is quite unrealistic for such a complex structure.

Since the households-workplaces model of Chapter 4 assumes homogeneous mixing within workplaces, the workplace is here thought to be something realistically small. In this respect, a class of 30 students is roughly considered as a possible upper bound for workplaces, with workplace sizes going down to 1 for those types of jobs where there is no relevant social interaction (or to model individuals who do not have a job).

Although some data concerning sizes of big institutions or companies may be available, there is little hope to obtain realistic data about office sizes or about other job environments. An exception is represented by school classes, about which data might be available: however, since the model considered here does not discriminate between adults and children, such data would still be insufficient to parameterise it properly. For all these reasons, researchers (e.g. Riley & Ferguson, 2006; Cauchemez *et al.*, 2008) often do not bother identifying workplaces, but instead distinguish between household transmission (and maybe school transmission) and community transmission, which includes all other modes. However, since the main purpose of this study is to try to disentangle the effect of workplaces from the other forms of transmission, then workplaces will be kept separate from other structures, despite all the issues arising in dealing with them.

The main exploration strategy consists in allowing the workplace sizes to vary and observe the impact this has on the spread of the epidemic.

The effect that the workplace size has on the household reproduction number R_H is clearly the easiest to study. In fact, denote by R_w the one-to-all average total infectiousness within a workplace, i.e. the average number of secondary cases that an infective would cause in his or her workplace if there were no saturation effects: once a particular form of disease history is assumed, R_w and the size n_W of a workplace univocally identify a value for μ_W and therefore, for fixed μ_H and R_g , a single value of R_H . When an exponentially distributed infectious period is assumed, exact analytical results about the real-time growth

rate obtained via continuous-time Markov chain theory (Section 5.5) can also be used in the investigation. Much more complex, instead, is the study of all the other relevant quantities (R_0 , the average final size, the incidence at the peak of the epidemic, the time of the peak, etc.).

The use of stochastic simulations allows the exploration of these quantities and the additional exploration of the effects of different disease histories on the epidemic real-time growth rate. Furthermore, it offers an instrument for cross-checking the other results.

For the purpose of this more theoretical exploration, four different types of disease histories with comparable generation times T_g are considered:

exp: constant (one-to-all) infectivity β during an infectious period that is exponentially distributed with average $T_g = 2.6$ days. The numerical value for the generation time is taken from Ferguson *et al.* (2005).

gamma: constant (one-to-all) infectivity β during an infectious period that is distributed according to a $\Gamma(\alpha, \delta)$ (gamma with shape parameter α and scale parameter δ). The parameter values are $\alpha = 4$ and $\delta = 1$, such that it approximates the disease history estimated in Cauchemez *et al.* (2004) with mean 3.8 days (approximated to 4) and standard deviation 2.0 days. The approximate value $T_g = 2.5$ days (instead of 2.4 of Cauchemez *et al.*, 2004) is comparable to that of the *exp* disease history.

SEIR: an ad-hoc constructed disease history given by an exponentially distributed latent period with mean 1 day, followed by a constant one-to-all infectivity β throughout an infectious period exponentially distributed with mean 1.6 days. For this choice, $T_g = 2.6$ days.

TVI: a time-varying infectivity profile $\beta(\tau) \sim \Gamma(\alpha, \delta)$, with $\alpha = 4$ (an ad-hoc choice) and $\delta = 1.54$: $T_g = 2.6$ days.

Some comments:

1. The *exp* disease history allows the real-time growth rate for the simulation to be compared with the numerical method described in Section 5.5.4.

2. Both the *exp* and the *gamma* histories are characterised by a positive infectivity at $\tau = 0$, i.e. immediately after infection, and therefore a quicker spread is expected from those distributions compared to the other two.
3. Because of the arguments in Chapter 3, in terms of the final size, the *TVI* disease history in a homogeneously mixing group of size n is equivalent to a Reed-Frost model with one-to-one escaping probability $q = e^{-R_0/(n-1)}$, where $R_0 = \int_0^{+\infty} \beta(\tau) d\tau$.
4. For fixed value of R_H , and therefore of μ_W (when μ_H and R_g are constant), the *exp* and *SEIR* disease histories are those with the greatest variability in the total infectivity across individuals (due to the exponentially distributed length of the infectious period), immediately followed by the *gamma* and then by the *TVI*. In fact, for fixed R_w , the more variability, the smaller the average final size (see Section 2.2.4 or Wallinga & Lipsitch, 2007). Therefore, to achieve the same μ_W , *exp* and *SEIR* require the largest R_w , followed by that for *gamma* and finally by that of *TVI*.

6.2.2 Effects of saturation

The presence of small groups where contacts are repeatedly made with the same individuals clearly has an impact on the disease spread. However, it is not at all clear what this impact is and, most importantly, when such impact is sufficiently small to conclude that the addition of such groups in the model does not cause results to be significantly different from those obtained with a homogeneous mixing model.

One reason why the effect of small groups on the spread of the infection is unclear is that there is no easy way to compare structurally different models: for example, there are problems in the definition of R_0 for a household model or, even when a definition is chosen, such a parameter is usually difficult to deal with analytically. In the case of a household model, one possible exploration of the effects of saturation is to tune the within- and between-household infectivity such that the reproduction number for households

$$R_* = R_g(1 + \mu_H) \quad (6.2.1)$$

is kept constant (other choices are possible, although the simple analytical expression for R_* makes it a convenient parameter to keep fixed). In this case, it is possible to highlight different contrasting effects, the strength of which is not easy to compare.

The following discussion focuses on pure households models, as they are simpler than the households-workplaces ones, but at the same time present local saturation effects within households. The additional presence of workplaces is considered in Section 6.2.3.

Depletion of susceptibles

The first effect of the presence of households is that there is a quick local depletion of susceptibles which causes most of the contacts to be with already infected individuals, thus reducing the number of new cases occurring. Furthermore, as time progresses, different cases in the group experience very different availabilities of susceptibles.

The transmission intensity is switched from between- to within-household transmission by decreasing R_g while R_h increases, such that R_* is kept constant (Equation (6.2.1)). When all the transmission is outside, an infective generates on average R_g cases, all in the next generation. When the transmission is concentrated inside the household, depending on its size, it may be impossible to infect the same number of cases or at least it may be that the same number of cases appears scattered across different generations because some infectious attempts fail to generate new cases in the first generation. In the limit of an infinite household size, all the cases are infected in the first generation as saturation effects disappear. Therefore, the “speed” at which new cases appear slows down with increasing saturation.

Furthermore, if the size of a household is n_H , the value of μ_H can never increase above $n_H - 1$. For this reason, the condition that R_* must remain fixed does not allow R_g to decrease below the value $R_g^{min} = R_*/n_H$. As this value is approached, the infectivity in households explodes in order to force μ_H closer and closer to $n_H - 1$ (see Figure 6.4).

Repeated contacts between the same individuals

A completely different effect is given by the fact that, when an infective repeatedly contacts the same individuals, only the first contact results in an infection. Therefore, although fewer cases occur in comparison to a non-saturated process, they tend to occur earlier and this speeds up the spread of the infection.

An analytically tractable example is given by the process occurring in a household of size 2 when considering the standard *SIR* model with constant recovery rate ν and within-household infection rate β_h and between-household infection rate β . In this case, outside the household, an infective generates new cases at an average rate $\beta(t) = \beta e^{-\nu t}$. Inside the household, the rate at which the primary case infects the other individual is $\beta_h(t) = \beta_h e^{-(\beta_h + \nu)t}$, since there is a competing hazard between infection of the secondary case and recovery of the primary case. Note that the bigger β_h , the faster the occurrence of the secondary case. Note also that repeated contacts with the same individuals are more likely for larger within-household infectivities and smaller household sizes.

Overlapping generations

When generations overlap, an infective that would have occurred in a certain generation may be infected earlier from individuals in subsequent generations. This effect again tends to speed up the spread of the infection and its contribution is more pronounced when it is more likely that generations overlap, i.e. when the latent period is short and the infectivity is widely spread in time, but also when the household sizes are not too small so that more generations are possible.

Global effect of saturation

In general, the effect of repeated contacts tends to be limited, unless the infectiousness is extremely high and the number of cases to contact very limited. Furthermore, unless infectivity is widely spread in time, the latent period is short and the size of the group is sufficiently large to allow different chains of generations to appear, the effect of overlapping

generations tends to be small too. Therefore, intuition can suggest that the speed-up caused by such effects is in general not sufficient to counterbalance the slowing down effect of the depletion of susceptibles. Fraser (2007) proposed a smart approach to take explicitly into account the effect of depletion of susceptibles in the dynamics of a household model and showed that the exact system behaviour is well approximated by an analytically tractable behaviour obtained when the other two effects are neglected.

The next section uses simulations to observe the effects of the considerations above in the case of the households-workplaces model.

6.2.3 The case of households and workplaces

The different assumptions of Sections 2.1.3 and 2.1.4 on the disease histories and the size of the workplaces are explored by fixing values for R_H , R_g and μ_H and, for each n_W , by tuning R_w such that μ_W and therefore R_H is kept constant. Then R_g varies, affecting R_w .

Extreme behaviour range

As noted above, since $0 \leq \mu_W < n_W - 1$, not all possible values of R_g can lead to a given R_H : when R_g is too large, it may cause also R_H to be too large, even when there is no transmission within workplaces; vice versa, when R_g is too small, the workplace size might be too small to guarantee the required value of R_H . What happens in this case is that R_w explodes, in the attempt to drive μ_W as close as possible to $n_W - 1$, and this explosion is more sudden for the *exp* and *SEIR* disease histories than for the others. In this extreme parameter range, the effect of repeated contacts can overcome the local depletion of susceptibles resulting in an overall speed-up of the spread of the infection. This type of extreme behaviour is unlikely to occur in a practical context, but the consequences on the real-time growth rate are here explored numerically to confirm the intuition.

First, the behaviour for the *exp* disease history is examined. In Figure 6.4, R_H has been fixed to 2 and the workplace size (equal for all workplaces) varies from 3 to 10. R_g is allowed to vary between 0 and 0.8 and the corresponding value of R_w is obtained. It

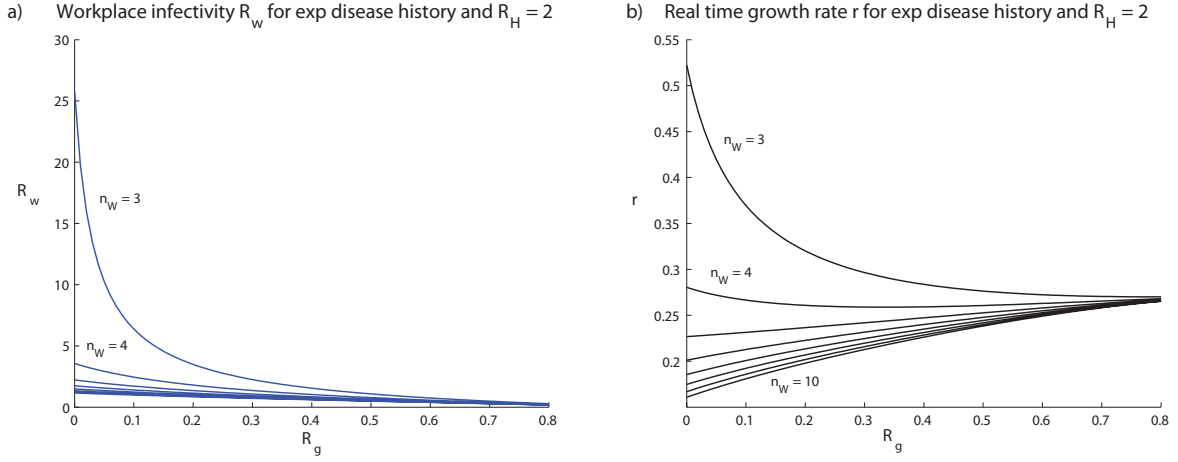


Figure 6.4: Values of (a) R_w and (b) the real-time growth rate r as functions of the global transmission R_g when the *exp* disease history is assumed (see text) and $R_H = 2$ is kept fixed. Different curves are obtained, from top to bottom, for workplace sizes n_W (equal for all workplaces) increasing from 3 to 10. The household size n_H is fixed to 3 and μ_H is fixed to roughly 1.064 (obtained from *exp* when $R_h = 2.5$). Note that, as $R_H \rightarrow \mu_H \mu_W$ for $R_g \rightarrow 0$, keeping R_H fixed at 2 for small R_g requires μ_W to be very close to 2, and in fact, when n_W is small (e.g. 3), R_w blows up for small R_g (a). For very large values of R_w , infections in workplaces occur almost immediately and determine an overall increase in the real-time growth rate (b).

appears evident the sharp increase in R_w required in small workplaces to keep $R_H = 2$ when R_g decreases to 0.

The consequences of this extreme behaviour is that the infection rate R_w is so high in workplaces that all infections occur almost instantaneously, and this is helped further by the fact that *exp* is characterised by a positive infectivity as soon as the individual is infected. Therefore, for small workplace sizes, small values of R_g force the real-time growth rate to increase as shown in Figure 6.4.

When *SEIR* is considered, then the behaviour of R_w is the same, but the real-time growth rate r is different. For *gamma* and *TVI* the values of R_w are given in Figure 6.5.

For $n_W = 3$ some simulation outputs are collected in Figure 6.6. The results for *exp* confirm the analytical results obtained with the method described in Section 5.5.4.

Note how the other disease histories do not show the extreme effect of r increasing

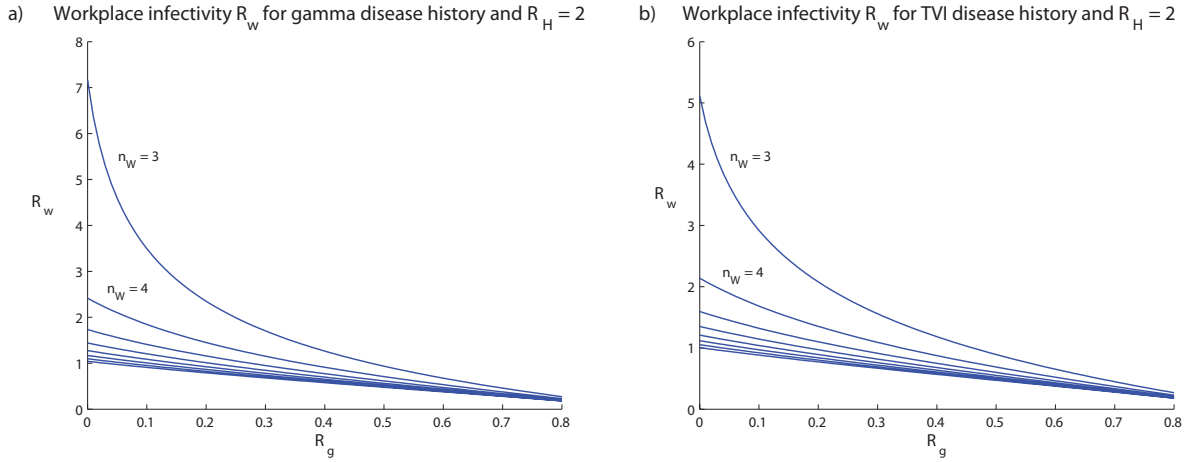


Figure 6.5: Values of R_w for (a) the *gamma* and (b) the *TVI* disease histories as functions of the global transmission R_g when $R_H = 2$ is kept fixed. Different curves are obtained, from top to bottom, for workplace sizes n_w (equal for all workplaces) increasing from 3 to 10. The household size n_H is fixed to 3 and μ_H is equal to 1.064, as in the case of *exp* (obtained in a) for $R_h = 1.84$ and in b) for $R_h = 1.67$). Note that, for small workplace sizes, R_w still increases as $R_g \rightarrow 0$, but less sharply than for the *exp* disease history.

when $R_g \rightarrow 0$. There are two different reasons. First, in the case of *SEIR* and *TVI*, the infectivity is zero immediately after infection and increases as time progresses: therefore, even with a very high infectivity (as in *SEIR*), new cases cannot occur too soon. Second, in the case of *gamma* and *TVI*, R_w is not required to be as large as for *exp* or *SEIR*, as it is easier to obtain a high average final size in a workplace than with *exp* or *SEIR* (Section 2.2.4 or Wallinga & Lipsitch, 2007). This double effect explains also why *TVI* realises the lower value of r .

Normal behaviour range

For realistic conditions of infection spread, a much less extreme behaviour than the one presented above is expected. Even for a highly infectious disease like measles, the number of infectious contacts an individual makes at home is limited. In Hope Simpson (1952), a one-to-one escaping probability of $q = 0.25$ is estimated under the assumption of density-dependent transmission: the correspondent value of R_h is highly dependent on the house-

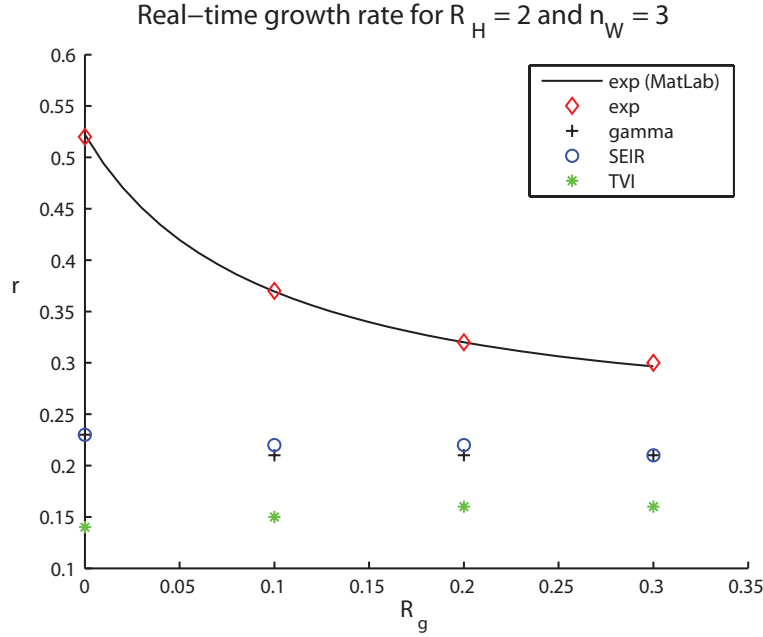


Figure 6.6: Comparison of the real-time growth rates obtained from different disease histories as a function of the global transmission R_g when the workplace size is $n_W = 3$ and $R_H = 2$ is kept fixed. The household size n_H is fixed to 3 and μ_H is roughly 1.064 in all cases: for *exp* and *SEIR*, $R_h = 2.5$, for *gamma* $R_h = 1.84$ and for *TVI* $R_h = 1.67$. Numerical results are an average over 100 simulations for a population of a million individuals. The black line is instead calculated in MatLab, using the technique of Section 5.5.4, thus providing an important cross-checking for the correctness of both the theoretical approach and the large scale simulation.

hold size, and can vary between $R_h = 1.4$ for a household of size 2 to $R_h = 8$ for a household of size 6. Despite not being a quantitatively reliable estimation, as more recent evidence suggests that the frequency-dependent transmission model is more appropriate even for small households (Cauchemez *et al.*, 2004), it suggests possible values of R_h which are in general much less than the value of 25 observed in Figure 6.4.

When the saturation effects are not too strong, then the normal behaviour is still given by an increase in R_w as R_g decreases, but this time accompanied by a decrease in r , since the depletion of susceptibles is the main consequence of the local saturation in workplaces. Results for the case of *exp* are reported in Figure 6.7, in the case of $R_H = 2$ and workplaces of size n_W with values ranging from 4 to 30.

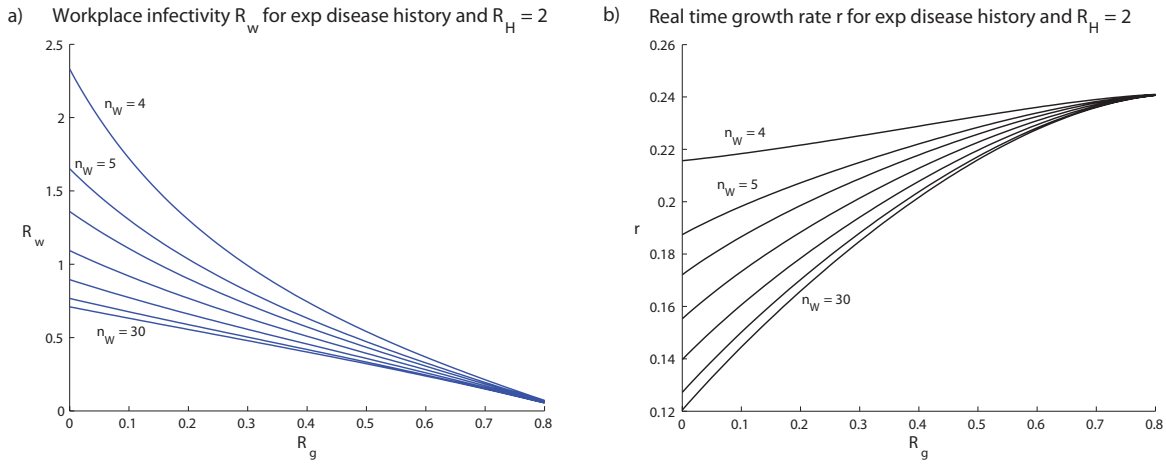


Figure 6.7: Values of (a) R_w and (b) the real-time growth rate r as functions of the global transmission R_g when the *exp* disease history is assumed and $R_H = 2$ is kept fixed. Different curves are obtained for a workplace size n_W (equal for all workplaces) taking values, from top to bottom, of 4, 5, 6, 8, 12, 20 and 30. The household parameters are $n_H = 4$ and $\mu_H = 2.34$ (obtained from $R_h = 2$ in *exp*); R_g varies from 0 to 0.8 and R_w is computed such that R_H remains constant.

Note that:

1. as n_W increases, the local saturation effects become less and less visible and R_w becomes more and more linearly related to R_g .
2. For fixed R_g , the reduction in R_w is smaller and smaller as n_W increases
3. For fixed R_g , the reduction in r is smaller and smaller as n_W increases.

Finally, note that, since the infectivity is not too strong, larger values of the real-time growth rate are obtained as the infectivity is shifted from within workplaces to the community (Figure 6.7b).

6.2.4 Ignoring the social structure

A very interesting and important issue concerning any model which is more complex than the homogeneous mixing one is the understanding of how much information is “lost” when neglecting such an additional complexity.

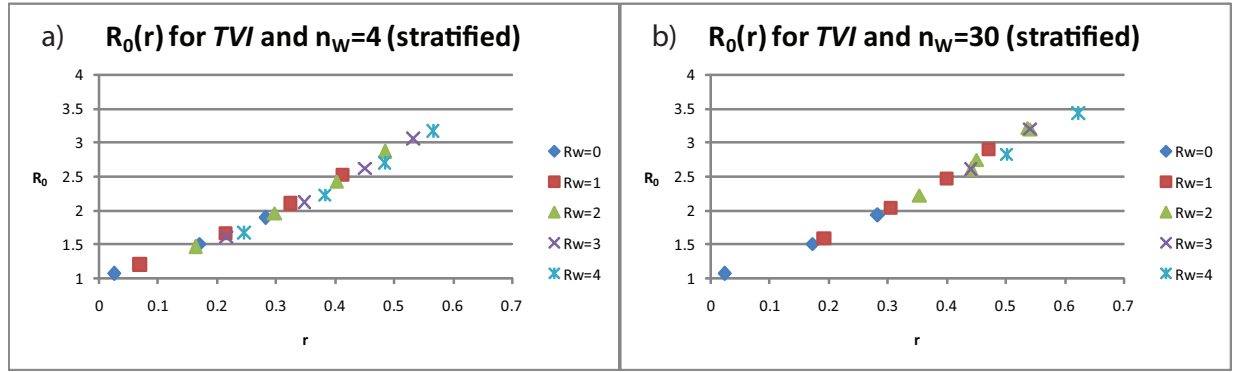


Figure 6.8: Plot of R_0 against the real-time growth rate r when the *TVI* disease history is assumed, stratified by the amount of infectivity in workplaces, for (a) workplaces of size 4 and (b) of size 30. Parameter values are: $R_g = 0.2, 0.6, 1$ and 1.4 , and $R_w = 1, 2, 3, 4$ and 5 . Not all combinations of parameters values are necessarily used, as when they are both large the exponentially growing phase becomes very short and the estimation of R_0 and r becomes less reliable. $R_h = 1.35$ and the household size distribution is obtained from the UK census data (2001). Numerical results are obtained by averaging across 100 epidemics in a population of a million individuals.

The question is of both theoretical and very practical interest. In fact, if predictions from oversimplified models are very inaccurate and more detailed models need to be used, it would be necessary to identify which additional data are necessary to collect and how to do so.

The problem is not trivial and essentially too generic to be answered in this form. In this section, a more specific problem is considered.

Estimating R_0 from the real-time growth rate

Assume an epidemic is observed in real-time and an estimation of the real-time growth rate r is obtained, and assume that information is available about the disease history and the within-household infectivity: what is the error made when a homogeneous mixing model is assumed and R_0 is estimated using the Lotka-Euler equation

$$R_0 = \left(\int_0^{+\infty} \omega(\tau) e^{-r\tau} d\tau \right)^{-1} \quad (6.2.2)$$

with respect to assuming the household-workplaces model of Chapter 4?

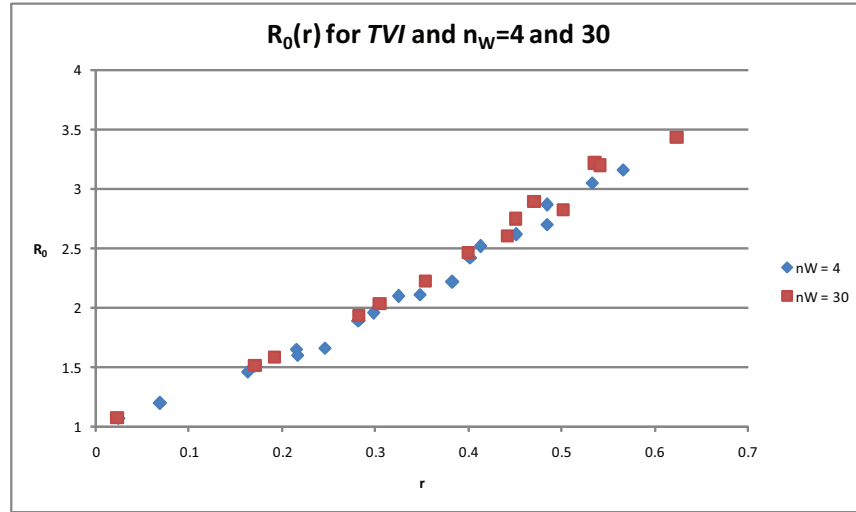


Figure 6.9: Plot of R_0 against the real-time growth rate r when the *TVI* disease history is assumed, stratified by the size of workplaces ($n_W = 4$ or 30). Parameter values are, as before, $R_g = 0.2, 0.6, 1$ and 1.4 , and $R_w = 1, 2, 3, 4$ and 5 , but not all combinations are necessarily used. $R_h = 1.35$ and the household size distribution is obtained from the UK census data (2001). Numerical results are obtained by averaging across 100 epidemics in a population of a million individuals.

Note that, since the model is more complex and involves two infectivity parameters instead of one (i.e. R_w and R_g , instead of R_0), it is likely that the same value of the real-time growth rate r is obtained for different combinations of the basic parameters of the model, and the same combinations may lead to substantially different values of R_0 in the two models. In this case, the disease history and the real-time growth rate would not be sufficient to estimate R_0 and further information about other model parameters would be required. Somewhat surprisingly, this is not necessarily the case.

Figure 6.8 gathers the outcomes of many simulations with the household-workplaces model for combinations of $R_w = 0, 1, 2, 3, 4$, $R_g = 0.2, 0.6, 1, 1.4$ and for workplaces sizes $n_W = 4$ and 30 (in order to experience strong and weak saturation effects, respectively). The disease history is initially assumed to be of the *TVI* form (as before, the generation time distribution is a $\Gamma(\alpha, \delta)$, with shape parameter $\alpha = 4$ and generation time $T_g = 2.6$ days) and the household size distribution is taken from the UK census data (2001). R_h is taken to be 1.35 , such that for the *TVI* disease history, the average epidemic size in

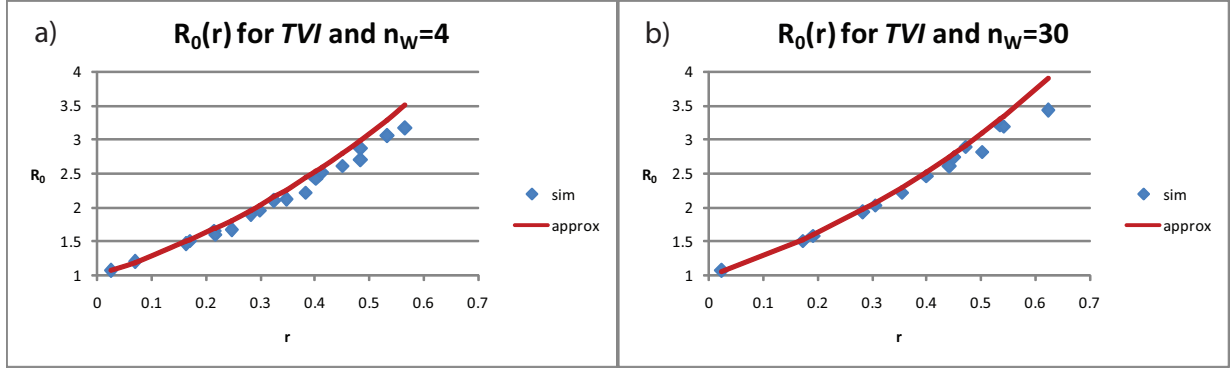


Figure 6.10: Plot of R_0 against the real-time growth rate r when the *TVI* disease history is assumed, together with the approximation obtained by neglecting the social structure, for (a) workplaces of size 4 and (b) of size 30. Parameter values are, as before, $R_g = 0.2, 0.6, 1$ and 1.4 , and $R_w = 1, 2, 3, 4$ and 5 , but not all combinations are necessarily used. $R_h = 1.35$ and the household size distribution is obtained from the UK census data (2001). Numerical results are obtained by averaging across 100 epidemics in a population of a million individuals. The approximation is computed in MatLab using the Lotka-Euler equation (see text).

the household of a randomly selected individual, including the primary case (i.e. $1 + \mu_H$), equals 1.87 as in Fraser (2007).

Not all parameter combinations are plotted, as when the overall infectivity is very large (large R_g and R_w , corresponding to large r and R_0) estimation via stochastic simulation becomes less reliable, because the population of a million individuals is not large enough to observe a clear exponentially growing phase. This is often the case when a large workplace size is considered.

The same data are reported in Figure 6.9 for both workplace sizes. Apparently, variations in the workplace sizes and in the amount of within-workplace infectivity (and therefore the intensity of local saturation), as well as in the global infectivity, seem to bear negligible effect on the relationship between r and R_0 , and this gives hope that all these details can be neglected when estimating R_0 in an emerging epidemic. Figure 6.10 shows how the same results are approximated by the estimation of R_0 under simple homogeneous mixing assumptions.

The approximation is remarkably good, although it is possible to observe a slight overes-

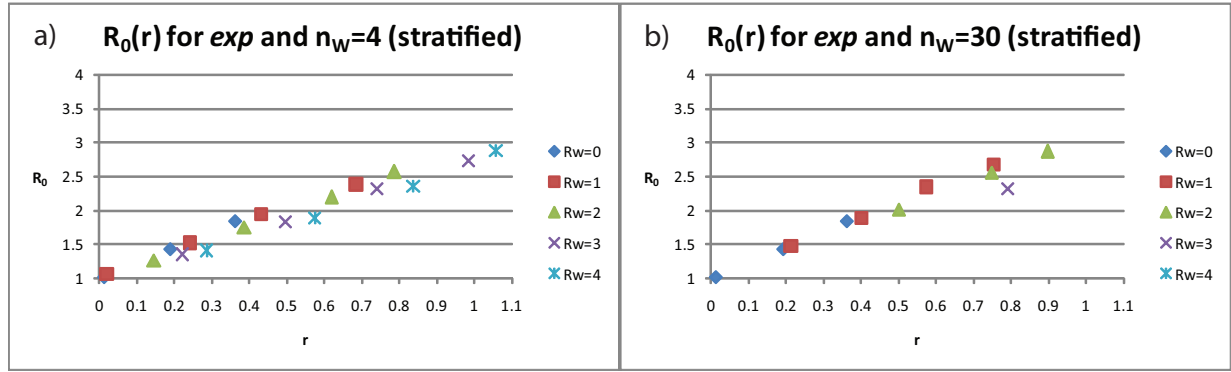


Figure 6.11: Plot of R_0 against the real-time growth rate r when the *exp* disease history is assumed, stratified by the amount of infectivity in workplaces, for (a) workplaces of size 4 and (b) of size 30. The other details are as before.

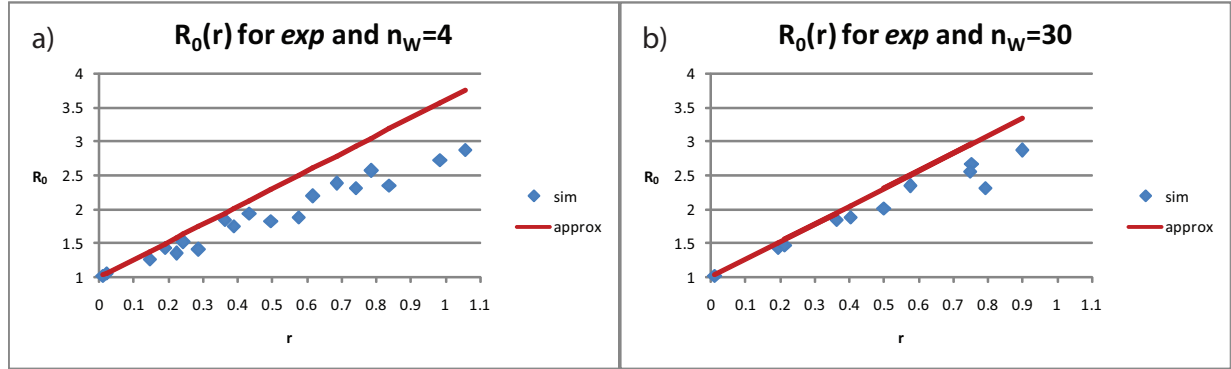


Figure 6.12: Plot of R_0 against the real-time growth rate r when the *exp* disease history is assumed, together with the approximation obtained by neglecting the social structure (see text), for (a) workplaces of size 4 and (b) of size 30. The other details are as before.

timination of R_0 when the overall infectiousness becomes very high. Such an overestimation occurs both for small and large workplace sizes, thus suggesting it is unlikely to be due to saturation effects, although they can contribute in stressing it.

The situation appears to be different for other disease histories. When *exp* is considered, the results are reported in Figure 6.11, stratified by intensity of workplace infection, for workplace sizes of 4 and 30 and in Figure 6.12, accompanied by the approximated estimation of R_0 .

Note that the approximation is much less satisfactory, even for small values of r and

R_0 . Despite fewer data points appear in Figure 6.12b, it is possible to observe a magnifying effect of local saturation.

The outputs in the case of *TVI* and *exp* disease histories are markedly different. In order to investigate the reasons for this, it is first necessary to note that the two disease histories differ for two main reasons: first, in *exp* there is much more variability in the total infectiousness between different individuals compared to *TVI*; second, the time intervals between the infection of a case and the times at which he or she transmits the infection to other cases are widely spread in time in *exp*, while they are much more concentrated around the generation time T_g in the *TVI* with shape parameter $\alpha = 4$. In order to gain understanding of how the two differences influence the outputs, a third disease history is considered, namely a form of *TVI* with exponentially decaying infectivity (*TVI* with $\alpha = 1$), for which there is no variability in the total infectivity between different individuals but the time intervals between infections can vary considerably (generations tend to easily overlap). Hereafter, it will be referred to as *TVIexp*.

As for the other disease histories, the simulation outputs are plotted stratified by workplace infectivity in Figure 6.13 for $n_W = 4$ and $n_W = 30$.

The approximated estimation of R_0 by using the Lotka-Euler equation is reported in Figure 6.14 together with the simulated values.

The outcome is much closer to that obtained from the *exp* disease history, thus suggesting that the fundamental cause of bad performance in the approximation is the fact that the infectivity in *exp* and *TVIexp* is more widely spread than that of *TVI*. On the contrary, the variability in total infectiousness across individuals has very limited effect on the relationship between r and R_0 .

Some intuitive insight in this relationship can be provided through the following naïve argument. Consider a disease history where the infectivity profile, random or not, is highly peaked around the generation time T_g : then generations of infectives in a local epidemic are well separated. Consider now an individual infected at time t , which would infect s cases, but because of local saturation effects, succeeds in infecting only l of them: those

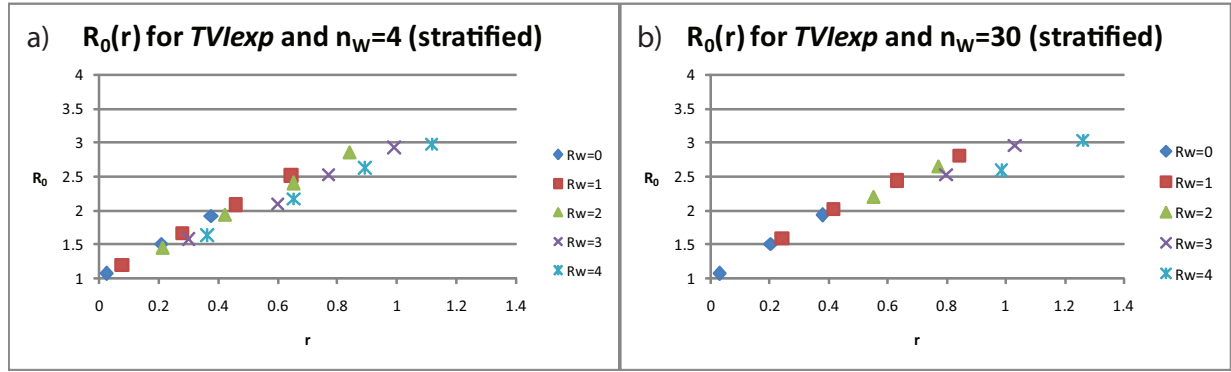


Figure 6.13: Plot of R_0 against the real-time growth rate r when the *TVIexp* disease history (i.e. exponentially decaying infectivity) is assumed, stratified by the amount of infectivity in workplaces, for (a) workplaces of size 4 and (b) of size 30. The other details are as before.

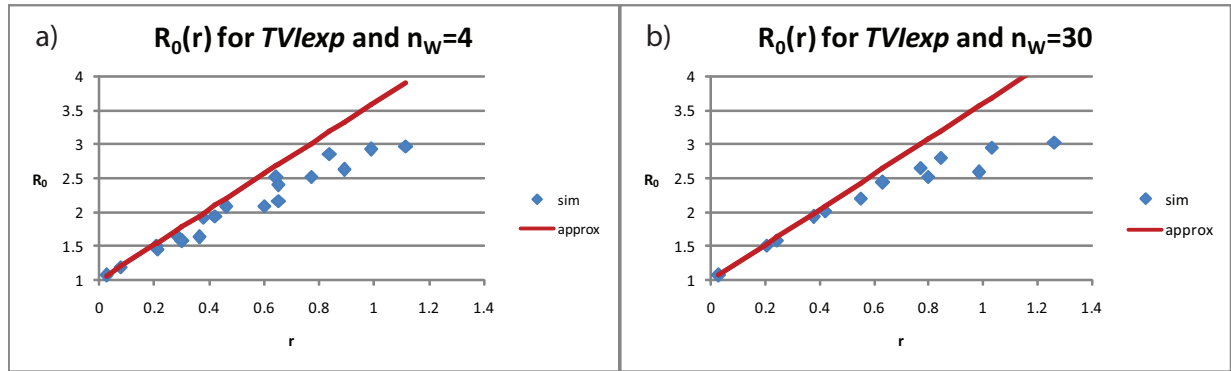


Figure 6.14: Plot of R_0 against the real-time growth rate r when the *TVIexp* disease history (i.e. exponentially decaying infectivity) is assumed, together with the approximation obtained by neglecting the social structure (see text), for (a) workplaces of size 4 and (b) of size 30. The other details are as before.

l cases would occur roughly T_g units of time after the infection of the infector, together with the remaining $s - l$ failed infectious attempts. If the infectivity profile (random or not) is widely spread in time, then it is more likely that, of all the s infectious contacts, the l that cause real infections are the first ones, as the ones occurring later have a much lower probability of hitting a susceptible individual. In other words, local saturation effects increase during the infectious life of an infective. This is true also when local groups are larger, although the effect is somewhat weaker.

Estimating the average final size

Apart from R_0 , another quantity of practical interest is the average final size of an epidemic (often shortened here in *afs*). Analytical results concerning the final size (Ball & Neal, 2002) could in some cases be used to investigate it, but the interest is now placed on the relationship between the average final size and the real-time growth rate, for which there are no exact analytical results in general. As in the previous section, the question addressed here could be: is the estimation of the average final size sufficiently accurate when the social structure is neglected?

When a *TVI* disease history is assumed, with the same parameters as before, Figure 6.15 suggests it is not necessarily the case: in fact, despite its little effect on R_0 (Figure 6.9), the final size seems to be highly affected by the intensity of local saturation (the stratification by amount of workplace transmission is more relevant for $n_W = 4$ than for $n_W = 30$).

An even more striking comparison is obtained when the parameters R_g and R_w are varied in opposite directions in the attempt of keeping fixed R_H , as shown in Figure 6.16.

If an approximated final size is obtained by solving the implicit equation

$$z = 1 - e^{-R_0 z}, \quad (6.2.3)$$

i.e. by neglecting the social structure, Figure 6.17 shows a small underestimation when workplaces are large (the disagreement is probably due to the saturation effects caused by the transmission in households), but a considerable underestimation when saturation in workplaces is much more intense.

Again, such a comparison is even more evident when the infectivity is moved from within workplaces to the community while R_H is kept fixed (Figure 6.18).

Note, however, that the approximation may still prove to be useful for a disease like influenza, given the fact that it is not very infectious compared to other transmissible diseases, like chickenpox or measles, and therefore that local saturation effects are limited.

It may be somewhat surprising that local saturation effects play no or little role when

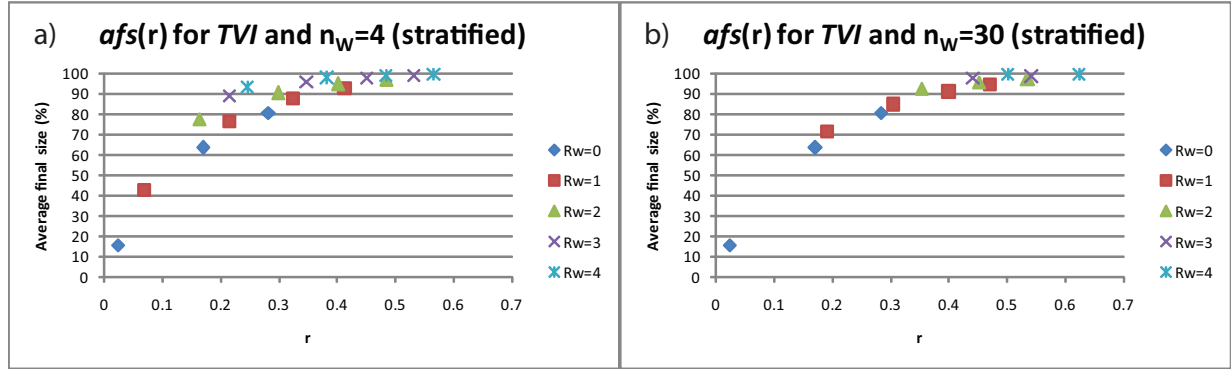


Figure 6.15: Plot of the average epidemic final size (afs) against the real-time growth rate r when the TVI disease history is assumed, stratified by the amount of infectivity in workplaces, for (a) workplaces of size 4 and (b) of size 30. Parameter values are: $R_g = 0.2, 0.6, 1$ and 1.4 , and $R_w = 1, 2, 3, 4$ and 5 . Not all combinations of parameters values are necessarily used, as when they are both large the exponentially growing phase becomes very short and the estimation of R_0 and r becomes less reliable. Numerical results are obtained by averaging across 100 epidemics in a population of a million individuals.

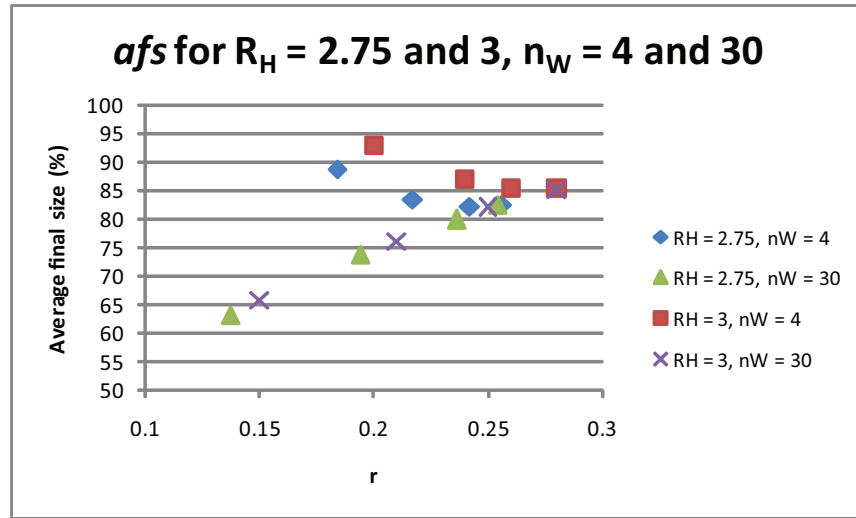


Figure 6.16: Plot of the average epidemic final size (afs) against the real-time growth rate r when the TVI disease history is assumed, for $R_g = 0.2, 0.6, 1$ and 1.4 (as before) and R_w obtained from a fixed value of R_H . Results are stratified by values of R_H (2.75 or 3) and sizes of the workplaces (4 or 30). $R_h = 1.35$ and the household size distribution is obtained from the UK census data (2001). Numerical results are obtained by averaging across 100 epidemics in a population of a million individuals.

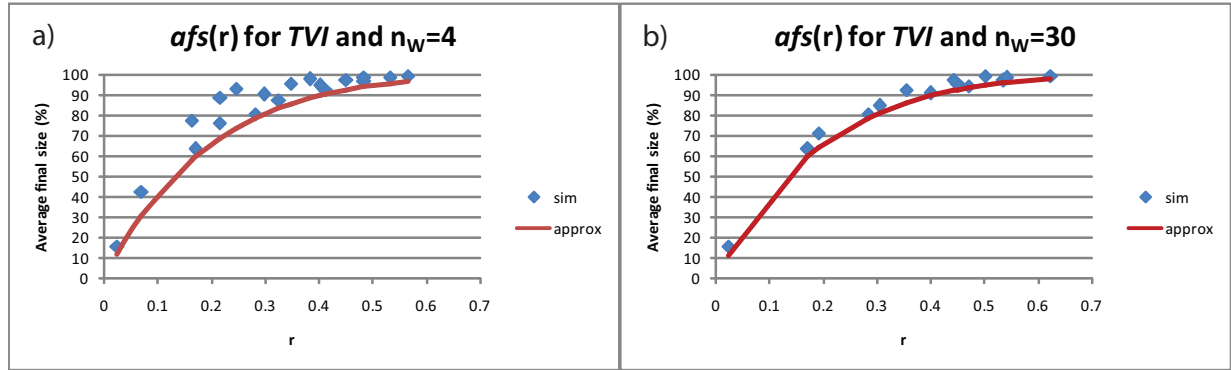


Figure 6.17: Plot of average epidemic final size (afs) against the real-time growth rate r when the *TVI* disease history is assumed, together with the approximation obtained by neglecting the social structure, for (a) workplaces of size 4 and (b) of size 30. Parameter values are, as before, $R_g = 0.2, 0.6, 1$ and 1.4 , and $R_w = 1, 2, 3, 4$ and 5 , but not all combinations are necessarily used. $R_h = 1.35$ and the household size distribution is obtained from the UK census data (2001). Numerical results are obtained by averaging across 100 epidemics in a population of a million individuals. The approximation is computed in MatLab using the Lotka-Euler equation (Equation (6.2.2)) to find an approximate estimate of R_0 and then iteratively solving the standard final size equation (Equation (6.2.3)) parameterised with the estimated value of R_0 .

estimating R_0 but have such a dramatic effect on the final size, although it should be pointed out that the relationship between the two epidemiological quantities is not expected to be trivial: in fact, while R_0 and r are quantities involved only during the exponentially growing phase of the epidemic, the final size is a property concerning the end of the epidemic, i.e. when a major fraction of the population is involved and reinfections of already infected households have to be taken into account.

When the *exp* disease history is considered instead, final size results are reported in Figure 6.19 for workplace sizes 4 and 30 and stratified by infectiousness in workplaces. The same results are collected altogether in Figure 6.20.

Apparently, saturation effects have little influence on the final size in this case (the leftmost point when $R_w = 0$ is an outlier, as R_0 is so close to 1 and the *exp* disease history has so much variability that the distinction between a large and a small epidemic becomes almost impossible; predictions from the final size equation in Figure 6.21 estimate the final

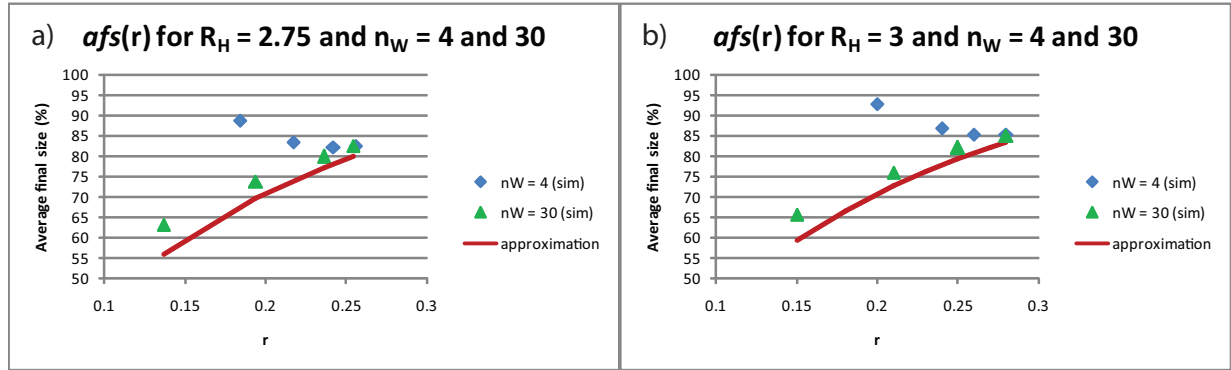


Figure 6.18: Plot of average epidemic final size (afs) against the real-time growth rate r when the TVI disease history is assumed, together with the approximation obtained by neglecting the social structure. As before, $R_g = 0.2, 0.6, 1$ and 1.4 , but R_w is obtained by fixing the value for R_H to 2.75 (a) and 3 (b). Results are stratified by workplace size (4 or 30). $R_h = 1.35$ and the household size distribution is obtained from the UK census data (2001). Numerical results are obtained by averaging across 100 epidemics in a population of a million individuals. The approximation is computed in MatLab using the Lotka-Euler equation (Equation (6.2.2)) to find an approximate estimate of R_0 and then iteratively solving the standard final size equation (Equation (6.2.3)) parameterised with the estimated value of R_0 .

size to be around 5% instead of the 30% shown above).

Approximations from the final size equation (Equation (6.2.3)) are obtained from the approximated value of R_0 reported in the previous section. Figure 6.21 shows a good agreement for both workplace sizes.

The additional set of data points in Figure 6.21a represents the average final size obtained from the final size equation (6.2.3), but when calculated with the “true” values of R_0 . The surprising result that the average final size estimated using the “wrongly” estimated R_0 provides indeed a much better approximation than the one obtained from the “true” value is most probably a coincidence.

Again, the additional analysis of the outcome in the case of the TVI_{exp} disease history can shed light on the reasons for such differences in the goodness of the approximate estimation of the average final size. Figure 6.22 reports the final size stratified by workplace infection, while Figure 6.23 reports the comparison with the approximate estimation obtained by ignoring the social structure.

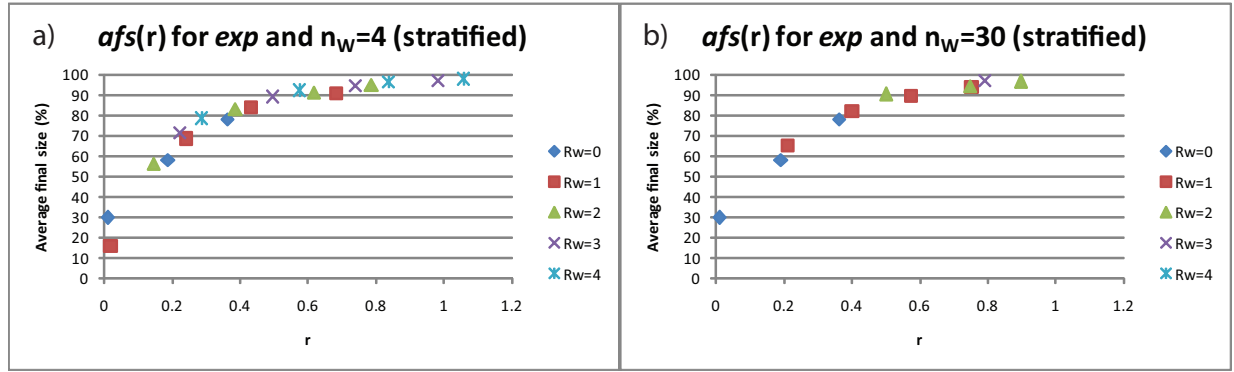


Figure 6.19: Plot of the average epidemic final size (afs) against the real-time growth rate r when the exp disease history is assumed, stratified by the amount of infectivity in workplaces, for (a) workplaces of size 4 and (b) of size 30. Parameter values are $R_g = 0.2, 0.6, 1$ and 1.4 , and $R_w = 1, 2, 3, 4$ and 5 , but not all combinations are necessarily used. $R_h = 1.35$ and the household size distribution is obtained from the UK census data (2001). Numerical results are obtained by averaging across 100 epidemics in a population of a million individuals.

This output is remarkably similar to that of TVI , thus suggesting that, when saturation effects are weak, the approximation to the final size works well for any type of disease history while, when saturation effects are strong, the main cause of the approximation failure seems to be lack of variability in the total infectiousness between individuals, more than the spreadness of infectivity in time. Particularly surprising is the fact that, with the TVI_{exp} disease history, the average final size is underestimated despite having been calculated with the overestimated value of R_0 (Figure 6.23a).

A possible intuitive explanation of this fact, which would also support the comment that the surprisingly good estimates of the average final size when the exp disease history is used (which are based on overestimated values of R_0) are merely a coincidence, is the following.

Saturation effects increase when R_w increases, but the final size can be kept constant by simultaneously reducing R_g . As discussed in Section 6.2.2, if this is the case, then the overall effect is a reduction in the real-time growth rate. If the average final size is plotted as a function of the real-time growth rate, this corresponds to a horizontal shift of a point towards the left, which is indeed what can be seen in Figure 6.15a (note, for example, the

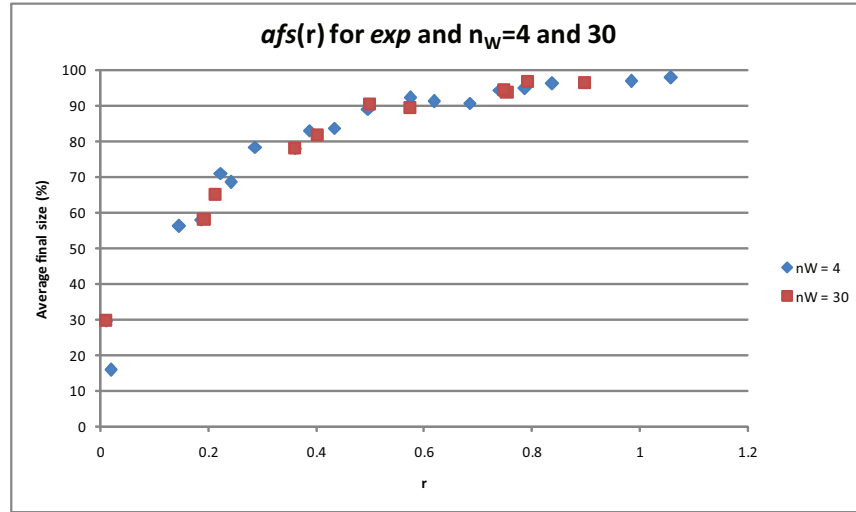


Figure 6.20: Plot of the average epidemic final size (afs) against the real-time growth rate r when the *exp* disease history is assumed, stratified by the size of workplaces ($n_W = 4$ or 30). Parameter values are, as before, $R_g = 0.2, 0.6, 1$ and 1.4 , and $R_w = 1, 2, 3, 4$ and 5 but not all combinations are necessarily used. $R_h = 1.35$ and the household size distribution is obtained from the UK census data (2001). Numerical results are obtained by averaging across 100 epidemics in a population of a million individuals.

points $(R_w, R_g) = (2, 0.6)$ and $(3, 0.2)$ corresponding to an average final size of about 90%).

Now, the disease histories *TVI* and *TVIexp* are characterised by a total infectivity which is the same for every individual and therefore, as already noticed at the end of Section 6.2.1, require a lower value of R_w compared to the *exp* disease history in order to achieve the same final size in a local epidemic. If R_w decreases and R_g increases, so that the overall final size remains constant, the *exp* disease history requires a much more pronounced increase in R_w with respect to the other two. Therefore, the same reduction in the real-time growth rate as described before (horizontal shift to the left) is mitigated, in the case of the *exp* disease history, by the fact that R_w increases more than with the other disease histories and therefore tends to accelerate the spread.

To sum up, the presence of saturation effects associates to the same final size smaller values of the real-time growth rate, but such a reduction is less evident in the case of the *exp* disease history. The fact that the final size is calculated from R_0 and that saturation effects lead to larger values of R_0 in the cases of *exp* and *TVIexp* suggests that a sort of

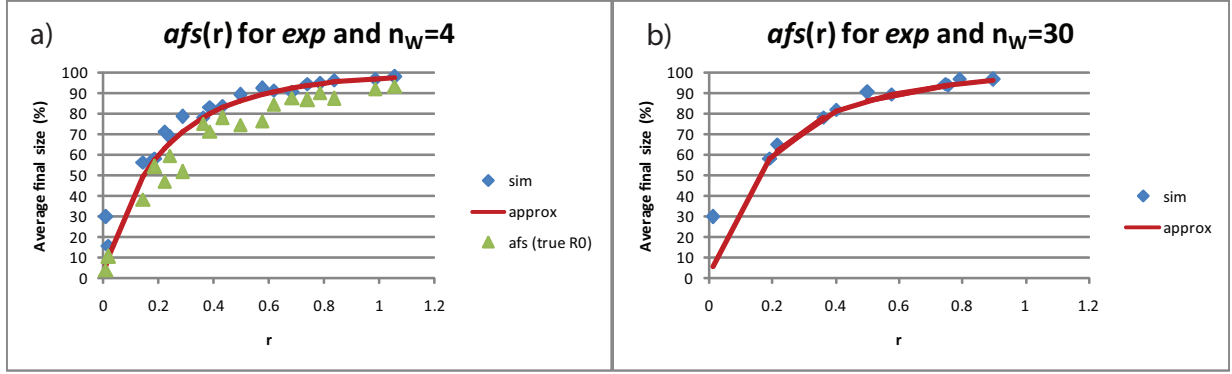


Figure 6.21: Plot of average epidemic final size (afs) against the real-time growth rate r when the *exp* diseases history is assumed, together with the approximation obtained by neglecting the social structure, for (a) workplaces of size 4 and (b) of size 30. The additional set of points in Figure 6.21a is computed by solving the standard final size equation (Equation (6.2.3)) when it is parameterised with the “true” value of R_0 as computed from the simulation, instead of the estimate obtained from the Lotka-Euler equation. Parameter values are, as before, $R_g = 0.2, 0.6, 1$ and 1.4 , and $R_w = 1, 2, 3, 4$ and 5 , but not all combinations are necessarily used. $R_h = 1.35$ and the household size distribution is obtained from the UK census data (2001). Numerical results are obtained by averaging across 100 epidemics in a population of a million individuals.

balance between opposite contributions is taking place.

Note however that, if one observes the real-time growth rate and tries to estimate R_0 and predict the average final size, then r has to be thought as fixed and the contribution of saturation effects needs to be read in terms of misestimation of R_0 and misprediction of the final size (i.e. the points in the figures are thought as moving vertically when saturation effects are added).

From this new perspective, the “true” values, i.e. those predicted when saturation effects are taken into account, differ from the approximated ones, i.e. those obtained when such effects are neglected, in the following fashion:

1. When *TVI* is assumed, the approximated and the true values of R_0 are very similar (Figure 6.10a); however, the approximated final size is an underestimation of the true final size (Figure 6.17a).

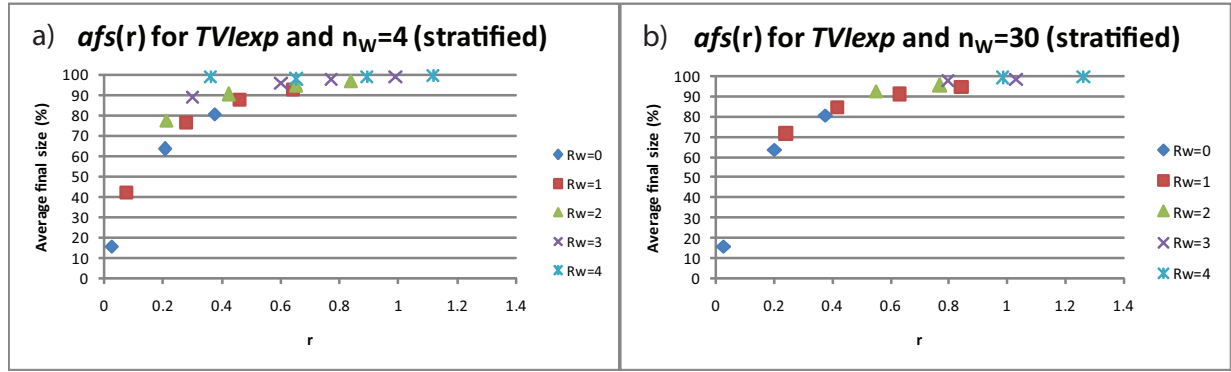


Figure 6.22: Plot of the average epidemic final size (afs) against the real-time growth rate r when the TVI_{exp} diseases history is assumed, stratified by the amount of infectivity in workplaces, for (a) workplaces of size 4 and (b) of size 30. Other details are as before.

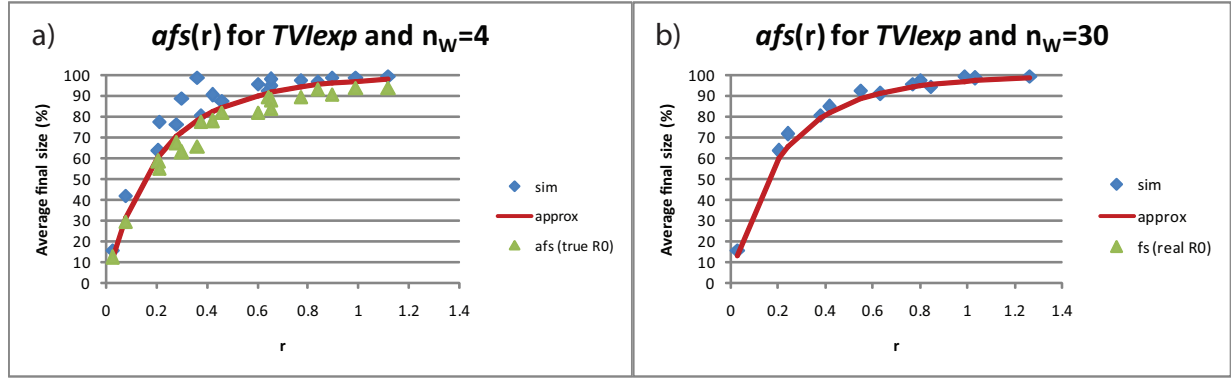


Figure 6.23: Plot of average epidemic final size (afs) against the real-time growth rate r when the exp diseases history is assumed, together with the approximation obtained by neglecting the social structure, for (a) workplaces of size 4 and (b) of size 30. The additional set of points in Figure 6.23a is computed by solving the standard final size equation (Equation (6.2.3)) parameterised with the “true” value of R_0 as computed from the simulation, instead of the estimate obtained from the Lotka-Euler equation. Other details are as before.

2. When TVI_{exp} is assumed, the approximated value of R_0 is larger than the true one (Figure 6.14a) because of the widely spread infectivity profile of TVI_{exp} . Now, as in the case of TVI , the approximated final size, computed from the true value of R_0 , is smaller than the true one computed from the same value of R_0 (green and blue points respectively, in Figure 6.23a); the fact that the approximated final size is computed from the approximated value of R_0 , which is larger than the true one,

reduces such a discrepancy, but the overall results remains a general underestimation of the predicted final size (Figure 6.23a).

3. When *exp* is assumed, the same scenario of *TVIexp* occurs, but the less pronounced underestimation in the final size results in an overall balance of the two contributions, thus leading to an unexpectedly accurate prediction of the final size (Figure 6.21a).

When workplaces have size 30, saturation effects are much less important, and apparently quantitatively similar between different disease histories. Therefore the approximated final sizes are very similar to the true ones: in particular, the one obtained for the *TVI* disease history is still slightly smaller than the true one, while the others are not (compare Figure 6.17b with Figures 6.21b and 6.23b), because the more peaked shape of the *TVI* does not visibly overestimate the true values of R_0 , as opposed to the other disease histories (compare Figure 6.10b with Figures 6.12b and 6.14b).

Although such heuristic arguments may provide some intuitive understanding of the factors involved in the relationships between these key epidemiological quantities, further work is needed to investigate them more rigorously.

6.3 Summary and conclusions

In order to explore those characteristics of the model about which analytical results are lacking or limited, and as a tool to cross-check analytical and semi-analytical results, an efficient individual-based stochastic simulator has been coded.

The simulator is very flexible, as it allows various different forms of disease history to be considered, and is exact in time, i.e. the time is not discretised in fixed time-steps. Such flexibility is obtained through the use of an event-based approach, in which all events are stored in a time-ordered linked list that is followed as the epidemic progresses and is continuously integrated with new events as soon as they are created.

Various “tricks” are exploited to improve efficiency. The overall result is a complexity which is roughly linear with the total population size. The drawback is that the code re-

quires an intensive memory usage, which limits the maximum allowed population size to a few millions on a common PC (processor: 1.86 GHz; RAM: 1 GB; operating system: Windows XP). The running time for a full epidemic with realistic infectivity parameter values in a population of a million individuals socially structured in households and workplaces is of the order of a few seconds on the same PC.

The simulation provides as outputs:

1. the average dynamics of the epidemic;
2. the values of the basic reproduction number R_0 and the effective reproduction number R_e as they change through time: they coincide during the exponentially growing phase;
3. the real-time growth rate;
4. the average final size;
5. the average sizes of a household and a workplace epidemic;
6. the value of R_H .

A Metropolis-Hastings algorithm has also been coded to allow the investigation of the consequences of relaxing the assumption of a random network connecting households and workplaces, a topic that will be treated in Section 7.2.

In Section 6.2.3 both the analytical results from Section 5.6.3 and the simulation are used to investigate the impact on the real-time growth rate of shifting the infectivity from the workplaces to the community, for various workplace sizes. As R_0 is analytically intractable and is a model output that is difficult to tune *a priori*, the choice adopted is to keep fixed R_H . The size distribution for households and the within-household infectivity R_h are fixed. The main insight gained is that increasing workplace infectivity while decreasing community infectivity (increasing R_w while decreasing R_g) generally reduces the speed of the epidemic, except when the workplace infectivity becomes so strong and the

size of workplaces so small that workplace transmission occurs almost immediately. Such behaviour, however, is not achieved for realistic transmission parameter values.

However, as R_0 represents a more significant concepts for the epidemiological community than R_H , another type of exploration has been performed. The full parameter space of R_g and R_w is explored (transmission in households and household size distribution are fixed), both for workplaces of size 4 and for workplaces of size 30, and the different outputs (r , R_0 and the average final size z) are directly compared. In particular, as the real-time growth rate r is usually one of the first quantities that can be estimated from field data, the exploration focused on the contribution of local saturation effects on the values of R_0 and the average final size for a wide range of values of r , for various assumptions concerning the transmission model.

In particular, the question of how important it is to take into account the social structure is investigated by comparing the real values of R_0 and the average final size z with the analytical estimates obtained when pure homogeneous mixing is assumed in the population.

The conclusion is that, when the average infectivity profile is peaked around the generation time (in particular when the infectivity start from 0 at the time of infection and builds up as time progresses), the estimation of R_0 from r using the Lotka-Euler equation

$$\int_0^{+\infty} \beta(\tau) e^{-r\tau} d\tau = 1,$$

where $\beta(\tau)$ is the average infectivity profile, such that

$$R_0 = \int_0^{+\infty} \beta(\tau) d\tau,$$

turns out to be surprisingly accurate and apparently unaffected by local saturation effects.

The average final size z , computed from the estimated value of R_0 as the solution of

$$1 - z = e^{-R_0 z},$$

is instead very sensitive to local saturation effects and tend to underestimate the true average final size. The underestimation effect is stronger when variability between the total infectivity spread by different individuals is reduced, and the worst situation is obtained when all individuals have the same infectivity profile.

Therefore, for pathogens characterised by a disease history which is unimodal and reasonably peaked, estimation of R_0 obtained when the population structured is ignored is remarkably accurate. If in addition the pathogen is characterised by a relatively low infectivity, as for example is the case of influenza, then saturation effects are limited even in small household and workplaces, and therefore the final size estimated by ignoring the social structure is also reasonably accurate. In general, this seems to be the case for influenza.

Further research is needed on this important topic: in particular, testing more formally the intuitive arguments proposed as possible explanations of the patterns observed in this chapter would be useful in order to address the important questions about the real impact of the social structure on the epidemic spread on a more rigorous basis.

Chapter 7

Application to influenza

The various methodological tools developed in the previous chapters are here applied to the case of influenza. More specifically, the approximate method developed in Fraser (2007) with the purpose of obtaining a numerical estimate for R_* in the households model is here extended to the households-workplaces model, by exploiting the general framework described in Section 5.6. The limited data availability concerning workplaces is reflected in a range of possible values of R_H , obtained for different assumptions on the structure of the population and the transmissibility within workplaces. Then, the Metropolis-Hastings-type algorithm described in Section 6.1.5 is used to explore how sensitive the results are with respect to the unrealistic assumption about random connections between households and workplaces. Finally, the same parameter values considered in the estimation of R_H are then used to offer a description of the possible effects of closing schools.

7.1 Approximate method for the estimation of R_H

The method proposed in Fraser (2007) has been used both in the households and in the households-workplaces model to estimate the real-time growth rate r when all other parameters are known. In this section, instead, the method is used to estimate the household reproduction number R_H for the households-workplaces model when r is observed, as done

in Fraser (2007) for the households model.

The same disease history as the one used in Fraser (2007) is considered, i.e. a *TVI* with a gamma distributed generation time $\omega(\tau)$ (parameter values are reported below, where the method is applied to a concrete example). Under these assumptions, since there is no variability between individuals' total infectiousness, infectious contacts from the same infective are independent of each other and the exact rank attribution ("generations" of infections, but in the more rigorous sense of Section 3.3) can be obtained using the standard Reed-Frost model. Possible extensions of the method can include other disease histories, with random total infectivity and random generation time distributions, which would require the rank attribution to be based on suitable forms of randomised Reed-Frost models.

7.1.1 Upper bound estimation

Recall that, for the households model, the approximation suggested by Fraser tends to provide larger times of occurrence of new cases and therefore a household average global infection rate $\beta_H(\tau)$ that tends to overestimate the times of occurrence of new global contacts. Let $R_* = \int_0^{+\infty} \beta_H(\tau) d\tau$ be the household reproduction number for the households model and let $\omega_H(\tau) = \beta_H(\tau)/R_*$ be the average household generation time distribution. Since the real-time growth rate is given by the Lotka-Euler equation

$$\int_0^{+\infty} \beta_H(\tau) e^{-r\tau} d\tau = R_* \int_0^{+\infty} \omega_H(\tau) e^{-r\tau} d\tau = 1,$$

different functions $\omega_H(\tau)$ in which the weight is increasingly shifted towards large values of τ associate to the same value of R_* smaller values of r or, alternatively, to the same r larger values of R_* . Therefore, the approximate method tends to overestimate the value of R_* . A possible lower bound is discussed in the next section.

For the households-workplaces model, the notation is as follows.

Let $\beta(\tau)$ represents the average global infectivity profile of an infective, let $R_g = \int_0^{+\infty} \beta(\tau) d\tau$ and let $\omega(\tau) = \beta(\tau)/R_g$ be the generation time distribution. By assumption,

the same distribution describes the times at which infectious contacts from an infective are made both within households and within workplaces.

Let γ_H and γ_W be the average rates at which new cases occur in household and workplace epidemics respectively, let μ_H and μ_W be the average final sizes in the household and workplace of a randomly chosen susceptible in the population and let μ_i^H and μ_i^W be the average numbers of cases in generation i of a typical household and workplace epidemic. Recall that, during the exponentially growing phase, multiple reintroductions within the same local group are rare and therefore $\mu_1^H = \mu_1^W = 1$. The primary case is not counted in μ_H and μ_W : therefore $\mu_H = \sum_{i=2}^n \mu_i^H$ and $\mu_W = \sum_{i=2}^n \mu_i^W$. Finally, let $\omega_{[i]}(\tau)$ represents the i -th autoconvolution of the generation time distribution $\omega(\tau)$, where $\omega_{[1]}(\tau) = \omega(\tau)$.

With this notation, recall from Section 5.6.3 that the real-time growth rate is obtained by considering the operator

$$K_r = \begin{pmatrix} \left(\sum_{i=2}^{n_H} \frac{\mu_i^H}{R^{i-1}} \right) \left(\sum_{j=2}^{n_W} \frac{\mu_j^W}{R^{j-1}} \right) & \left(1 + \sum_{i=2}^{n_H} \frac{\mu_i^H}{R^{i-1}} \right) \left(\sum_{j=2}^{n_W} \frac{\mu_j^W}{R^{j-1}} \right) \\ R_g \left(\sum_{i=1}^{n_H} \frac{\mu_i^H}{R^i} \right) & R_g \left(\sum_{i=1}^{n_H} \frac{\mu_i^H}{R^i} \right) \end{pmatrix}$$

where

$$R = \frac{1}{\int_0^{+\infty} \omega(\tau) e^{-r\tau} d\tau}$$

represents the basic reproduction number as estimated from the real-time growth rate in absence of the social structure. Given that r is observed from data and that $\omega(\tau)$ and the within-household and within-workplace transmission coefficients are assumed to be known (together with the household and workplace size distributions), the only unknown in K_r is given by R_g . Requiring that $\rho(K_r) = 1$ gives an implicit condition for R_g .

Lemma 7.1.1. *Assume M is a 2-by-2 matrix of the form*

$$M = \begin{pmatrix} a & b \\ xc & xd \end{pmatrix}, \quad (7.1.1)$$

with $d > 0$, and assume that $\det M \leq 0$. Then the dominant eigenvalue of M is an increasing function of x .

Proof. The dominant eigenvalue of M , $\rho(M)$, is given by

$$\rho(M) = \frac{1}{2} \left[(a + xd) + \sqrt{(a + xd)^2 - 4x \det M} \right].$$

The conditions $d > 0$ and $\det M \leq 0$ imply that $\rho(M)$ increases monotonically with x . $\square$

Theorem 7.1.2. *The solution $R_g = \overline{R}_g$ of the equation $\rho(K_r) = 1$ exists and is unique. In addition, $\overline{R}_g$ is strictly positive if and only if*

$$\left(\sum_{i=2}^{n_H} \frac{\mu_i^H}{R^{i-1}} \right) \left(\sum_{j=2}^{n_W} \frac{\mu_j^W}{R^{j-1}} \right) < 1. \quad (7.1.2)$$

Proof. K_r has the same structure as in Lemma 7.1.1, with $R_g = x$, $c = d > 0$ and $b > a$, which implies $\det M = ad - bc = c(a - b) < 0$. Therefore $\rho(K_r) = \rho(K_r(R_g))$ is a continuous and monotonically increasing function of R_g . Given that $\lim_{x \rightarrow -\infty} \rho(K_r(x)) = -\infty$ and $\lim_{x \rightarrow +\infty} \rho(K_r(x)) = +\infty$, the solution of $\rho(K_r(R_g)) = 1$ exists and is unique.

Furthermore, when $R_g = 0$,

$$\rho(K_r(0)) = \left(\sum_{i=2}^{n_H} \frac{\mu_i^H}{R^{i-1}} \right) \left(\sum_{j=2}^{n_W} \frac{\mu_j^W}{R^{j-1}} \right),$$

and therefore $\overline{R}_g$ is strictly positive if and only if condition (7.1.2) is satisfied. $\square$

If condition (7.1.2) is not satisfied, the solution $\overline{R}_g$ is negative and therefore not acceptable. This reflects the fact that not all values of r are compatible with the assumptions concerning within-household and within-workplace infectivities (and the household and workplace sizes). In fact, it is known from Section 2.3.4 that $r > 0 \Leftrightarrow R_H > 1$. Assume now that $R_g = 0$: then $R_H = \mu_H \mu_W = \rho(K_0(0))$. For a sufficiently strong infectivity in households and workplaces, $R_H > 1$ even if there is no global transmission (this is the case when the average clump size is infinite). In this case, $\rho(K_r(0)) = 1$ is solved by a positive value of r , which represents a lower bound under which no value of r can correspond to a positive value of $\overline{R}_g$.

Once $\overline{R}_g$ is found, R_H is then computed using Equation (4.3.4).

The procedure just described cannot be expressed in a closed formula as nicely as in Fraser (2007), where the particularly simple expression of R_* for the households model

allows direct computation of R_* without requiring an intermediate value for R_g , which cancels out from the main formula. In the present case, instead, a two-type next generation matrix approach involves computing the eigenvalues of a matrix and, in addition, the complex Equation (4.3.4) for R_H is not nicely factorised.

7.1.2 Lower bound estimation

Neglecting the effects of repeated contacts between the same individuals and of overlapping generations results in an overestimation of R_H . In order to control such an approximation, Fraser (2007) suggests a lower bound estimate obtained when all non-primary cases in a local epidemic are lumped in the second generation (see also Section 5.6.1).

The same procedure as above is repeated, but a worse performance than that found in Fraser (2007) is expected for two main reasons: first, chaining two local epidemics where all non-primary cases are lumped in the second generation is worse than doing it only once; second, a workplace (e.g. a class) can be larger than a household, and thus the average final size can be much larger too: therefore, placing all cases in the second generation leads to drastically different dynamics, much more than what would occur for a smaller local structure, as it is the case for a household.

However, Fraser (2007) showed that, in the case of influenza, the upper approximation is much closer to the exact value than the lower bound, because influenza is not particularly infectious. For this reason, and because the lower bound obtained with this method is excessively low, it is arguably not worth bracketing the solution between two bounds and accept instead the upper bound as the approximate result. Although such a decision is not completely satisfying, it must be first noted that more precise results would still be only indicative, given the lack of information concerning workplace transmission, and second that the large local structures, which are responsible for an excessively low lower bound, have at the same time the effect of weakening the saturation effects (repeated contacts between the same infector and infectee are rare) and therefore making the upper bound much closer to the real value of R_H .

Finally, although the lack of a lower bound sufficiently close to the upper bound weakens the predictive power of such an approach, nevertheless the method provides a range of reasonable values for R_H when both households and workplaces are taken into account. This is invaluable because R_H provides information about how much between-household transmission needs to be blocked in order to stop an epidemic (for example providing a secure vaccination coverage, see Section 4.5), and not only there are no estimates of it in the literature, when households and workplaces are taken into account, but there is even no intuitive range of possible values for it.

7.1.3 The case of influenza

Any possible application of this method to a practical example is likely to be very artificial, given the lack of data concerning workplaces. Some data are available about schools, though, and it is reasonable to assume that classes are larger than workplaces (interpreted as small groups, e.g. offices, for which homogeneous mixing can be assumed; see comments in Section 6.2.1) and contact rates between children are larger than between adults.

Possible numerical values of transmission rates for influenza in schools are obtained from Cauchemez *et al.* (2008): the authors assume a disease history of type *TVI* with the form $\beta(\tau) = \beta f(\tau)/n$, where $f(\tau)$ is the survival function of the same *gamma* of Section 6.2.1 (obtained from Cauchemez *et al.*, 2004, to which they refer), n is the size of the workplaces and values of β are estimated in different conditions in Table SI5 and Table SI10 to be 0.22 day^{-1} and 0.28 day^{-1} (Cauchemez *et al.*, 2008). Since in Cauchemez *et al.* (2004) the average duration of the infectious period is assumed to be of 3.8 days, these values correspond roughly to $R_w = 0.84$ and $R_w = 1.1$, respectively. Note how these values are just around the threshold 1 and therefore, despite being close to one another, their effects can be considerably different.

The disease history is taken from Fraser (2007) to be of type *TVI*, where there is no variability across individuals and, after infection, the generation time distribution $\omega(\tau)$ follows a $\Gamma(\alpha, \delta)$ distribution with $\alpha = 9.39$ and $\delta = 3.29$ (mean 2.85 and standard deviation

0.93). These values are here approximated to $\alpha \approx 9$ and $\delta \approx 3.16$, so that α is integer and the generation time T_g is fixed to 2.85.

The size distribution of households and the within-household infectivity are again taken according to the influenza data presented in Fraser (2007): the household size distribution is that of the UK census data (2001) and $R_h = 1.35$: the average epidemic size in the household of a randomly selected individual including the primary case ($1 + \mu_H$) is then equal to 1.87.

Further assumptions are unavoidably strong:

1. The “workplace” of a child is the class, where homogeneous mixing can be reasonably assumed, and not the entire school.
2. However, estimated transmission rates in schools are assumed for the transmission in classes.
3. Classes are taken to be of 20 students: although they could be larger, it is numerically challenging to compute generations of infectives in the Reed-Frost model with populations larger than 20.
4. Roughly 20% of the population goes to school and the other 80% to workplaces (according to the UK National Statistics, 2005, roughly 10 million individuals in the UK are less than 16 years old, out of a total population of 50 millions).
5. The same transmission rate in schools occurs also in workplaces.
6. A lower bound for the effective workplace size is obtained when $n_W = 1$, i.e. there is no pure workplace transmission but only school transmission. Note that in this case, since there is no pure workplace transmission, R_g is expected to be higher than if $n_W > 1$, since it effectively contains also the transmission that is likely to occur at work.

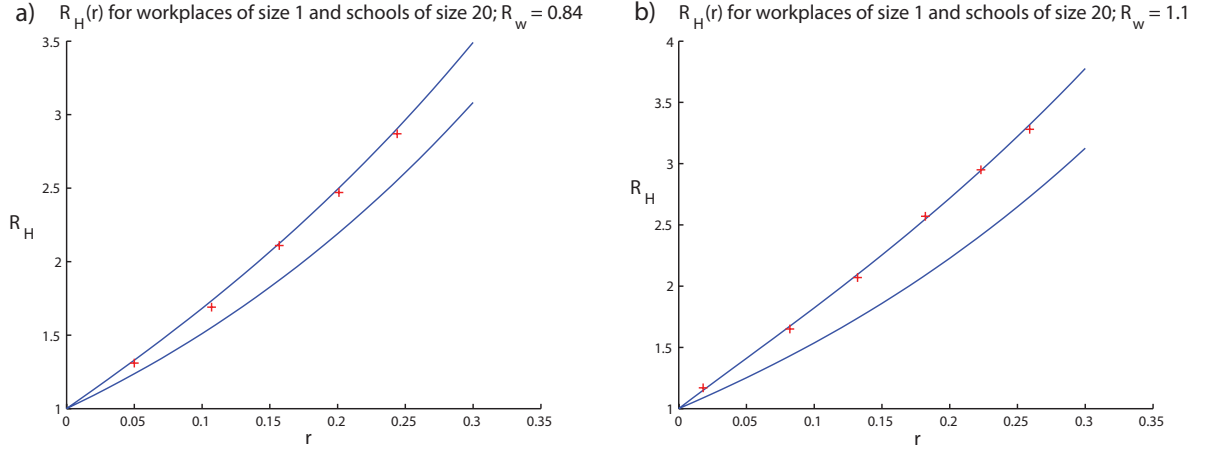


Figure 7.1: The household reproduction number R_H as a function of the real-time growth rate r , obtained when 20% of the population (children) go to workplaces of size 20 (school classes) and the remaining 80% (adults) go to workplaces of size 1. The blue lines represent the main approximation described in the text (top) and the lower bound approximation obtained by lumping all non-primary cases in the second generation (bottom). The red crosses are the result of the individual-based stochastic simulation. The household size distribution is obtained from the UK census data (2001) and the within households infectivity is given by $R_h = 1.35$. In *a*), $R_w = 0.84$ and the simulation results are obtained, for increasing r , for values of R_g of 0.4, 0.6, 0.8, 1 and 1.2; in *b*), $R_w = 1.1$ and the simulation results are obtained, for increasing r , for values of R_g of 0.2, 0.4, 0.6, 0.8, 1 and 1.2. Note that the main approximation should be an upper bound, but this is not always apparent as it is very close to the real value and the precision in reading the outcome of the stochastic simulation is not particularly high.

7. Another possible value for the workplace size is taken to be $n_W = 4$. This is probably an upper bound, given that the effective size of a household appears to be lower (the average household size of a randomly selected individual is roughly 3; see Fraser, 2007) and that a realistic workplace transmission is likely to be lower than the assumed school-like transmission. A much lower value of R_g is expected in this case.

Figures 7.1, 7.2 and 7.3 show the approximated value of R_H as a function of r computed using a Matlab built-in iterative method (the function *fzero*), together with the less useful lower bound described before.

In Figure 7.1*a*, R_w is taken to be 0.84, while in Figure 7.1*b*, $R_w = 1.1$. In both cases school classes have size 20 and workplaces have size 1. From Figure 7.1*a*, a realistic real

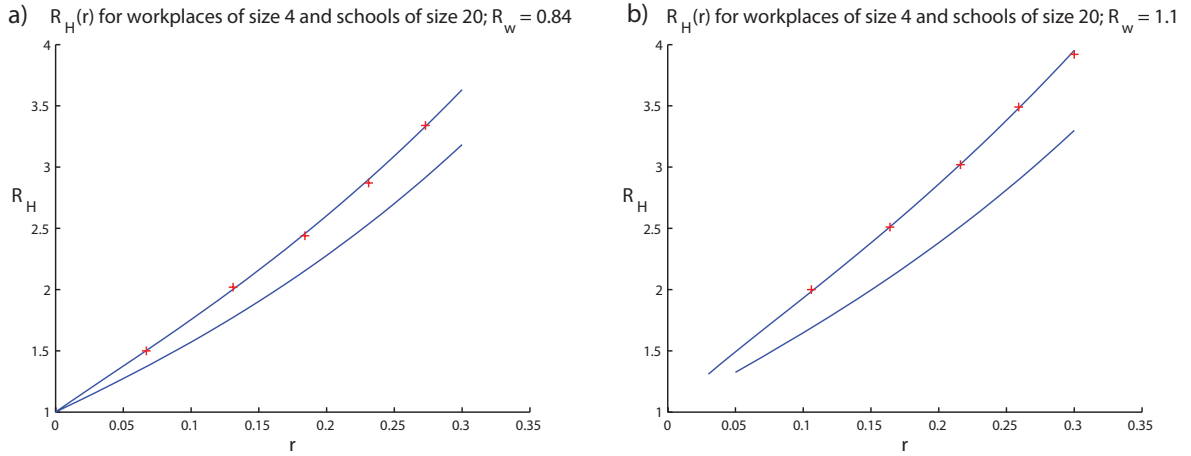


Figure 7.2: The household reproduction number R_H as a function of the real-time growth rate r , for $R_w = 0.84$ (a) and $R_w = 1.1$ (b), obtained when 20% of the population (children) go to workplaces of size 20 (school classes) and the remaining 80% (adults) go to workplaces of size 4. The blue lines represent the main approximation described in the text (top) and the lower bound approximation (bottom). The red crosses are the result of the individual-based stochastic simulation and are obtained, for increasing r , for values of R_g of 0.2, 0.4, 0.6, 0.8 and 1 in both cases. The household size distribution is obtained from the UK census data (2001) and the within households infectivity is given by $R_h = 1.35$. Note that r cannot decrease below a certain level, given that $R_g \geq 0$ (see text).

time growth rate $r = 0.2$ for influenza (Fraser, 2007; Grassly & Fraser, 2008, and references therein) gives a value of $R_H = 2.49$ and a corresponding value of $R_g = 1.1$ out of an R_0 of 1.74 (not shown), i.e. global infections count for more than 60% of the total transmission of an infective. From Figure 7.2b, the same $r = 0.2$ leads to $R_H = 2.71$ and $R_g = 0.9$ out of the same value of R_0 (the same r corresponds to the same R_0 , as shown in Section 6.2.4 for the *TVI* disease history), giving a contribution of the global infections to R_0 of more than 50%.

Figure 7.2 reports the same analysis in the case of workplaces of size 4, for the same values of R_w used before. The outcomes of the stochastic simulation (red crosses) are obtained from values of R_g of 0.2, 0.4, 0.6, 0.8 and 1 in both cases.

For $r = 0.2$, the values of R_H obtained in Figure 7.2a and 7.2b are, respectively, $R_H = 2.60$ and $R_H = 2.86$, with corresponding values of R_g of 0.71 and 0.54, i.e. of 40%

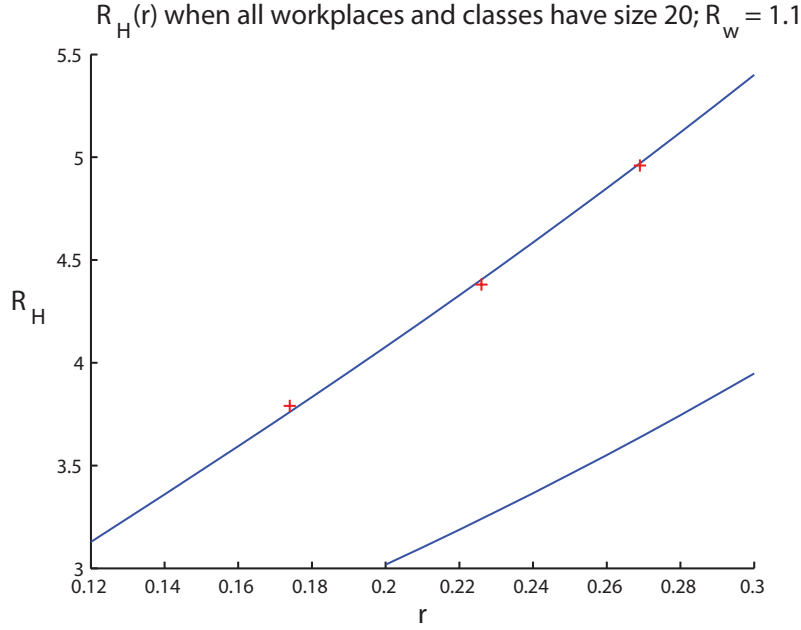


Figure 7.3: The household reproduction number R_H as a function of the real-time growth rate r , for $R_w = 1.1$, obtained when all workplaces have size 20 (population of children). The blue lines represent the main approximation described in the text (top) and the lower bound approximation (bottom). The red crosses are the result of the individual-based stochastic simulation and are obtained, for increasing r , for values of R_g of 0.2, 0.4 and 0.6. The household size distribution is obtained from the UK census data (2001) and the within households infectivity is given by $R_h = 1.35$ (see text). Note the limited predictive power of the method due to presence of many large workplaces.

and less than 30% of the same $R_0 = 1.74$. Separating the transmission out of school in workplace and pure community transmission, apart from substantially decreasing R_g as expected, associates to the same r a larger value of R_H .

As an extreme case and to show how the lower approximation can be become hardly informative to be useful in the cases of large workplaces, the same analysis as before is reported in Figure 7.3 for $R_w = 1.1$ and assuming a full population of children, i.e. in which all individuals have workplaces of size 20 with school-like transmission rates. The red crosses are obtained from values of R_g of 0.2, 0.4 and 0.6.

For $r = 0.2$, a value of $R_H = 4.1$ is obtained, and $R_g = 0.35$ represents only 20% of the total transmission of an infective.

7.2 Spatial constraints on the social structure

7.2.1 Motivation

The model of Chapter 4 is based on the unrealistic assumption that households and workplaces are connected at random, which in a large population guarantees a negligible clustering coefficient (see e.g. Keeling, 1999) for the network connecting local structures. This assumption implies a worst case scenario, where the probability of reintroduction of the infection in the same local structure multiple times is minimised. Without this assumption, the model becomes mathematically intractable.

It is however interesting to explore the implications of this assumption for the spread of the disease and this investigation can be performed by exploiting the individual-based stochastic simulator.

The rewiring algorithm used to introduce some degree of spatial constraint in the network is described in Section 6.1.5.

In the present section, the random social structure is modified by attributing a position to each household and each workplace and by forcing individuals to go to work close to where they live. This type of spatial constraint on the distance travelled by each individual has been considered before (e.g. Riley & Ferguson, 2006) and is motivated by the availability of data concerning social commuting patterns (although these data are not used here, given the theoretical character of this exploration).

The effect of the different social structures on the spread of the infection is then analysed.

7.2.2 The social structure

The population is assumed to live in a unit square, and households and workplaces are randomly scattered on the available surface, so that the density is uniform on the square and the maximum possible distance travelled by an individual is $\sqrt{2}$ (along the diagonal). When individuals select their workplace at random, the commuting patterns are described

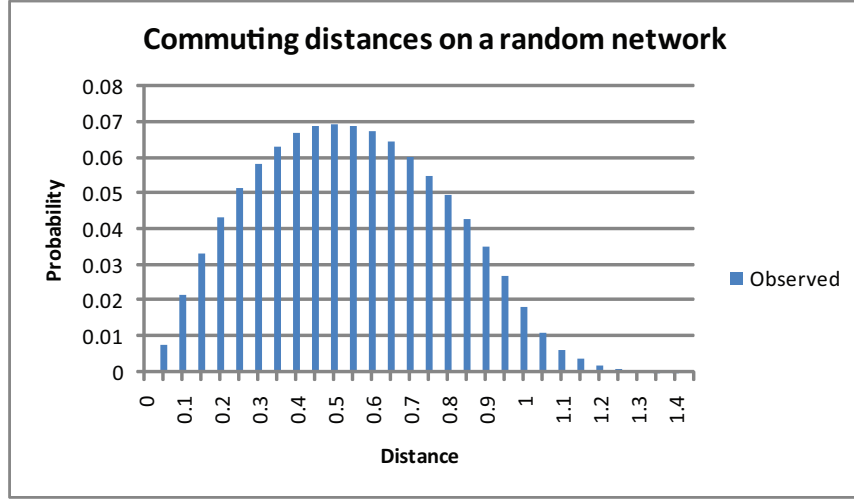


Figure 7.4: Discretised probability distribution for the commuting distance between the household and the workplace of a randomly selected individual in a population of a million individuals, when the households-workplaces network is random. The “city” is assumed to be a square of size 1, the discretisation step is taken to be $\Delta d = 0.05$.

in Figure 7.4 and well approximate the distribution of distances between randomly selected points on the square (not shown).

The same epidemic of the previous section, when classes are assumed to have size 20 and workplaces to have size 4, is considered. The parameters are recalled to be $R_w = 1.1$, $R_g = 0.54$ and the usual household size distribution from the UK census data (2001) is used with $R_h = 1.35$; the value of R_g is selected to obtain $r = 0.2$, and the other associated outputs are $R_0 = 1.74$, $R_H = 2.86$ and a final size of 78%.

The distance kernel that is used to introduce spatial constraints on social structure (see sec 6.1.5) is assumed to be a power-law function of the form

$$k(d) = \frac{1}{1 + \left(\frac{\delta}{d}\right)^\alpha}, \quad (7.2.1)$$

(as in Riley & Ferguson, 2006). The distribution is represented in Figure 7.5 for $\delta = 0.5$ and $\alpha = 3$, together with the observed commuting pattern after having rewired the social structure. The parameter α controls how flat the distribution is for small distances, how

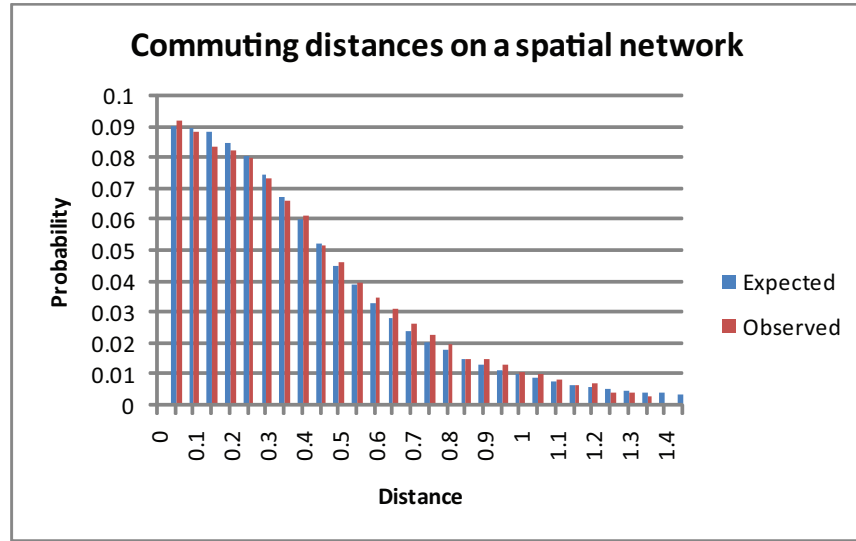


Figure 7.5: Distribution for the commuting distance between the household and the workplace of a randomly selected individual in a population of a million individuals, when the households-workplaces network has been rewired using the Metropolis-Hastings algorithm described in Section 6.1.5. The blue bars represent the target distribution to match (Equation (7.2.1) with $\delta = 0.5$ and $\alpha = 3$), while the red bars represent the distribution observed in the synthetic population after the rewiring process. The choice of the target distribution is purely exemplificative.

steeply it decreases in the middle and how quickly it goes to zero for large distances. The parameter δ controls how large the “bell” shape (obtained by rotation of the distribution around the axis $d = 0$) is and therefore decides how “localised” the kernel is. As α increases, the distribution becomes more and more similar to the step function

$$k(d) = \begin{cases} 1 & \text{for } d < \delta \\ 0 & \text{for } d > \delta. \end{cases}$$

Note the limited precision obtained in particular for very small and very large distances to the expected values.

As workplaces are uniformly distributed on the square and their number is m_W , for a selected δ , an individual sees around roughly $m_W \pi \delta^2$ workplaces to choose from. If all workplaces have size 20, $m_W = 50000$ and, for $\delta = 0.5$ the effective *workplace neighbourhood size* is about 40000.

For $\delta = 0.1$, 0.05 and 0.01, the effective workplace neighbourhood size is 1500, 400 and

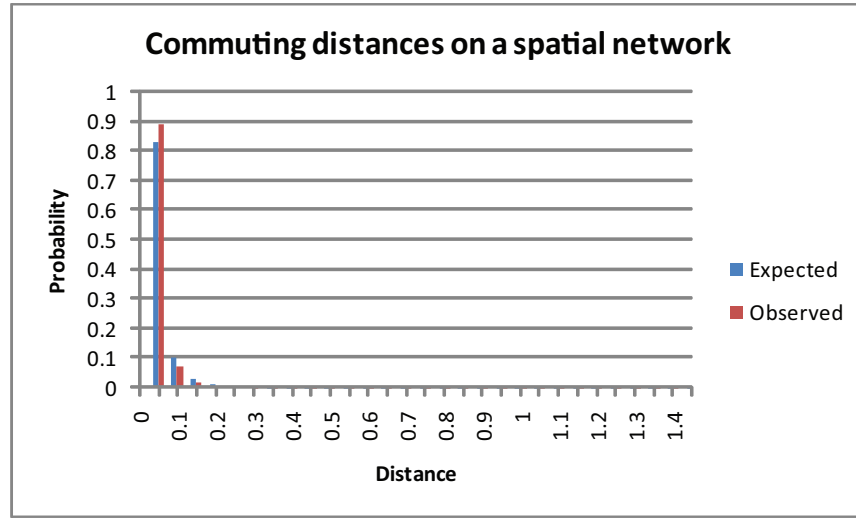


Figure 7.6: Expected and observed distributions for the commuting distance between the household and the workplace of a randomly selected individual in a population of a million individuals, after the rewiring process. The target distribution (Equation (7.2.1) with $\delta = 0.01$ and $\alpha = 3$) has been selected to be as much localised as possible without excessively compromising the convergence of the rewiring method.

15 respectively. Therefore, for smaller δ , convergence of the rewiring process is compromised as the small variations in the density of workplaces may generate situations where individuals see too few or even no workplaces sufficiently close around.

Figure 7.6 represents the commuting patterns obtained for $\delta = 0.01$ and $\alpha = 3$ (the power α is roughly in agreement with Riley & Ferguson, 2006), which are markedly more localised than those obtained in the random network.

However, when an epidemic is run on this new network with the same parameters as before, the overall spread of the epidemic is exactly the same as before: $r = 0.2$, $R_0 = 1.74$, $R_H = 2.86$ and a final size of 78%. The daily incidence at the peak also remains the same, although the peak can be delayed by a couple of days. Figure 7.7 compares the outcomes for r and R_H in the two cases.

Although the overall dynamics are not affected, the spatial spread may be significantly different, as the very localised spread between households and workplaces is expected to result in a “wave-like” front of infection. However, the additional presence of global infections tends to seed the infection in other points of the square and, if R_g is sufficiently

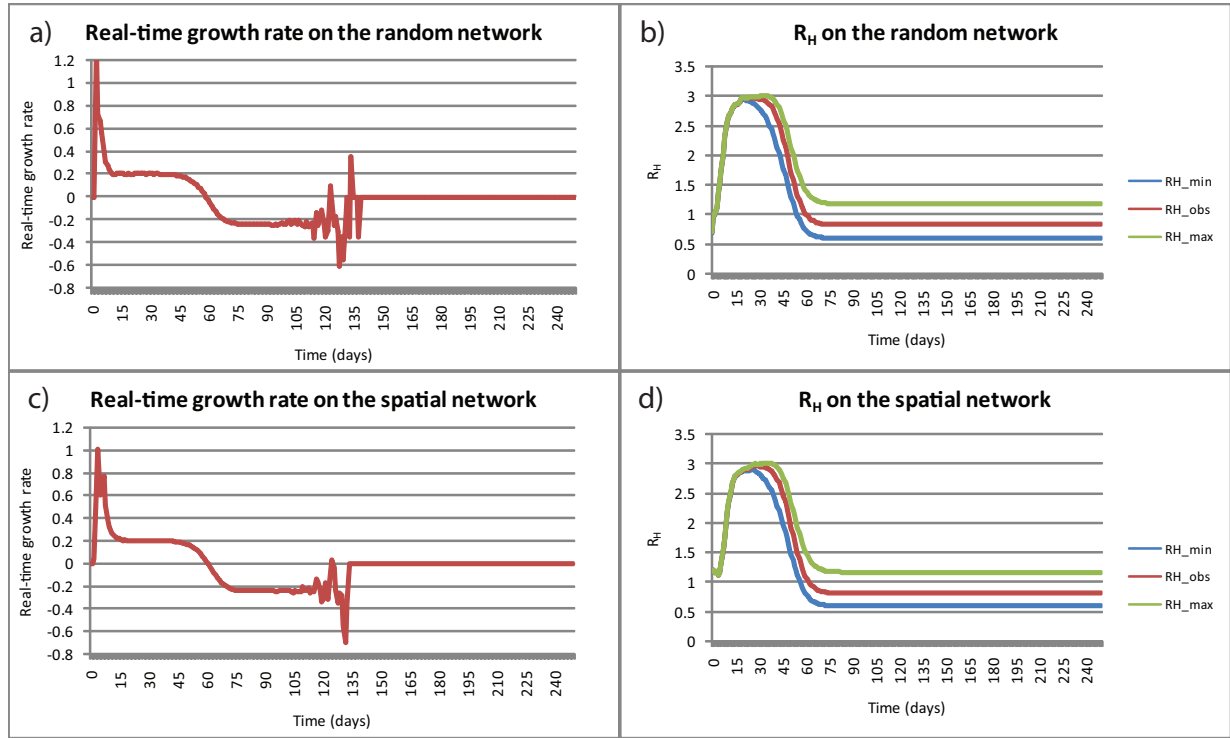


Figure 7.7: The real-time growth rate r (a and c) and the household reproduction number R_H (b and d) plotted as a function of time: the values stabilise on the required value during the exponentially growing phase; for R_H , three outputs are computed and they should coincide during the exponentially growing phase. Note that, except from small differences due to stochastic variations, the two outputs are indistinguishable. The infectivity parameters are: $R_h = 1.35$, $R_w = 1.1$ and $R_g = 0.54$; the households size distribution is taken from the UK census data (2001) and 20% of the population (children) go to workplaces of size 20 (school classes) and the remaining 80% (adults) go to workplaces of size 4. The output is an average over 100 simulations in a population of a million individuals.

high, such additional long-range connections may make the spatial spread indistinguishable from that on a random network. This would easily justify the observation of the same outcome on the random and spatially constrained network.

For this reason, an epidemic is simulated again on both networks, with parameters $R_h = 1.35$, $R_w = 1.5$ and $R_g = 0$ (R_w has been increased in order to observe large epidemics with high probability even when $R_g = 0$). With no global contacts, the infection on the spatially constrained network spreads in waves, but the overall dynamics are the almost the same anyway: $r = 0.08$, $R_0 = 1.25$, the final size is 51% in both cases and the

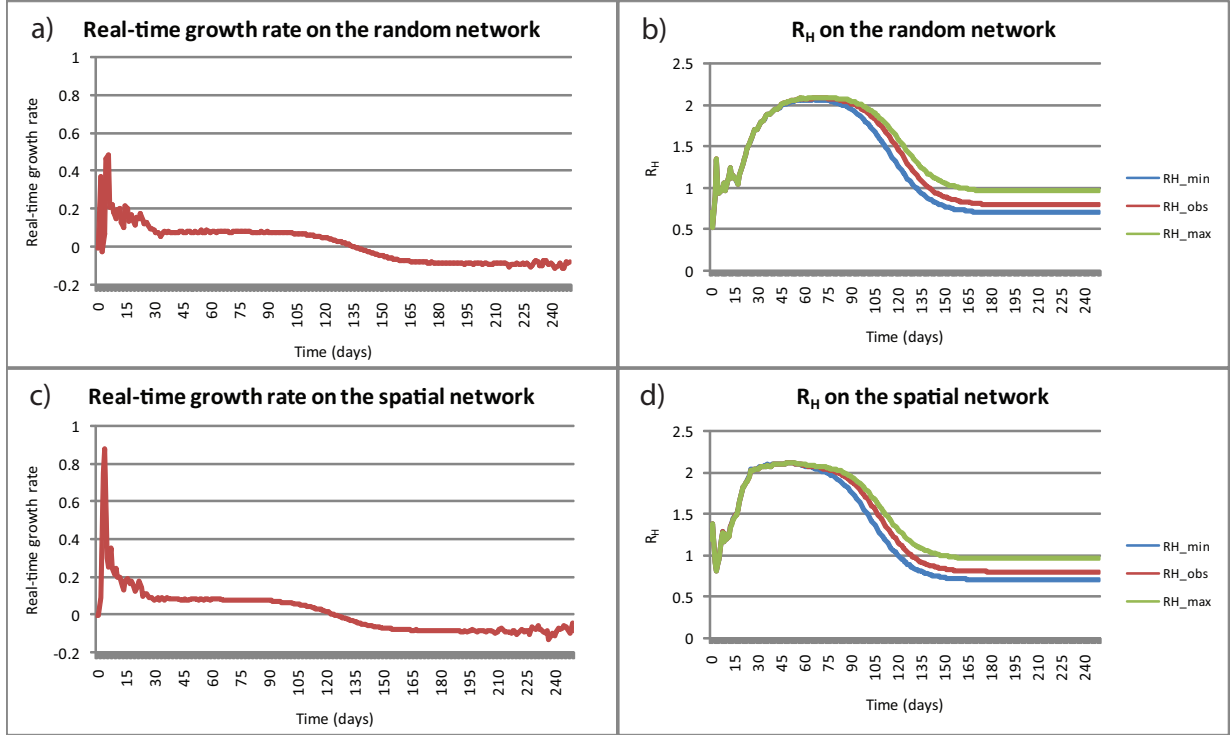


Figure 7.8: The real-time growth rate r (a and c) and the household reproduction number R_H (b and d) plotted as a function of time: the values stabilise on the required value during the exponentially growing phase; for R_H , three outputs are computed and they should coincide during the exponentially growing phase. Note that, except from small differences due to stochastic variations, the two outputs are indistinguishable. The infectivity parameters are: $R_h = 1.35$, $R_w = 1.5$ and $R_g = 0$; the households size distribution is taken from the UK census data (2001) and 20% of the population (children) go to workplaces of size 20 (school classes) and the remaining 80% (adults) go to workplaces of size 4. The output is an average over 100 simulations in a population of a million individuals.

daily incidence at the peak is of 1%; R_H experiences a negligible increase from 2.08 to 2.11 from the random to the spatially constrained network, while the peak time (with a million individuals) is delayed by a couple of days.

The comparison is reported in Figure 7.8. Even in this last numerical simulation, where long-range global infections do not occur, no detectable difference appears between the overall dynamics of the epidemic spread on the random and on the spatially rewired network. The conclusion is that, for parameter values characteristic of influenza, the spatial

spread may be different, but spatial constraints are insufficient to detectably modify the overall dynamics of the epidemic. For the same reason, this results would also suggest that it is not possible to obtain spatial information about the social network and the disease spreading on it by only looking at the overall dynamics.

Other parameter values have not been explored and may lead to different results in general. However, no detectable difference is seen in Figure 7.8 between the epidemics on the random and the spatially rewired networks (when no long-range global transmission can occur). Therefore, even if other slightly different parameter values were able to generate some visible difference when $R_g = 0$, it appears improbable that such a difference would remain visible when realistic casual long-range global transmissions are included.

7.3 School closure

One of the main reasons for disentangling the school/workplace transmission from other routes is that it is possible to explore the effects of closing schools during an outbreak. However, the effectiveness of such a control policy is highly dependent on the numerical value of model parameters. Quantitative estimates from epidemiological data are very limited, though, and this explains the relatively wide range of effects predicted by different studies (see Cauchemez *et al.*, 2008, and citations therein). A further complication derives from the behavioural change subsequent to the introduction of the control policy, which is unknown beforehand. Forms of compensatory behaviour consisting in a wide range of increased within-household and community transmission rates have been explored in Cauchemez *et al.* (2008), where numerical values are inspired by estimated transmission rates during school and holiday periods. It is however debatable how similar the social behaviour during holiday period is to that during a pandemic context.

In this section, the effect of closing schools is studied for two social structures considered in the previous section (and corresponding parameter values), in order to highlight how sensitive the effect of this control policy is to the particular assumptions concerning transmission in workplaces and in the community.

With the insight provided by the previous section, the effects of school closure are studied on the bipartite random network, without any spatial constraint imposed on it. Even if there were a sensible difference between the random and the rewired network, the use of the random network would assure that the worst possible scenario is considered and therefore that the strongest likely impact of school closure is estimated.

Note that the households-workplaces model does not distinguish between adults and children, and therefore between schools and workplaces. It might therefore appear that the model is not particularly suited for such an analysis. However, school closure often means that at least one of the parents needs to stay at home to take care of the children, thus reducing the workplace transmission too. Therefore, although very idealised, the model might be more representative about reality than what would seem at first.

What the model is unable to capture, though, is the assortativity due to the fact that larger households tend to contain more children and therefore that school classes, which are much larger than offices, are preferentially connected to larger households, as they tend to contain more children.

As before, the disease history is taken from Fraser (2007) to be of type *TVI*, where there is no variability across individuals and, after infection, the generation time distribution $\omega(\tau)$ follows a $\Gamma(\alpha, \delta)$ distribution with $\alpha \approx 9$ and $\delta \approx 3.16$, so that the generation time is $T_g = 2.85$. The size distribution of households and the within-household infectivity are the same as before.

As in the case with higher school transmission rates of previous section, R_w is taken to be 1.1 (Cauchemez *et al.*, 2008) and the effect of school closure is studied in the case of classes of size 20 and both when workplaces have size 1 and when they have size 4.

7.3.1 School closure only

When workplaces have size 1, only pure school transmission is disentangled from the global transmission, which therefore contains also adult transmission occurring in workplaces. In

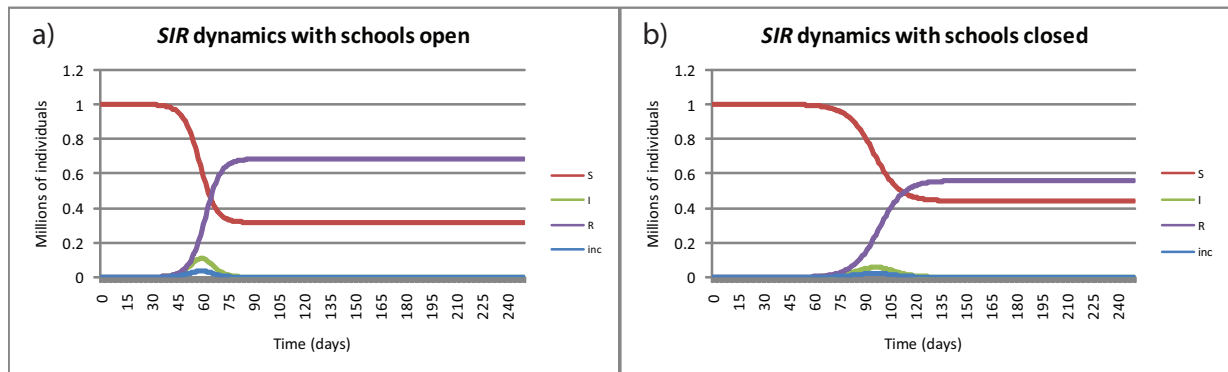


Figure 7.9: Average dynamics of the households-workplaces model. The output describes the average number of susceptibles (S , red), infectives (I , green), recovered (R , violet) and the incidence (inc , light blue) as a function of time (in days). The infectivity parameters are: $R_h = 1.35$ and $R_g = 0.9$; In a), school transmission occurs with $R_w = 1.1$, while in b) schools are closed, i.e. $R_w = 0$. The households size distribution is taken from the UK census data (2001) and 20% of the population (children) go to workplaces of size 20 (school classes) and the remaining 80% (adults) go to workplaces of size 1. The output is an average over 100 simulations in a population of a million individuals.

the previous section, for $r = 0.2$, R_H was estimated to be 2.72 and R_g to be 0.9 out of an R_0 of 1.74 (see Figure 7.1b).

Figure 7.9a presents the average epidemic dynamics for these parameter values. The final size is of about 68% of the population while the peak of the epidemic is obtained (in a population of a million individuals) between 50 and 62 days after the beginning of the epidemic (the lower bound does not appear in the figure, as epidemics are synchronised at the peak: the longest epidemic starts at time 0, while the other ones later on).

School closure is then simulated by setting $R_w = 0$. This is assumed to happen before the start of the epidemic, in order to maximise its effect, which would otherwise be reduced by delays in policy implementation. No compensatory behaviour is assumed, again with the aim of maximising the policy effect. The new dynamics are reported in Figure 7.9b.

A noticeable slowdown in the spread is observed, moving the peak from between days 50 and 62 to between days 77 and 100 (as before, lower bounds are not shown in the figure), and reducing the daily incidence at the peak from 3.5% to 1.8% of the population. This is accompanied by a reduction of the real-time growth rate from 0.2 to 0.12, of R_0 from 1.74

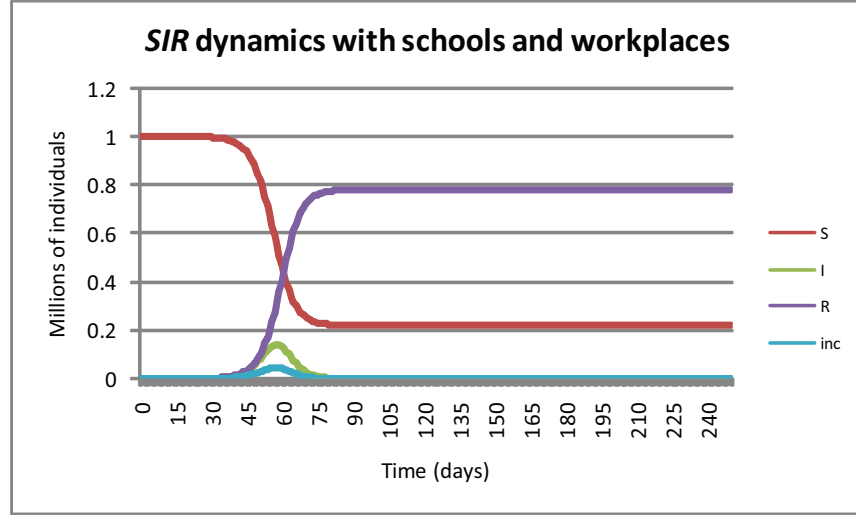


Figure 7.10: Average dynamics of the households-workplaces model when 20% of the population (children) go to workplaces of size 20 (school classes) and the remaining 80% (adults) go to workplaces of size 4. The infectivity parameters are: $R_h = 1.35$, $R_w = 1.1$ and $R_g = 0.9$. The output describes the average number of susceptibles (S, red), infectives (I, green), recovered (R, violet) and the incidence (inc, light blue) as a function of time (in days). The households size distribution is taken from the UK census data (2001) and the output is an average over 100 simulations in a population of a million individuals.

to 1.4 and of R_H from 2.72 to 1.68. However, the reduction in the final size remains quite limited, from 68% to 56% and might appear to be insufficient to accept the economical implications of such a control policy, especially since its implementation is likely to achieve a smaller result because of logistic delays and compensatory behaviours.

7.3.2 School and workplace closure

The same analysis is repeated in a population with workplaces of size 4. In this case, both school and workplace transmissions are disentangled from the community transmission, resulting in much lower values of R_g . In this case, school closure is associated with workplace closure, thus simulating the need for adults to stay at home to take care of their children.

Figure 7.2b reports the baseline scenario of no control policy, for the same parameter values as before and a real-time growth rate $r = 0.2$. In this case, $R_g = 0.54$ and $R_H = 2.86$ and the final size is of 78%. The peak occurs at day 57 with a daily incidence at the peak

of 4.3%.

This time, the control policy is almost sufficient to halt the epidemic spread, causing a reduction of R_H to 1.01. This highlights the drastic difference in the effect of school closure when it is accompanied by a simultaneous stop in workplace transmission. Despite all the limitations of this approach and the quantitative uncertainties associated with the lack of realistic data, the argument highlights how limited effects in pure school closure could be enhanced by the social distancing effects indirectly caused by it. This makes clear estimation of school closure efficacy even more difficult as highly dependent on other transmission factors which are difficult to control quantitatively.

7.4 Summary and conclusions

In this chapter, the approximate method presented in Fraser (2007) for the estimation of R_* from the real-time growth rate in the households model is extended to the households-workplaces model. Although the application of the method to influenza is not reliable from a quantitative point of view, as it is based on limited numerical data and some rather strong assumptions, it nevertheless provides some indicative numerical values for the household reproduction number R_H , ranging from 2.5 to 2.86 for educated guesses of within school transmission and arguably reasonable workplace and school sizes. A wider but safer interval of values ranges from $R_H = 2.26$, as estimated in Fraser (2007) for the pure households model, and R_H as high as 4 when the whole population mixes according to a children-like behaviour.

These estimates, however, do not take into account the form of assortativity in the social structure which results from the fact that larger households tend to contain more children and therefore that school classes, which are typically much larger than offices, are preferentially connected to larger households. Since in the case of multitype models it is recognised that increasing assortative mixing between individuals is generally associated to a larger value of R_0 , an increased assortative mixing at the level of local structures may lead to an increase in the epidemic severity. However, if children living in the same household

go to the same school or class, the infection is reintroduced multiple times in the same local structures thus slowing down the spread and reducing the final size (in the terminology of Section 4.2.3, the network linking households and workplaces is not random, and presents many cycles of size 2). The overall effect of this assortativity is therefore unclear, and further research should focus on including it in the model by distinguishing between adults and children.

In any case, it is important to note how the method shows that the presence of workplaces in the model results in larger values of R_H compared to the simpler household model (Fraser, 2007), thus highlighting the risk of underestimating the effort required to stop the epidemic by vaccinating households when the presence of workplaces is neglected.

The Metropolis-Hastings rewiring process described in Section 6.1.5 allowed the investigation of the likely impact of the simplifying assumptions that households and workplaces are linked randomly. For parameter values reasonable for influenza, the conclusion is that even in the case of very localised commuting patterns, pure spatial constraints on the social structure are insufficient to affect the overall dynamics of the epidemic. The result may be surprising, but some intuitive insight can be gained by referring to simpler models of epidemics on networks (Keeling, 1999; Keeling & Eames, 2005; Keeling & Rohani, 2008).

It is well known that an infection spreading through a network with a positive *clustering coefficient* (e.g. see Keeling & Eames, 2005; Keeling & Rohani, 2008) can manifest the lack of a pure exponential phase, given that small loops in the network may cause repeated contacts towards already infected individuals, and that therefore the concept of reproduction number is not necessarily well-defined or particularly useful for these models.

In the model studied here, the usual concept of clustering (Keeling, 1999) does not apply directly as it is defined as the fraction of triplets that form a triangle. When the network is bipartite, only loops of even size can occur: a loop of size 2 corresponds to two individuals sharing the same households and working in the same place, a loop of size 4 corresponds to two individuals living in the same household, who work in different places but who have work colleagues sharing the same household themselves, and so on.

Although local saturation occurs at the individual level (i.e. within households and workplaces), when households and workplaces are connected at random and the population is large, the resulting bipartite network shows no clustering and appears locally as tree-like. With the spatial rewiring, it is reasonable to expect a certain amount of clustering to be introduced: it is still unclear how much, though. It could be that the amount of loops has not increased considerably, thus providing a sufficient explanation for the similarities observed above in the epidemic dynamics before and after the rewiring process.

However, another explanation is possible. The rewiring process could have added a considerable amount of short loops but the transmissibility of the infection could be insufficient for the effects of clustering to manifest. For example, a household of size 5 surrounded by workplaces of size 20 would be able to reach other households through $5 \times 20 = 100$ different paths. Even if some of them lead to the same household or go back to the initial household itself because of the presence of loops in the network, the influenza-like parameters of the previous sections suggests that R_H is of the order of a few units and therefore that only a few of these paths are actually involved in the transmission process. In other words, even if loops exist, the spread could be simply too weak to experience saturation effects. This is probably the case for influenza, given the limited infectivity that characterises it.

It is also worth noting that real epidemic often do not show a clear exponentially growing phase, but the motivation is not particularly well understood and has not been properly investigated to date. The exploration performed here suggests that pure spatial factors are unlikely to be the main cause of the often observed lack of exponentially growing phase in real epidemics. Note however that other different forms of social clustering occur in reality (e.g. work colleagues deciding to register their children to the same school; cultural, religious or political affinities and so on), which are not taken into account by adding purely spatial constraints only. Future research is needed to find data and methods to rigorously quantify such factors and include them into models in order to test if they can provide sufficient explanation for the observed epidemic patterns that do not show an exponentially growing phase.

Given that no detectable differences appear between epidemics spreading on the random and the spatially constrained networks, the random network is used to explore the effects of closing schools. The numerical values for the parameters and the same assumptions on the structure of the population used for the estimation of R_H are used. Again, quantitative results have to be taken with caution, given all the strong assumptions made. However, they highlight that the effectiveness of closing schools is extremely difficult to estimate. The main reasons are:

1. lack of data about workplaces and workplace transmission;
2. little data availability concerning school transmission at present;
3. difficulty in assessing the type and amount of compensatory behaviour after school closure;
4. poor quantitative understanding of the implications of school closure on workplace transmission.

The last statement is equivalent to a poor quantitative understanding of how much transmission occurs in the workplace with respect to the community: for the same R_0 , since the contribution of household transmission is the same in both cases, high values of R_g suggest that little transmission occurs at work/school and it is therefore associated with a low effectiveness of school closure; vice versa, low values of R_g suggest that the control policy can be very effective.

In addition, it is still necessary to keep in mind that the model neglects the fact that larger households are preferentially linked with school classes, which could play an important role in estimating school closure efficacy.

Note however that the case when workplaces have size 1, i.e. only schools are closed, is substantially in agreement with the results of Cauchemez *et al.* (2008): despite not having taken into account any form of compensatory behaviour or delay in school closure, the estimated reduction in the final epidemic size is somewhat limited (12% in Figure 7.9),

although it is accompanied by a delay in the peak and a reduction in the daily incidence at the peak, which could alleviate the stress on the health care system and buy it more time to react.

Finally, note that in the case of workplaces of size 4, community transmission contributes to around $1/3$ of the whole transmission of an individual; since household transmission also contributes for roughly $1/3$ (see Fraser, 2007), this choice of parameter values supports the *a priori* assumptions appearing in earlier studies (Ferguson *et al.*, 2006; Wu *et al.*, 2006) that transmission is equally partitioned between the household, the workplace and the community. When workplaces have size 1, the proportion of community transmission can increase up to $1/2$ or even more, thus reducing the contribution of workplace transmission to the total. It needs to be stressed however that such an observation is highly dependent on the definition of workplace adopted here. If workplaces are defined to be much larger (entire schools, companies, universities), then the contribution of workplace transmission to the total could be enhanced. For this reason, the answer to the question about how much transmission occurs via different routes is much more than a simple matter of parameter values which have not been estimated up to now: it is difficult to obtain because it is connected to a more fundamental problem concerning the definition of a workplace.

Chapter 8

Conclusions and future research

8.1 A look backwards

This thesis has merely added a drop into the ocean of the contributions that mathematical models have brought and continue to bring to the field of infectious disease epidemiology. After almost a century of development, such contributions are now widely recognised by the epidemiological community.

Throughout this period, the hypothesis of homogeneous mixing required for mathematical convenience, although unrealistic, has been at the base of entire generations of models and has provided remarkable insights into the study of infectious disease spread and surprisingly accurate predictions.

However, mathematical techniques and the computational power available today offer much scope for an improvement in this direction, with the purpose of making qualitative and quantitative results more robust and generalisable. For example, understanding which types of heterogeneity in a population are responsible for particular epidemic trends allows educated predictions of the spread of similar pathogens into other populations with different characteristics: it is the case, for example, of a possible future influenza pandemic, the dynamics of which are difficult to predict not only because the characteristics of the virus are unknown, but also because past influenza pandemics occurred in populations with very

different demographic characteristics.

For this reason, this thesis offered a broad overview of the methods and techniques used in the attempt of relaxing the hypothesis of homogeneous mixing, in addition to a detailed investigation of a particular model in which the population is structured into households and workplaces. Such a model, although still highly idealised, departs from the homogeneous mixing assumption as it includes some of the social characteristics that are deemed to be most relevant for the spread of many directly transmissible infections, like influenza, SARS, measles, chickenpox, and so on. Being at the limit of what can be studied analytically, it benefitted from the joint contribution of both the analytical approach and the stochastic simulation.

Original contributions have involved:

1. the definition of a new household reproduction number R_H for the model with households and workplaces, which overcomes some of the limitations of a previously defined reproduction number (R_* , see Section 4.3.2);
2. the study of the properties of R_H , in particular its relationship with the effort needed to control an epidemic when control policies are targeted at households;
3. the extension of a methodology for the exact or approximate computation of the real-time growth rate for the households model to the model with households and workplaces;
4. the investigation of the impact that different assumptions on the disease history have on the model outcomes;
5. the investigation of the risk of misestimating R_0 and the average final size from the real-time growth rate when the social structure is ignored: for pathogens that are not particularly infectious, as it is the case for influenza, the presence of the social structure seems to have a modest effect in shaping the spread;

6. the estimation of a range of realistic values for R_H in the case of influenza, which highlights the risk of underestimating the effort needed to stop the epidemic spread when interventions are targeted at households: within a safe range of 2.26 to 4, the most probable values for R_H are between 2.5 and 2.86;
7. the observation that pure spatial constraints applied to the population are generally insufficient to have a visible impact on the overall dynamics of the spread of a pathogen which is not particularly infectious, as it is the case of influenza: however, they might considerably affect the spatial spread and therefore open further interesting questions on spatially targeted interventions;
8. the analysis of the possible range of effects of school closure depending on the particular model assumptions: although predictions are very difficult, an interesting insight gained by this analysis is that the effectiveness of such a control policy is highly dependent on the amount of reduction in between-adults transmission that is obtained as a consequence of parents needing to stay at home to take care of children not attending school.

8.2 A look forward

Most of these results are theoretical in nature or offer more a qualitative insight than a robust quantitative one. The main reasons are the lack of reliable data and the limited amount of heterogeneity that the idealised modelling approaches discussed in this thesis are able to capture. This highlights the need for further improvements in all aspects involved in this research field, from larger-scale and more reliable data collection to development of robust statistical method for parameter estimation, from theoretical improvement in epidemic models to extensions and enhanced robustness of large-scale simulations.

More specifically, much more research is needed in the development of idealised but mathematically tractable models, as they allow a neat exploration of the causal determinants of particular characteristics observed in epidemic spread.

In particular, it would be interesting to explore the limits of standard analytically tractable models, for which the concept of R_0 is almost always deemed to be a fundamental concept. In fact, the presence of a reproduction number is strictly linked with the presence of an exponentially growing phase, while observed epidemics often do not show such a pattern. More research is needed to gain understanding of the causal determinants behind the lack of an exponentially growing phase: such research would be invaluable in reducing the distance between the more theoretical epidemic models and the models that aim at providing more direct answers to contingent epidemiological questions.

Furthermore, not much effort has been spent to date in the comparison between various modelling approaches. Fundamentally different approaches (e.g. households, network, multitype, age-structured, spatial, pair, metapopulation models, etc.) has been taken in different studies, usually offering invaluable insights into important questions. However, investigation on the differences and of possible forms of mapping between them mostly remains an open research question. The most common attempts in this direction are surely those concerning the difference between network models and homogeneous mixing models, as they are somewhat easier to map into one another. Important contributions in the exploration of the patterns of epidemic spread on different network models, and therefore of the differences they show with respect to the homogeneous mixing models, include for example Keeling & Rohani (2008); Keeling (1999); Keeling & Eames (2005), and Bansal *et al.* (2007). Other examples of important attempts in the direction of model selection are represented by the comparison between households and homogeneous mixing models in Becker & Utev (1998) (and to a certain extent also in Fraser, 2007) and the approximation of spatial models with network models and pair approximations in Parham & Ferguson (2006) and Parham *et al.* (2008).

Despite the important role of idealised but analytically tractable models, however, such models usually offer only partial insight in the realistic effect of various forms of control policies. In fact targeted interventions, which are more efficient, less expensive, and faster to implement than interventions at the population level, require the model

to include enough details to allow their investigation, resulting in a level of complexity that can often be dealt with only through simulations: it is usually the case of spatially targeted interventions, but often also of reactive interventions to detected or suspected cases, of various forms of contact tracing, etc. In all these cases, realistic delays in policy implementation have a major impact on their efficacy, which is therefore difficult to study analytically. Therefore further research is much needed also in the direction of individual-based realistic simulation modelling approaches.

Finally, given their computational intensity, such simulation models are often not as fast as it might be needed in the case of real-time reaction to an ongoing epidemic, although today's speed in computational power increase is extremely promising. Therefore analytically tractable approximate tools are still invaluable when a quick analysis and answer is needed. Much more research should therefore be spent in simple and approximate statistical inference techniques for parameter estimation in models which tend to capture only those forms of heterogeneities that are deemed to be driving an ongoing epidemic but for which sufficient data can be realistically collected in conditions of public health stress.

In particular it is worth highlighting the present lack of statistical tools for parameter estimation able to include even basic forms of social heterogeneities. Unfortunately, this limits the use of more sophisticated models than those assuming homogeneous mixing during an emerging epidemic, thus relegating such models into the area of the tools that are theoretically interesting to explore but that bear little practical usefulness in the decision making process.

Furthermore, the development of such statistical tools would highlight which data need to be collected during an epidemic in addition to that already included in standard protocols, an onerous task which is unlikely to be implemented without a scientifically grounded argument in support of its advantages and cost-effect benefits in terms of public health.

Bibliography

- ALBERT, R., & BARABÁSI, A.-L. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics*, **74**(1), 47 – 97.
- ALDIS, G. K., & ROBERTS, M. G. 2005. An integral equation model for the control of a smallpox outbreak. *Mathematical Biosciences*, **195**(1), 1.
- ANDERSON, R. M., & MAY, R. M. 1991. *Infectious diseases of humans: dynamics and control*. Oxford University Press.
- ANDERSSON, H., & BRITTON, T. 2000. *Stochastic epidemic models and their statistical analysis*. Lecture notes in statistics ; 151. New York ; London: Springer.
- ANDERSSON, M. 1999. The asymptotic final size distribution of multitype chain-binomial epidemic processes. *Advances in Applied Probability*, **31**(1), 220.
- BAILEY, N. T. J. 1975. *The mathematical theory of infectious diseases and its applications*. 2nd ed. edn. London: Griffin.
- BALL, F. 1983a. A Threshold Theorem for the Reed-Frost Chain-Binomial Epidemic. *Journal of Applied Probability*, **20**(1), 153–157.
- BALL, F. 1986. A Unified Approach to the Distribution of Total Size and Total Area under the Trajectory of Infectives in Epidemic Models. *Advances in Applied Probability*, **18**(2), 289–310.

- BALL, F., & NEAL, P. 2002. A general model for stochastic *SIR* epidemics with two levels of mixing. *Mathematical Biosciences*, **180**, 73–102.
- BALL, F., MOLLISON, D., & SCALIA-TOMBA, G. 1997. Epidemics with two levels of mixing. *The Annals of Applied Probability*, **7**(1), 46–89.
- BALL, F. G. 1983b. The Threshold Behaviour of Epidemic Models. *Journal of Applied Probability*, **20**(2), 227.
- BALL, F. G., & BRITTON, T. 2007. An epidemic model with infector-dependent severity. *Advances in Applied Probability*, **39**(4), 949.
- BALL, F. G., & LYNE, O. D. 2002. Optimal vaccination policies for stochastic epidemics among a population of households. *Mathematical Biosciences*, **177**, 333–354.
- BALL, F. G., & NEAL, P. 2008. Network epidemic models with two levels of mixing. *Mathematical Biosciences*, **212**(1), 69–87.
- BALL, F. G., BRITTON, T., & LYNE, O. D. 2004. Stochastic multitype epidemics in a community of households: Estimation of threshold parameter R_* and secure vaccination coverage. *Biometrika*, **91**(2), 345–362.
- BANSAL, S., GRENFELL, B. T., & MEYERS, L. A. 2007. When individual behaviour matters: homogeneous and network models in epidemiology. *Journal of the Royal Society Interface*, **4**(16), 879.
- BARBOUR, A. D. 1975. The Duration of the Closed Stochastic Epidemic. *Biometrika*, **62**(2), 477–482.
- BARTOSZYŃSKI, R. 1972. On a certain model of an epidemic. *Applicationes Mathematicae*, **13**(2), 139–151.
- BECKER, N. G., & DIETZ, K. 1995. The effect of household distribution on transmission and control of highly infectious diseases. *Mathematical Biosciences*, **127**, 207–219.

- BECKER, N. G., & DIETZ, K. 1996. Reproduction numbers and critical immunity levels for epidemics in a community of households. *In: Athens Conference on Applied Probability and Time Series*. Applied Probability, Lecture Notes in Statistics 114, p. 267, vol. Volume 1.
- BECKER, N. G., & STARCZAK, D. N. 1998. Optimal vaccination strategies for a community of households. *Mathematical Biosciences*, **139**(2), 117.
- BECKER, N. G., & UTEV, S. 1998. The effect of community structure on the immunity coverage required to prevent epidemics. *Mathematical Biosciences*, **147**(1), 23–39.
- BECKER, N. G., BAHRAMPOUR, A., & DIETZ, K. 1995. Threshold parameters for epidemics in different community settings. *Mathematical Biosciences*, **129**(2), 189–208.
- BECKER, N. G., GLASS, K., LI, Z., & ALDIS, G. K. 2005. Controlling emerging infectious diseases like SARS. *Mathematical Biosciences*, **193**(2), 205–221.
- BRISSON, M., EDMUNDS, W. J., GAY, N. J., LAW, B., & DE SERRES, G. 2000. Modelling the impact of immunization on the epidemiology of varicella zoster virus. *Epidemiology and Infection*, **125**(03), 651–669.
- CAUCHEMEZ, S., CARRAT, F., VIBOUD, C., VALLERON, A. J., & BOELLE, P. Y. 2004. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Statistics in Medicine*, **23**(22), 3469–3487.
- CAUCHEMEZ, S., VALLERON, A. J., BOELLE, P. Y., FLAHAULT, A., & FERGUSON, N. M. 2008. Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature*, **452**(7188), 750–754.
- CHAN-YEUNG, M., & YU, W. C. 2003. Outbreak of severe acute respiratory syndrome in Hong Kong Special Administrative Region: case report. *BMJ*, **326**(7394), 850–852.

- COLIZZA, V., BARRAT, A., BARTHELEMY, M., VALLERON, A. J., & VESPIGNANI, A. 2007. Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions. *PLoS Medicine*, **4**(1), e13.
- DATTA, S., HALLORAN, M. E., & LONGINI JR, I. M. 1999. Efficiency of Estimating Vaccine Efficacy for Susceptibility and Infectiousness: Randomization by Individual Versus Household. *Biometrics*, **55**(3), 792.
- DE JONG, M. C. M., DIEKMANN, O., & HEESTERBEEK, J. A. P. (eds). 1995. *How does transmission of infection depend on population size?* ed. cambridge edn. Epidemic models: their structure and relation to data. Cambridge University Press.
- DEMIRIS, N., & O'NEILL, P. D. 2005. Bayesian inference for stochastic multitype epidemics in structured populations via random graphs. *Journal of the Royal Statistical Society*, **67**, 731.
- DIEKMANN, O., & HEESTERBEEK, J. A. P. 2000. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Wiley series in mathematical and computational biology. Chichester: Wiley.
- DIEKMANN, O., HEESTERBEEK, J. A., & METZ, J. A. 1990. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, **28**(4), 365–82.
- DODD, P. J., & FERGUSON, N. M. 2007. Approximate disease dynamics in household-structured populations. *Journal of the Royal Society interface*, **4**(17), 1103.
- EDMUNDS, W. J., O'CALLAGHAN, C. J., & NOKES, D. J. 1997. Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proceedings of the Royal Society of London. Series B*, **264**(1384), 949–57.

- ELVEBACK, L. R., ACKERMAN, E., GATEWOOD, L., & FOX, J. P. 1971. Stochastic two-agent epidemic simulation models for a community of families. *American Journal of Epidemiology*, **93**(4), 267–80.
- ELVEBACK, L. R., FOX, J. P., ACKERMAN, E., LANGWORTHY, A., BOYD, M., & GATEWOOD, L. 1976. An influenza simulation model for immunization studies. *American Journal of Epidemiology*, **103**(2), 152–65.
- FERGUSON, N. M., DONNELLY, C. A., & ANDERSON, R. M. 2001. The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science*, **292**(5519), 1155–60.
- FERGUSON, N. M., KEELING, M. J., EDMUNDS, W. J., GANI, R., GRENFELL, B. T., ANDERSON, R. M., & LEACH, S. 2003. Planning for smallpox outbreaks. *Nature*, **425**(6959), 681–685.
- FERGUSON, N. M., CUMMINGS, D. A. T., CAUCHEMEZ, S., FRASER, C., RILEY, S., MEEYAI, A., IAMSIRITHAWORN, S., & BURKE, D. S. 2005. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, **437**, 209–214.
- FERGUSON, N. M., CUMMINGS, D. A. T., FRASER, C., CAJKA, J. C., COOLEY, P. C., & BURKE, D. S. 2006. Strategies for mitigating an influenza pandemic. *Nature*, **442**(7101), 448.
- FINE, P. E. M. 1977. A commentary on the mechanical analogue to the Reed-Frost epidemic model. *American Journal of Epidemiology*, **106**(2), 87–100.
- FRASER, C. 2007. Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic. *PLoS ONE*, **2**(8), e758.
- FRASER, C., RILEY, S., ANDERSON, R. M., & FERGUSON, N. M. 2004. Factors that make an infectious disease outbreak controllable. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(16), 6146.

- GERMANN, T. C., C., TIMOTHY, KADAU, K., LONGINI, I. M., & MACKEN, C. A. 2006. Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(15), 5935–40.
- GILLESPIE, D. T. 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, **22**(4), 403–434.
- GILLESPIE, D. T. 1977. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry. A*, **81**(25), 2340.
- GRABOWSKI, A., & KOSIŃSKI, R. A. 2004. Epidemic spreading in a hierarchical social network. *Physical Review E*, **70**(3), 031908. TY - JOUR M1 - Copyright (C) 2009 The American Physical Society Please report any problems to prola@aps.org.
- GRABOWSKI, A., & KOSIŃSKI, R. A. 2005. The *SIS* Model of Epidemic Spreading in a Hierarchical Social Network. *Acta physica Polonica. B*, **36**(5), 1579.
- GRABOWSKI, A., & KOSIŃSKI, R. A. 2006. Epidemic control in a hierarchical social network. *International Journal of Modern Physics C*, **17**(1), 133–145.
- GRABOWSKI, A., & ROSIŃSKA, M. 2006. The *SIS* model for assessment of epidemic control in a social network. *Acta physica Polonica. B*, **37**(5), 1521.
- GRASSLY, N. C., & FRASER, C. 2008. Mathematical models of infectious disease transmission. *Nature Reviews, Microbiology*, **6**(6), 477–487.
- GRASSLY, N. C., FRASER, C., WENGER, J., DESHPANDE, J. M., SUTTER, R. W., HEYMANN, D. L., & AYLWARD, R. B. 2006. New strategies for the elimination of polio from India. *Science*, **314**(5802), 1150–1153.
- GRIMMETT, G., & STIRZAKER, D. 2001. *Probability and random processes*. 3rd ed. edn. Oxford: Oxford University Press.

- GRIMMETT, G., & WELSH, D. 1986. *Probability: an introduction*. Oxford: Clarendon.
- HALLORAN, M. E. 2003. Estimating Vaccine Efficacy from Secondary Attack Rates. *Journal of the American Statistical Association*, **98**(461), 38.
- HALLORAN, M. E., COCHI, S. L., LIEU, T. A., WHARTON, M., & FEHRS, L. 1994. Theoretical Epidemiologic and Morbidity Effects of Routine Varicella Immunization of Preschool Children in the United States. *Am. J. Epidemiol.*, **140**(2), 81–104.
- HALLORAN, M. E., HAYDEN, F. G., YANG, Y., LONGINI, I. M., & MONTTO, A. S. 2007. Antiviral effects on influenza viral transmission and pathogenicity: Observations from household-based trials. *American Journal of Epidemiology*, **165**(2), 212–221.
- HAMER, W. H. 1906. Epidemic disease in England. *The Lancet*, **i**, 733–739.
- HEESTERBEEK, J. A. P. 2002. A brief history of R_0 and a recipe for its calculation. *Acta Biotheoretica*, **50**(3), 189–204.
- HEESTERBEEK, J.A.P., & ROBERTS, M. G. 2007. The type-reproduction number T in models for infectious disease control. *Mathematical Biosciences Alcalá Special Issue*, **206**(1), 3–10.
- HETHCOTE, H. W. 2000. The mathematics of infectious diseases. *SIAM Review*, **42**(4), 599–653.
- HETHCOTE, H. W., & DRIESSCHE, P. 1991. Some epidemiological models with nonlinear incidence. *Journal of Mathematical Biology*, **29**(3), 271–287.
- HETHCOTE, H. W., & TUDOR, D. W. 1980. Integral equation models for endemic infectious diseases. *Journal of Mathematical Biology*, **9**(1), 37–47.
- HOLLINGSWORTH, T. D., ANDERSON, R.M., & FRASER, C. 2008. HIV-1 Transmission, by Stage of Infection. *The Journal of Infectious Diseases*, **198**(5), 687–693.

- HOPE SIMPSON, R. E. 1952. Infectiousness of communicable diseases in the household (measles, chickenpox, and mumps). *Lancet*, **260**, 549–554.
- HOPE SIMPSON, R. E., & SUTHERLAND, I. 1954. Does influenza spread within the household? *The Lancet*, **266**(6814), 721–6.
- HOUSE, T., & KEELING, M. J. 2008. Household structure and infectious disease transmission. *Epidemiology and Infection*. Published online.
- JANSEN, V. A. A., STOLLENWERK, N., JENSEN, H. J., RAMSAY, M. E., EDMUNDS, W. J., & RHODES, C. J. 2003. Measles Outbreaks in a Population with Declining Vaccine Uptake. *Science*, **301**(5634), 804–.
- KEELING, M. J. 1999. The effects of local spatial structure on epidemiological invasions. *Proceedings - Royal Society. Biological sciences*, **266**(1421), 859.
- KEELING, M. J. 2005. The implications of network structure for epidemic dynamics. *Theoretical Population Biology*, **67**(1), 1–8.
- KEELING, M. J. 2008. On methods for studying stochastic disease dynamics. *Journal of the Royal Society Interface*, **5**(19), 171.
- KEELING, M. J., & EAMES, K. T. D. 2005. Networks and epidemic models. *Journal of the Royal Society Interface*, **2**(4), 295–307.
- KEELING, M. J., & ROHANI, P. 2008. *Modeling infectious diseases in humans and animals*. Princeton, N.J. ; Woodstock: Princeton University Press.
- KENDALL, W. S., & SAUNDERS, I. W. 1983. Epidemics in Competition II: The General Epidemic. *Journal of the Royal Statistical Society. Series B (Methodological)*, **45**(2), 238–244.
- KERMACK, W. O., & MCKENDRICK, A. G. 1927. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A*, **115**, 700–721.

- KURTZ, T. G. 1971. Limit Theorems for Sequences of Jump Markov Processes Approximating Ordinary Differential Processes. *Journal of Applied Probability*, **8**(2), 344–356.
- KURTZ, T. G. 1981. The Central Limit Theorem for Markov Chains. *The Annals of Probability*, **9**(4), 557–560.
- LEFÈVRE, C., & PICARD, P. 1990. A non-standard family of polynomials and the fsd of Reed-Frost mode. *Advances in Applied Probability*, **22**, 25–48.
- LEFÈVRE, C., & PICARD, P. 1999. Abel-Gontcharoff Pseudopolynomials and the Exact Final Outcome of *SIR* Epidemic Models (III). *Advances in Applied Probability*, **31**(2), 532–550.
- LEFEVRE, C., & UTEV, S. 1999. Branching Approximation for the Collective Epidemic Model. *Methodology and Computing in Applied Probability*, **1**(2), 211.
- LONGINI, I. M., & KOOPMAN, J. S. 1982. Household and community transmission parameters from final distributions of infections in households. *Biometrics*, **38**(1), 115–126.
- LONGINI, I. M., KOOPMAN, J. S., MONTÓ, A. S., & FOX, J. P. 1982. Estimating household and community transmission parameters for influenza. *American Journal of Epidemiology*, **115**(5), 736–51.
- LONGINI, I. M., HALLORAN, M. E., NIZAM, A., & YANG, Y. 2004. Containing Pandemic Influenza with Antiviral Agents. *American Journal of Epidemiology*, **159**(7), 623.
- LONGINI, I. M., NIZAM, A., XU, S., UNGCHUSAK, K., HANSHAOWORAKUL, W., CUMMINGS, D. A. T., & HALLORAN, M. E. 2005. Containing Pandemic Influenza at the Source. *Science*, **309**(5737), 1083–1087.
- LUDWIG, D. 1974. *Stochastic population theories*. Lecture notes in biomathematics 3. Berlin: Springer-Verlag.

- LUDWIG, D. 1975. Final size distributions for epidemics. *Mathematical Biosciences*, **23**, 33.
- MARKEL, H., LIPMAN, H. B., NAVARRO, J. A., SLOAN, A., MICHALSEN, J. R., STERN, A. M., & CETRON, M. S. 2007. Nonpharmaceutical Interventions Implemented by US Cities During the 1918-1919 Influenza Pandemic. *JAMA*, **298**(6), 644–654.
- MCCALLUM, H., BARLOW, M., & HONE, J. 2001. How should pathogen transmission be modelled? *Trends in Ecology & Evolution*, **16**(6), 295.
- MCKENDRICK, A.G. 1926. Application of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, **44**, 98.
- MINISTRY OF EDUCATION AND MINISTRY OF HEALTH. 2003. *Joint press release on School Closure*. Tech. rept. <http://www.moe.gov.sg/media/press/2003/pr20030326.htm>.
- MOLLISON, D. 1991. Dependence of epidemic and population velocities on basic parameters. *Mathematical Biosciences*, **107**(2), 255–87.
- MOORE, C., & NEWMAN, M. E. J. 2000. Epidemics and percolation in small-world networks. *Physical Review E*, **61**(5), 5678–5682.
- MURRAY, J. D. 2002. *Mathematical biology*. 3rd ed. edn. Interdisciplinary applied mathematics ; v. 17. New York ; London: Springer.
- MURRAY, J. D., STANLEY, E. A., & BROWN, D. L. 1986. On the Spatial Spread of Rabies among Foxes. *Proceedings of the Royal Society of London. Series B, Biological Sciences (1934-1990)*, **229**(1255), 111–150.
- NEAL, P. 2006. Multitype randomized ReedFrost epidemics and epidemics upon random graphs. *The Annals of Applied Probability*, **16**(3), 1166–1189.
- NEWMAN, M. E., MOORE, C., & WATTS, D. J. 2000. Mean-field solution of the small-world network model. *Phys. Rev. Lett.*, **84**(14), 3201–4.

- NEWMAN, M. E. J. 2002. Spread of epidemic disease on networks. *Physical Review E*, **66**(1).
- NEWMAN, M. E. J. 2003. The Structure and Function of Complex Networks. *SIAM Review*, **45**(2), 167–256.
- PARHAM, P. E., & FERGUSON, N. M. 2006. Space and contact networks: capturing the locality of disease transmission. *Journal of the Royal Society interface*, **3**(9), 483.
- PARHAM, P. E., SINGH, B. K., & FERGUSON, N. M. 2008. Analytic approximation of spatial epidemic models of foot and mouth disease. *Theoretical Population Biology*, **73**(3), 349–368.
- PELLIS, L., FERGUSON, N. M., & FRASER, C. 2008. The relationship between real-time and discrete-generation models of epidemic spread. *Mathematical Biosciences*, **216**(1), 63–70.
- PELLIS, L., FERGUSON, N. M., & FRASER, C. 2009. Threshold parameters for a model of epidemic spread among households and workplaces. *Journal of The Royal Society Interface*. published online.
- PICARD, P., & LEFÈVRE, C. 1990. A Unified Analysis of the Final Size and Severity Distribution in Collective Reed-Frost Epidemic Processes. *Advances in Applied Probability*, **22**(2), 269.
- PRESS, W. H. 2007. *Numerical recipes : the art of scientific computing*. 3rd ed. edn. Cambridge ; New York: Cambridge University Press.
- RILEY, S. 2007. Large-Scale Spatial-Transmission Models of Infectious Disease. *Science*, **316**(5829), 1298 –1301.
- RILEY, S., & FERGUSON, N. M. 2006. Smallpox transmission and control: spatial dynamics in Great Britain. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(33), 12637–42.

- ROBERTS, M. G. 2004. Modelling strategies for minimizing the impact of an imported exotic infection. *Proceedings of the Royal Society B: Biological Sciences*, **271**(1555), 2411–2415.
- ROBERTS, M. G., & HEESTERBEEK, J. A. P. 2003. A new method for estimating the effort required to control an infectious disease. *Proceedings of the Royal Society of London. Series B (Biological Sciences)*, **270**(1522), 1359–1364.
- ROBERTS, M. G., BAKER, M., JENNINGS, L. C., SERTSOU, G., & WILSON, N. 2007. A model for the spread and control of pandemic influenza in an isolated geographical region. *Journal of the Royal Society Interface*, **4**(13), 325–330.
- ROSS, R. 1911. Some quantitative studies in epidemiology. *Nature*, **87**(2188), 466.
- RVACHEV, L. A., & LONGINI JR, I. M. 1985. A mathematical model for the global spread of influenza. *Mathematical biosciences*, **75**(1), 3.
- SCALIA-TOMBA, G. 1985. Asymptotic Final-Size Distribution for Some Chain-Binomial Processes. *Advances in Applied Probability*, **17**(3), 477.
- SCALIA-TOMBA, G. 1986. Asymptotic Final Size Distribution of the Multitype Reed-Frost Process. *Journal of Applied Probability*, **23**(3), 563.
- SCALIA-TOMBA, G. (ed). 1990. *On the asymptotic final size distribution of epidemics on heterogeneous populations*. Springer-Verlag edn. Stochastic processes in epidemic theory, vol. 86.
- SELLKE, T. 1983. On the Asymptotic Distribution of the Size of a Stochastic Epidemic. *Journal of Applied Probability*, **20**(2), 390.
- SVENSSON, Å. 2007. A note on generation times in epidemic models. *Mathematical Biosciences*, **208**(1), 300–311.

- SVENSSON, Å., & SCALIA-TOMBA, G. 2001. *Competing epidemics in closed populations*. Research Report / Mathematical Statistics, Stockholm University, 2001:8. Stockholm: Mathem. Stat., SU.
- UK CENSUS DATA. 2001. <http://www.ons.gov.uk/census/>.
- UK NATIONAL STATISTICS. 2005. <http://www.statistics.gov.uk/STATBASE/ssdataset.asp?vlnk=9390>.
- VON BAHR, B., & MARTIN-LÖF, A. 1980. Threshold limit theorems for some epidemic processes. *Advances in Applied Probability*, **12**(2), 319–349.
- WALD, A. 1947. *Sequential Analysis*. John Wiley and Sons Inc.
- WALLINGA, J., & LIPSITCH, M. 2007. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, **274**(1609), 599–604.
- WATTS, D. J., & STROGATZ, S. H. 1998. Collective dynamics of 'small-world' networks. *Nature*, **393**(6684), 440–2.
- WATTS, D. J., MUHAMAD, R., MEDINA, D. C., & DODDS, P. S. 2005. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(32), 11157–11162.
- WAWER, M.J., GRAY, R.H., SEWANKAMBO, N.K., SERWADDA, D., LI, X., LAEYEN-DECKER, O., KIWANUKA, N., KIGOZI, G., KIDDUGAVU, M., LUTALO, T., NALUGODA, F., WABWIRE-MANGEN, F., MEEHAN, M.P., & QUINN, T.C. 2005. Rates of HIV-1 Transmission per Coital Act, by Stage of HIV-1 Infection, in Rakai, Uganda. *The Journal of Infectious Diseases*, **191**(9), 1403–1409.
- WU, J. T., RILEY, S., FRASER, C., & LEUNG, G. M. 2006. Reducing the impact of the next influenza pandemic using household-based public health interventions. *PLoS Medicine*, **3**(9), 1532–1540.

- YANG, Y., LONGINI JR, I. M., & HALLORAN, M. E. 2006. Design and evaluation of prophylactic interventions using infectious disease incidence data from close contact groups. *Applied Statistics*, **55**(3), 317–330.