Topics in structured host-antagonist interactions

by

Maria Annichia Riolo

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Applied and Interdisciplinary Mathematics)
in The University of Michigan
2014

Doctoral Committee:

Professor Charles Doering, Co-Chair
Professor Pejman Rohani, Co-Chair
Professor Mark Hunter
Professor Aaron King
Professor Mark Newman
Professor Carl Simon

For my parents

ACKNOWLEDGEMENTS

Many thanks to my research collaborators and the members of my dissertation committee for making this dissertation possible.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

LIST OF APPENDICES

Appendix

ABSTRACT

Topics in structured host-antagonist interactions

by

Maria Annichia Riolo

Co-chairs: Charles Doering and Pejman Rohani

In many complex systems, simple parts interact to produce the large scale patterns we observe. The structure of these interactions can have a dramatic effect on the behavior of the system, and many systems which have simple dynamics under the assumption of well-mixed interactions display vastly different behaviors when embedded on a discrete network or a continuous space.

Often, the desynchronization of local dynamics, natural delays of information transfer, and higher dimensionality of the structured system can result in extending the duration of transient dynamics, enabling the stable persistence of heterogeneous solutions, and rendering optimal control challenging. This dissertation explores these phenomena in the context of the dynamics and control infectious diseases and agricultural pests. In particular, I focus on using compartmental models to investigate the effects of age-structured social mixing in the transmission of pertussis and spatially structured mixing and resource heterogeneity in plant-herbivore-parasitoid interactions.

In chapter II, I investigated the potential of age-structure in social contacts to

explain the resurgence of pertussis in highly vaccinated populations. and found that strong age-assortative mixing and a past history of vaccination coverage insufficient for eradication were sufficient to generate a slow resurgence in older age-groups.

In chapter III, I searched for efficient age-targeted booster vaccination strategies using a genetic algorithm, under several simulated modes of vaccine failure. and found that the type of booster schedules most successful in reducing disease strongly depended on the mechanisms of failure.

With an eye to finer scale targeting of vaccination, I derived and presented a multi-way spectral graph partitioning algorithm in chapter IV.

In chapter V, I investigated the effects of spatial variation in plant quality on populations of herbivore hosts and their parasitoids, finding that variation in plant quality occurring at a fine spatial scale decreased overall herbivore populations.

Finally, I explored the relationship between patterns of dispersal and abundance of a population in a one-dimensional space reproduces locally according to the logistic map and disperses with a Gaussian kernel and derived the conditions under which I expect certain classes of behavior to be stable.

CHAPTER I

Introduction

In many complex systems, simple parts interact to produce the large scale patterns we observe. The structure of these interactions can have a dramatic effect on the dynamics of the system. In this dissertation, I focus on the role of structured interactions in host-parasitoid and host-pathogen interactions.

In the first two chapters of this dissertation, we investigated the resurgence of pertussis in the context of age-structured social contacts. Pertussis incidence has been rising in some countries, including the UK, despite sustained high vaccine coverage. We questioned whether it is possible to explain the resurgence without recourse to complex hypotheses about pathogen evolution, subclinical infections, or trends in surveillance efficiency. In particular, we investigated the possibility that the resurgence is a consequence of the legacy of incomplete pediatric immunization, in concert with cohort structure and age-dependent transmission. We constructed a model of pertussis transmission in England and Wales based on data on age-specific contact rates and historical vaccine coverage estimates. We evaluated the agreement between model-predicted and observed patterns of age-specific pertussis incidence under a variety of assumptions regarding the duration of immunity. Under the assumption that infection-derived immunity is complete and lifelong, and regardless of the duration of vaccine-induced immunity, the model consistently predicts a resurgence of pertussis incidence comparable to that which has been observed. Interestingly, no resurgence is predicted when infection- and vaccine-derived immunities wane at the same rate. These results were qualitatively insensitive to rates of primary vaccine failure. We conclude that the alarming resurgence of pertussis among adults and adolescents in Britain and elsewhere may simply be a legacy of historically inadequate coverage employing imperfect vaccines. Indeed, we argue that the absence of resurgence at this late date would be more surprising. Our analysis shows that careful accounting for age dependence in contact rates and susceptibility is prerequisite to the identification of which features of pertussis epidemiology want additional explanation.

Although the mechanisms behind the resurgence of pertussis remain an ongoing point of contention, many countries have nevertheless recommended additional booster vaccinations against pertussis, the timing and number of which vary widely. In the second chapter of this dissertation, we investigated the optimal scheduling of booster vaccinations for the control of a resurgence of a vaccine preventable disease in a population with age-structured social mixing. We explored several scenarios under which infant vaccination alone might fail to eradicate a disease. In each case, we used a genetic algorithm to search for efficient booster vaccination strategies. We found that which booster schedules most successfully controlled transmission strongly depended on why the disease had persisted even with routine infant vaccination, suggesting that booster vaccination schedules are most effective when tailored to the problem at hand and that that misidentifying the causes of pertussis resurgences may be costly.

The problem of optimally targeted vaccination is closely related to that of graph partitioning. One can think of a vaccine that provides sterilizing immunity (preventing both symptomatic disease and transmission) as removing susceptible individuals from the network of risky contacts. While a randomized mass vaccination reduces transmission among unvaccinated individuals by lowering the average number of susceptible contacts they have, targeted vaccination program, such as the age-targeted vaccination strategies in chapter III, aims to exploit the structure of the contact network in deciding which individuals to remove to most effectively interrupt the spread of disease. Reducing the connectivity of a network by targeted removal of nodes is a task very similar in spirit to traditional graph partitioning problems.

However, even finding a bisection of a network by removing the minimum possible number of edges is an NP-hard problem. While there are many well-established graph partitioning algorithms that give good results in practice, multi-way partitioning algorithms are typically presented without derivation and proved to perform well. One downside to this approach is that the derivation of an algorithm can often provide insights into why it works, or even into the underlying behavior of the system at hand. In chapter IV of this dissertation, we derived a multi-way spectral partitioning algorithm for minimum-cut graph partitioning from first principals. We found that the resulting algorithm is a new one, though it shares some similarities with other multi-way spectral partitioning algorithm, and that our algorithm reduces to Fiedler's Laplacian spectral bisection algorithm in the case of division into two equally sized groups. Finally, we demonstrated that our algorithm performs comparably well to the k-means algorithm in some benchmark cases.

In the final two chapters of this dissertation, I consider the role of spatial structure and spatial heterogeneity in shaping population dynamics. Spatial variation in ecological systems can arise both as a consequence of variation in the quality and availability of resources and as an emergent property of spatially structured interactions. In the context of host-parasitoid interactions, spatial structure in interactions has long been of interest as a mechanism for promoting stable coexistence of hosts and parasitoids. Insect parasitoids, whose larvae develop in or on a single host and kill or sterilize it in the process, are a large, diverse group of insects that have been frequently employed in the biological control of agricultural pests and invasive insect herbivores. Though spatially explicit models of host-parasitoid have been studied for decades, this research has focused almost exclusively on homogeneous landscapes, where each location is identical in resources and connectivity. In contrast, there is substantial evidence from field studies of natural and managed systems that spatial variation in resource quality can impact the control of herbivores by their natural enemies.

In order to investigate the effects of spatial heterogeneity in resource quality on the large-scale patterns of herbivore and parasitoid populations, we added local variation in herbivore fecundity to a spatially explicit model of host-parasitoid dynamics. We found that small variation in plant quality at a fine spatial scale decreased overall herbivore populations, as parasitoid populations on low-quality plants were subsidized by high-quality neighbors. On landscapes with large, homogeneous regions of high- and low-quality plants, herbivore populations increased with variation in plant quality. Overall, our results demonstrate that local variation in resource quality profoundly influences global population dynamics. In particular, fine-scale variation in plant quality enhanced biological control of herbivores by parasitoids, suggesting that adding plant genetic variation back into agroecosystems might increase rates of parasitism on pest insects.

We also observed differences in the spatial structure of host and parasitoid population dynamics based on the spatial arrangement of high and low quality plants. In landscapes with large blocks of homogeneous plants, the population dynamics in each homogeneous block reflected the local plant quality. However, the spatial population dynamics on a landscape with fine-scale variation in plant quality more resembled those at the average plant quality, with traveling waves of hosts and parasitoids "overrunning" the spatial variation in plant quality.

To investigate more generally the role of spatial scale in the dynamics of dispersing populations, we consider the toy model of a population that reproduces and diffuses according to a simple integrodifference equation, with reproduction occurring locally according to the logistic map and dispersal using a Gaussian kernel. Using analytical and simulation methods, we discovered conditions under which certain forms of spatially heterogeneous population distributions are stabilized. Even in parameter regimes where a well–mixed population would exhibit chaotic dynamics, we found that different dispersal parameters could lead to a range of spatially and temporally periodic patterns in population density. However, we also find that dispersal limits the extent to which variation in population density can be maintained at high spatial frequencies, and derive a constraint on the power spectrum of the spatially varying population density at any given time.

Chapter II of this dissertation has been published in full (along with Appendix A) as Riolo, M. A., A. A. King, and P. Rohani. 2013. Can vaccine legacy explain the british pertussis resurgence? Vaccine 31:5903–5908. A version of chapter IV has been published as Riolo, M. A., and M. E. J. Newman. 2012. First-principles multiway spectral partitioning of graphs. CoRR abs/1209.5969

CHAPTER II

Can vaccine legacy explain the British pertussis resurgence?

2.1 Introduction

Pertussis resurgence in some highly-developed countries has caused a good deal of alarm. In the United Kingdom, the unexpectedly large outbreak of 2012 — responsible for thirteen infant deaths — has prompted consideration of new prevention measures, including vaccination of pregnant women and a booster dose for adolescents. (United Kingdom. Health Protection Agency., 2012*a*,*b*; Gulland, 2012) Figure 2.1 depicts annual pertussis incidence against the background of vaccine uptake in England and Wales.(United Kingdom. Health Protection Agency., 2008, 2012*b*) Since 2000, a gradual increase in incidence among adults has been apparent. More recently, a sharp rise in incidence among infants and toddlers has become evident. This pattern of increasing incidence, especially among adults and adolescents, has emerged over the past two decades in a number of countries where pertussis had been considered under control. (Celentano et al., 2005; Güriş et al., 1999; Skowronski et al., 2002; Quinn and McIntyre, 2007; Crowcroft and Pebody, 2006; Tan et al., 2005; Wood and MacIntyre, 2008)

A variety of candidate mechanisms have been proposed to explain this phenomenon.

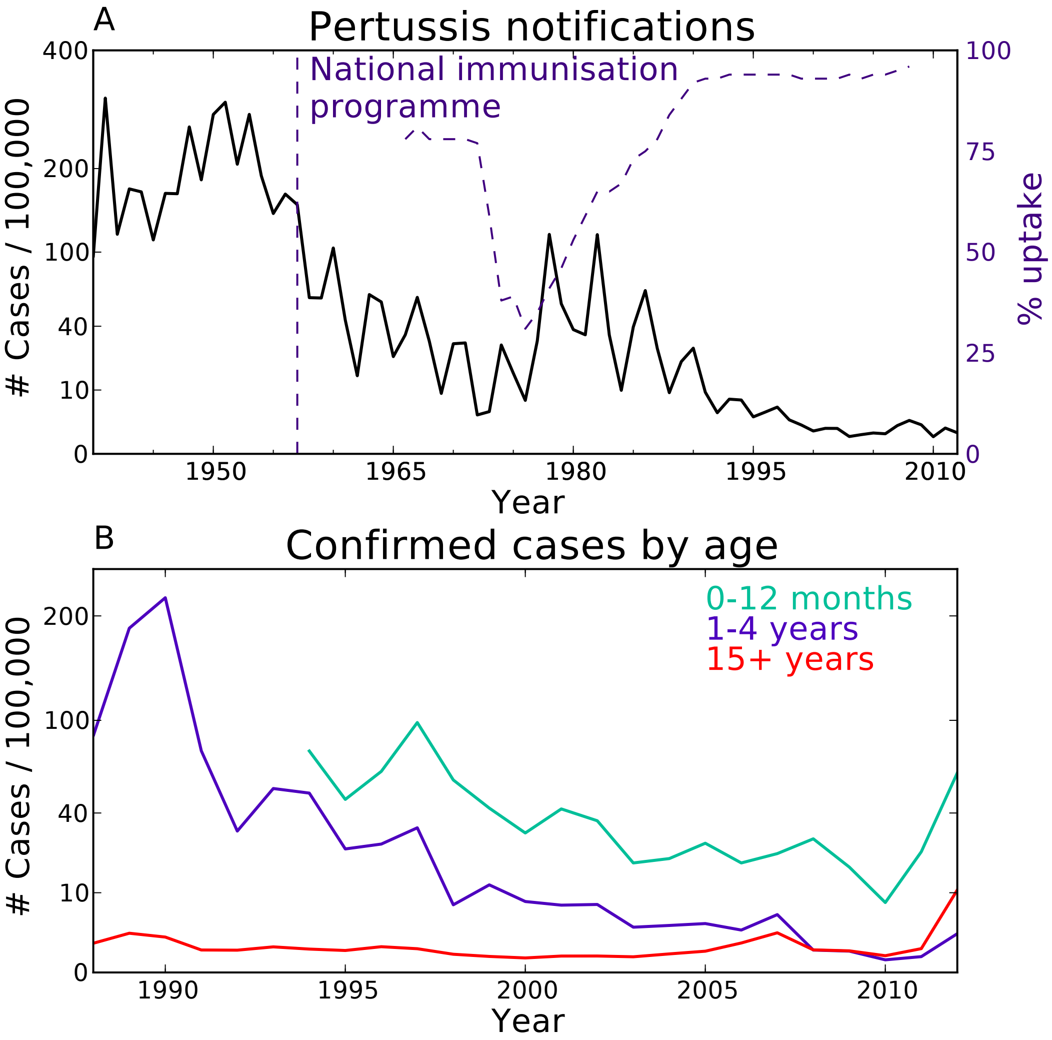


Figure 2.1: Incidence of pertussis in England and Wales over time based on total notifications (panel A, solid line) and lab confirmed cases by age (panel B). (United Kingdom. Health Protection Agency., 2013, 2012*b*) Incidences are plotted on a square root scale for clarity. Estimated vaccine uptake for each birth year is plotted in panel A (dashed line).(United Kingdom. Health Protection Agency., 2008) Although the national immunization program began in 1957, uptake data is unavailable between 1957 and 1966.

Chief among these are the vaccine-driven evolution of *Bordetella pertussis* (the main aetiological agent),(Mooi et al., 2001) improved surveillance, (Cherry, 2003) changes in diagnostic tests,(Cherry, 2012) cessation of natural immune boosting, (Águas, 2006; Lavine et al., 2011) and the switch from whole-cell to acellular vaccines,(Shapiro, 2012) with concomitant changes in the nature and duration of protection.(Cherry, 2012; Shapiro, 2012) Less attention has focused on the long-term consequences of inadequate coverage with an imperfect vaccine. Though effective vaccines have been widely used in the England & Wales since 1957, their efficacy has never been perfect and vaccine coverage has only exceeded 90% since the 1990s. As we show here, the gradual accumulation of a population of individuals who have avoided both infection and vaccination — and thus have escaped receiving protection — sets the stage for a resurgence, even in the absence of the aforementioned complexities. Focusing squarely on the recent pertussis epidemiology in England and Wales, we developed a transmission model to determine the extent to which observed patterns of incidence are a predictable consequence of this legacy of imperfect vaccination.

2.2 Methods

We constructed an age-stratified compartmental model of pertussis transmission dynamics. Individuals are categorised by yearly age groups up to seventy five, with an additional category for infants under five months of age (i.e. too young to have received at least two doses of pertussis vaccine under the pre-1990 vaccine schedule in the UK). For convenience, these age categories are labeled with indices starting from zero, so that N_0 designates the number of 0–5 month olds, N_1 is the number of 6 month–1 year olds, N_2 is the number of 1–2 year olds, and so on up to N_{75}. The total population is designated by N with no subscript.

All ages except for 0–5 mo olds are tracked as yearly cohorts, the cutoff occurring at the start of each school year. Newborns are assumed to age continuously at rate

$a = \frac{12}{5}$ yr^{-1}, corresponding to the assumption that a newborn spends on average 5 mo in the 0–5 mo age category. Susceptible newborns aging at time t have probability $u(t)\, e$ of being protected by vaccination, where $u(t)$ is the vaccine uptake at time t and e is the vaccine efficacy.

The stochastic model is initialized with conditions from the pre-vaccine era and proceeds by updating the numbers of individuals in each age category who are susceptible, latently infected, infectious, recovered, or vaccinated, respectively. Those in the recovered and vaccinated classes are protected from infection for a period, the duration of which is a random variable, as we detail below. The dynamics of susceptible (S_i), exposed (E_i), infectious (I_i), recovered (R_i and R_i'), and vaccinated (V_i and V_i') individuals in age group i are given by:

$$\begin{aligned}
\frac{dS_i}{dt} &= w_V V_i' + w_R R_i' - \lambda_i(t) S_i + \\
&\quad (bN - aS_0)\delta_{i,0} + a(1 - eu(t))S_0\delta_{i,1} && (2.1)\\
\frac{dE_i}{dt} &= \lambda_i(t) S_i - \gamma E_i + aE_0(\delta_{i,1} - \delta_{i,0}) && (2.2)\\
\frac{dI_i}{dt} &= \gamma E_i - r I_i + aI_0(\delta_{i,1} - \delta_{i,0}) && (2.3)\\
\frac{dR_i}{dt} &= r I_i - w_R R_i + aR_0(\delta_{i,1} - \delta_{i,0}) && (2.4)\\
\frac{dR_i'}{dt} &= w_R R_i - w_R R_i' + aR_0'(\delta_{i,1} - \delta_{i,0}) && (2.5)\\
\frac{dV_i}{dt} &= eu(t) a S_0 \delta_{i,1} - w_V V_i + aV_0(\delta_{i,1} - \delta_{i,0}) && (2.6)\\
\frac{dV_i'}{dt} &= w_V V_i - w_V V_i' + aV_0'(\delta_{i,1} - \delta_{i,0}) && (2.7)
\end{aligned}$$

where $\delta_{i,j}$ is the Kronecker delta, which is one if i and j are equal and zero otherwise.

Age group i gains susceptible members through immune waning and, in the first two age categories, births and the ageing of susceptible newborns, respectively. The birth rate $b = \frac{1}{75}$ yr^{-1} is chosen to keep the population steady given the 75 year lifespan. Individuals leave the susceptible category by becoming exposed or, in the infant category, aging.

The force of infection acting on age group i at time t is

$$\lambda_i(t) = q\Sigma_k F_{hk}(t) c_{hk} \frac{\tilde{I}_k}{\tilde{N}_k}$$

where c_{hk} is the average rate (in contacts per year) at which an individual who is $5h$–$(5h+5)$ years old makes contact with $5k$–$(5k+5)$ year olds and q is the probability of infection given exposure. The number of infected individuals and total individuals in the k^{th} five-year age block are denoted by

$$\tilde{I}_k = \Sigma_{5k<i\leq 5k+5}\ I_i \quad \text{and} \quad \tilde{N}_k = \Sigma_{5k<i\leq 5k+5}\ N_i$$

with I_0 and N_0 included in the calculation of $\tilde{I}_0$ and $\tilde{N}_0$, respectively.

Values of c_{ij} and q were adopted from an earlier study. (Rohani et al., 2010) In particular, rates of daily contacts c_{ij} were obtained from the POLYMOD study,(Mossong et al., 2008) (see Figures A.1D and 2.2A) and q was fixed at 4% as in Ref. Rohani et al., 2010, leading to a prevaccine era mean age of first infection consistent with historical estimates (Fig. 2.2B). The necessary steps for obtaining contact rates c_{ij} from the data are described in detail in Section A.1 of the supplementary material.

To capture the strong seasonality in children's social contacts,(Eames et al., 2012) we incorperated an age-dependent seasonal forcing term $F_{hk}(t)$ based on school holidays. For $0 < h < 3$ or $0 < k < 3$ (*i.e.* when either party is 5–15 years old), $F(t) = \kappa(1 \pm 0.2)$, with $+$ when school is in session and $-$ during school holidays. Because there are more school days than holidays, we use the normalization constant κ to ensure that $F(t)$ has a mean of 1.0 over the whole year. The school holidays used in our simulations were July 19 – September 8, October 28 – November 3, December 21 – January 10, and April 10 – 25. If neither party is 5–15 years old, $F_{hk}(t) = 1$, leaving the contact rate unaltered year round.

Beginning in 1957, we assume that infants are vaccinated at six months of age. From 1966, we used available estimates of vaccine uptake for the UK (Figure 1A).(United Kingdom. Health Protection Agency., 2008; Rohani et al., 2000) Uptake data for the period prior to 1966 are unavailable; we assumed uptake ramped up linearly from 1957 to 1966. The results presented in the main text assume a value of 60% coverage in 1957. We explored other values; our results are qualitatively unaffected (see appendix S2). In the absence of a serological marker for protection, the efficacy of pertussis vaccines and the durations of infection- and vaccine-derived immunity are highly uncertain. (Mills et al., 1998; Wendelboe et al., 2005) The results presented in the main text assume a vaccine efficacy of 85%, biologically plausible for both the acellular and whole cell vaccines, (Greco et al., 1996) though other efficacies produced qualitatively similar results (see appendix S3). In further tests of the robustness of our results, we similarly varied our assumptions regarding human lifespan and the shape of the contact matrix (see appendix S1 and S6).

In our simulations, individuals exposed to pertussis become infectious after an average of 8 days ($\gamma = \frac{365}{8}$ yr^{-1}) and the infectious period lasts 15 days on average ($r = \frac{365}{15}$ yr^{-1}), again as in the model of (Rohani et al., 2010).

We model two stages of resistence, R_i and R_i', so that the duration of immunity is gamma distributed with shape parameter two. The waning rate is given by $w_R = \frac{2.0}{d_R}$ yr^{-1} where d_R denotes the mean duration of infection derived immunity. Like infection derived immunity, the duration of vaccine derived immunity is gamma distributed with shape parameter two. In simulations with lifelong natural (or vaccine–derived) immunity, w_R (or w_V) is set to zero. We varied the assumed durations of vaccine- and infectionderived immunity, respectively, between lifelong immunity and durations gamma-distributed with means of 70, 40, and 10 years (see Table 2.1).

Mean Duration of Immunity	Probability of remaining immune after 10 years	Probability of remaining immune after 25 years	Probability of remaining immune after 50 years
70 years	0.97	0.84	0.58
40 years	0.91	0.64	0.29
10 years	0.41	0.04	0.0005

Table 2.1: Biological interpretation of average immune durations.

For the initial conditions of our simulations, we used the population at the end of the 150th year of a run with lifelong immunity, no vaccination, and a total population of approximately sixty-three million individuals. All simulations were run for 250 years, with vaccination beginning in the 157th year. At the end of each year, for each age category we recorded the population, number of susceptibles, number of successful vaccinations, and number of cases. The pre-vaccine behavior among runs with the same duration of natural immunity was very consistent (see Figure 2.3 and Figure S7).

Because of the computational cost of using Gillespie's direct algorithm with so many age-categories, we use a multinomial τ–leaping method,(Pettrgrew and Resat, 2007) in which we move forward by a fixed time step τ and determine the set of events that occurred during that time step. All simulations presented in this paper use $\tau = \frac{1}{365}$ yr.

At each step, we consider all the ways an individual can leave each category as a set of competing events. For a sufficiently small time step τ, we can approximate an individual's probability of leaving a category as the total rate at which individuals leave multiplied by the length of the time step. For example, susceptible newborns leave the category by aging at rate a or by becoming exposed at rate $\lambda_i(t)$, so each of the S_0 susceptible newborns has probability $(a + \lambda_0(t))\tau$ of leaving the category. We determine the total number who leave the category drawing from a binomial distribution, $X \sim B(S_0, p_0)$ in our example. The expected fraction of these individuals leaving via each event is proportional to that event's rate. Continuing our example, the X individuals leaving the susceptible newborn category are aging with probability $\frac{a}{a+\lambda_0(t)}$ and have been exposed with probability $\frac{\lambda_0(t)}{a+\lambda_0(t)}$, so we draw the numbers of aging and exposure events (X_a, X_E) from a multinomial distribution with X trials and probabilities $(\frac{a}{a+\lambda_0(t)}, \frac{\lambda_0(t)}{a+\lambda_0(t)})$. For aging infants, we perform one more binomial draw with probability $eu(t)$ to determine how many aging infants are successfully vaccinated.

The number of births, which don't deplete any population categories, is determined by drawing from binomial distribution $B(N, b\tau)$, where b is the per capita annual birth rate. Because this is a discrete stochastic model, we also include an immigration rate of one infected individual per year, uniformly distributed among age categories to help distinguish between stable eradication and a chance extinction in an easily re–invaded population. Once the set of events taking place has been determined, the whole population is updated according to those events and the time t is incremented by τ.

2.3 Results

In Figure 2.2, we show a typical realization of the model, under the conservative assumption that infection-derived immunity is lifelong and that vaccination protects, on average, for 70 years (note that we did not take into account under-reporting, thus all cases are included in model incidences). Vaccine efficacy is assumed to be 85%, consistent with estimates from Ref. Blennow et al., 1988. Aggregate annual incidence, the estimated annual vaccine uptake and the percentage of the population susceptible to pertussis are shown in Figure 2.2B.

Consistent with observed patterns in England & Wales, incidence in our model declines with the onset of paediatric vaccination and rebounds during the early 1980s, after several years of low uptake due to the mid-'70s vaccine scare, before eventually returning to lower overall incidence and a long inter-epidemic period. Although high

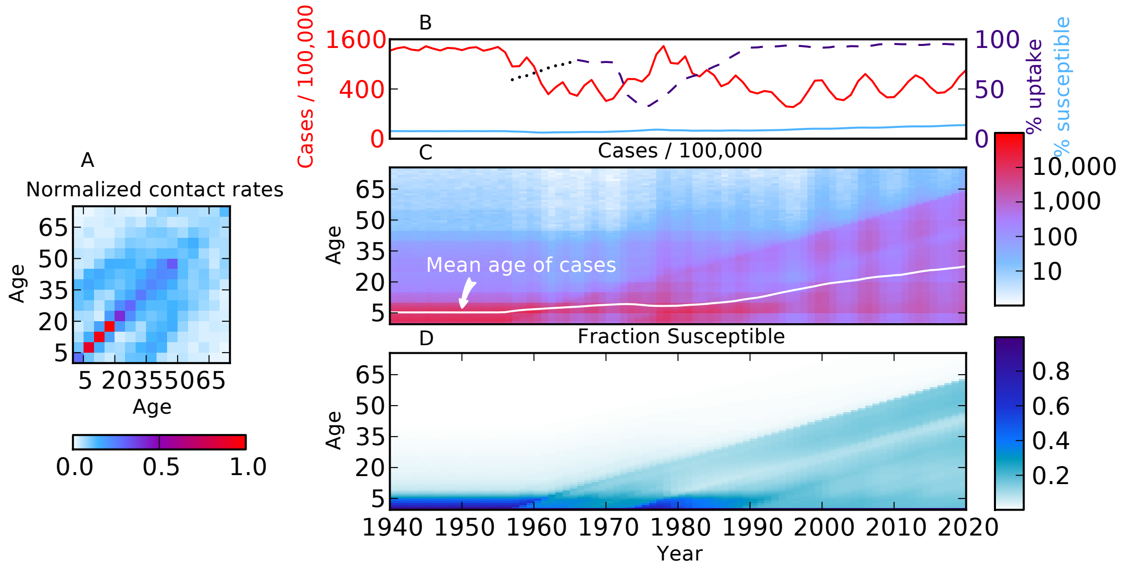


Figure 2.2: A. Age-specific pattern of contacts per day, normalized to run between 0 (white) and 1 (red). Infants, toddlers, and adults are involved in fewer contacts per day than school aged children. Furthermore, most contacts are between people of similar age (the strong, central diagonal). Of the remaining contacts, most appear to be inter-generational (the two weaker diagonals), largely composed of household contacts between parents and children. Panels B, C & D depict the results of a single realization of our transmission model with lifelong vaccine-derived immunity and 85% vaccine efficacy. B. Overall incidence (red), vaccine uptake (purple), and the overall fraction of the population susceptible to pertussis (cyan) by year. C. Incidence (in cases/100,000) by age during each year of a realization of the model. D. The fraction of each age group that is susceptible to pertussis, plotted over time.

vaccine coverage is maintained, our model predicts a gradual increase in overall incidence beginning in the 1990s (Fig. 2.2B). In Figure 2.2C, we dissect these incidence data by age group, demonstrating that, as expected, in the pre-vaccine era, pertussis was most common in young children (with a mean age at infection in our model of 5 years). This figure also shows that the onset of immunization was accompanied by a rise in the age of cases and that, crucially, the increase in overall incidence over the past two decades appears to be primarily among adults. From the rollout of the national infant vaccination programme in 1957 to the present day, the mean age of infection in our model climbed steadily, excluding the slight dip following the vaccine scare of the mid-1970s.

This shift in age distribution was also apparent in the immunological profile of our simulated population (Fig 2.2D). In the pre-vaccine era, the proportion of susceptibles among the population fell sharply with age, as expected for a childhood infection. With the onset of paediatric vaccination, the fraction of susceptible children decreased substantially, but those who escaped protection (either due to incomplete coverage or primary vaccine failure) remained susceptible into adulthood. This effect is clearly visible in Figure 2.2D as a spillover of susceptibles into older age groups over time, with cohorts born at the start of the vaccine era on the leading edge of the wave.

Crucially, this rising incidence of pertussis in adults and adolescents occurred regardless of the assumed duration of vaccine-derived immunity. We provide support for this statement in Figure 2.3, which depicts the annual pertussis incidence in different age groups for varying durations of vaccine- and infection-derived immunity. When infection-derived immunity was lifelong, an increase in incidence among adolescents and adults was inevitable if vaccine-induced protection was not permanent (Fig. 2.3D). With long-lasting vaccine-derived immunity, there was also a notable increase in infant (<1 year old) cases (Fig. 2.3B).

2.4 Discussion

Our results suggest that rising pertussis incidence among adults and adolescents should not be surprising. Indeed, our simulations, even with the conservative assumptions of lifelong natural immunity, a 70-year mean duration of vaccine immunity and 85% efficacy predicted a long-lasting honeymoon period (McLean and Anderson, 1988) followed by a resurgence among older age groups. This pattern is a legacy of incomplete vaccination with an imperfect vaccine: individuals born in the vaccine era are less likely to be infected as children and more likely to remain susceptible as teens and adults than their pre-vaccine predecessors. Thus, during the first few decades of vaccination, the population benefits both directly from vaccine protection

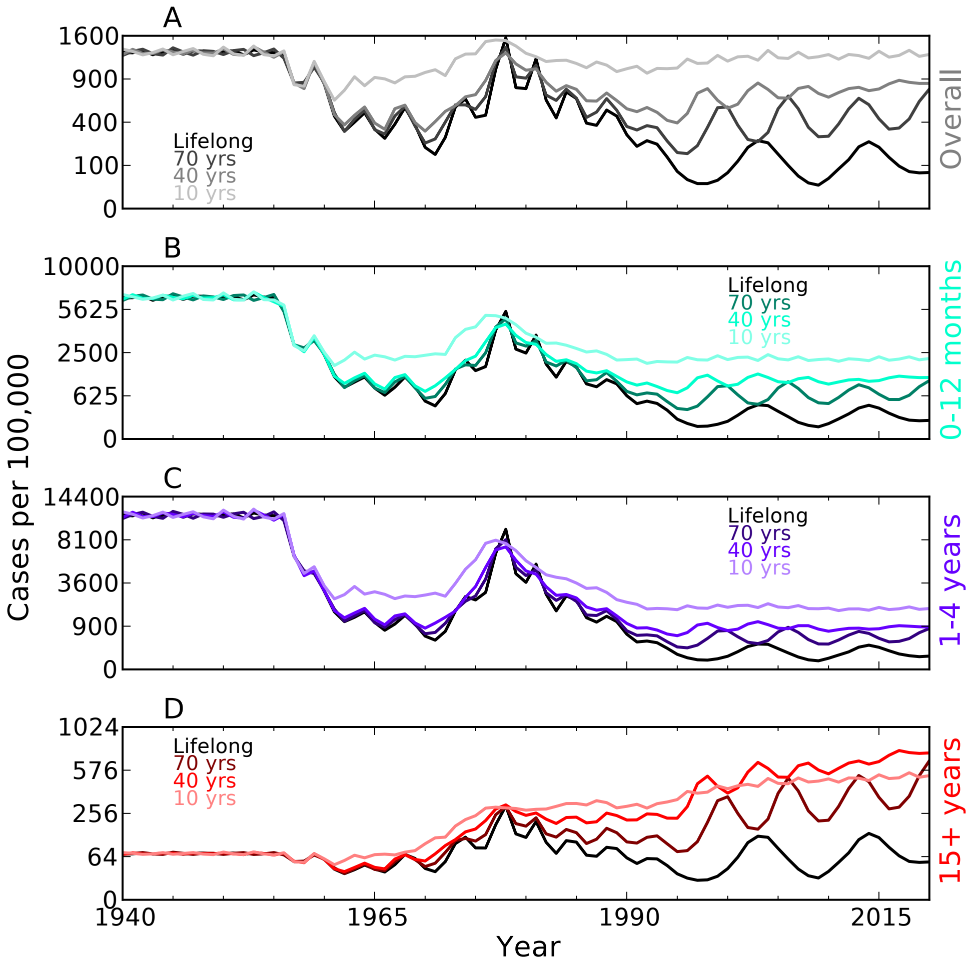


Figure 2.3: Pertussis incidence through time for (A) the whole population, (B) infants under 1 year old, (C) toddlers 1-4 years old, and (D) adults and adolescents over 15 years old, plotted for model realizations assuming lifelong natural immunity. In each panel, different coloured lines indicate predicted incidence when the duration of vaccine-derived immunity is varied from lifelong (black) to a mean of 10 years (lightest). For clarity, only one realization is plotted for each set of parameters, but other realizations matched closely (see Fig. A.8). In the online supplementary materials, we present analogous figures under alternative assumptions of the durations of vaccine-derived and natural immunity.

of children and indirectly from herd immunity established by natural infection in the pre-vaccine era. As cohorts of children born in the vaccine era grow up, the latter effect diminishes and incidence among adults inevitably rises.

In view of the considerable uncertainty surrounding the nature and duration of acquired immunity to pertussis,(Mills, 2001) it is important to assess the robustness of this conclusion. Accordingly, we varied model assumptions regarding the durations of natural and vaccine-induced immunity. When these were assumed equal, there was no resurgence. Rather, the model predicted substantially higher and steady incidence among all age groups (Fig A.8). This prediction contradicts observation.

Our use of a model in which infection-derived immunity can be lifelong may at first glance appear to be at odds with studies documenting reinfection.(Wendelboe et al., 2005) However, a number of population-level studies have found that incidence data are best explained by transmission models with long-lasting natural immunity. (Wearing and Rohani, 2009; Broutin et al., 2010; Rohani et al., 2010) Moreover, evidence suggesting that pertussis reinfections are frequently less transmissible than are primary infections, (Mertsola et al., 1983; Wendelboe et al., 2007) goes some way toward resolving the apparent discrepancy between immune durations estimated from population-level data on the one hand and clinical studies on the other.

The demographics of our model are a rough caricature, with a constant birth rate and type I demographics (life expectancy of seventy five years). The observed slow rise in adult incidence driven by the aging of vaccine-era cohorts, however, is reasonably robust: the same trends emerge under a variety of demographic assumptions (Fig. A.10), though the cohort effects exhibited in Figure 2.3 are less sharply defined. In addition, incorporating immigration and population expansion would enhance the legacy effect described here. Similarly, the legacy of incomplete vaccination is still observed when alternative contact matrix structures are assumed (Fig. A.2). However the detailed epidemiological picture, including the shape of the population's immunity profile and the timing and speed of the predicted resurgence, is affected by the precise structure of the contact network.

These caveats notwithstanding, the principal conclusion of our accounting should be clear. The legacies of infection and vaccination are visible in a population's immunity profile for decades and leave long-lived signatures on incidence dynamics. We submit that the recent resurgence in adult and adolescent pertussis cases may be best understood as the end of a long honeymoon period, during which infection-derived herd immunity and imperfect vaccine protection combined to greatly reduce incidence. In the years to come, as the children who benefitted from infection-induced immunity die, more effective vaccines and vaccination campaigns will likely be required if we are to regain the upper hand on this disease.

CHAPTER III

Troubleshooting pertussis vaccination: one booster schedule doesn't fit all

Pertussis has become a major public health concern in many countries where it was once a plausible candidate for vaccine-based elimination. Although the mechanisms behind these resurgences remain elusive, many countries have nevertheless recommended additional booster vaccinations against pertussis, the timing and number of which vary widely. We searched for cost-effective booster vaccination strategies using a genetic algorithm and found that which booster schedules most successfully controlled pertussis transmission strongly depended on why pertussis transmission had persisted even with routine infant vaccination. Our results suggest that booster vaccination schedules are most effective when tailored to the problem at hand and that that misidentifying the causes of pertussis resurgences may be costly.

Reconciling the recent resurgences of pertussis in many highly-vaccinated countries with the initial successes of routine pertussis vaccination programs and the global heterogeneity in recent trends of pertussis incidence has proved difficult (Jackson and Rohani, 2013), and a wide variety of competing explanations have been proposed. There are many mechanisms which could contribute to increases in pertussis incidence, including improvement in the surveillance and diagnosis of pertussis (Cherry, 2003, 2012) or reduction of the protection afforded by vaccination, whether due to the evolution of *Bordetella pertussis* (Mooi et al., 2001) or the switch from whole cell to acellular vaccines (Shapiro, 2012). Even without any changes in the nature of pertussis transmission, control, or reporting, a resurgence of pertussis might be expected in some countries, as a simple mathematical consequence of a history of insufficient vaccination (Riolo et al., 2013).

Disentangling the many pathways to pertussis resurgence is particularly difficult because pertussis immunity is not well-understood. With no known, reliable markers of pertussis immunity (Farizo et al., 2014), the properties of infection- and vaccine-derived protection against pertussis must be inferred indirectly. However, the models of pertussis immunity that best fit individual level clinical data and population level incidence data, respectively, are strikingly different. While the data from reinfection studies (Jenkinson, 1988) and animal models (Warfel et al., 2014) can be explained by a vaccine that protects against disease for a limited duration and may not protect against transmission at all, large scale pertussis incidence data are more consistent with long-lasting vaccine-derived immunity and sufficiently little transmission of pertussis among vaccinated and previously infected individuals as to confer some protection on unvaccinated individuals via herd-immunity (Nielsen and Larsen, 1994; Wearing and Rohani, 2009; Blackwood et al., 2013).

Despite the uncertainty surrounding the properties of pertussis vaccines, several countries experiencing increased pertussis incidence have supplemented existing infant vaccinations with additional booster vaccinations (Zepp et al., 2011). We investigated which booster schedules afforded the greatest reduction in disease burden for the least logistical and monetary cost. Because the causes of the pertussis resurgence remain unclear, and need not be the same in every location, we consider four scenarios in which infant vaccination alone may fail to eliminate pertussis:

Scenario I: Insufficient vaccine coverage. Perhaps the simplest reason a disease may persist despite infant vaccination is if too few infants are vaccinated to curtail transmission. This scenario would describe, for example, Thailand and Italy for several decades after the introduction of pertussis vaccines, where an increase in vaccine coverage, such as has occurred in the past three decades in both countries, has the potential to substantially reduced pertussis incidence. In many other regions, low vaccine coverage is likely to remain an important factor in the persistence of pertussis transmission. As of October 2013, WHO estimates of vaccine coverage suggest that in 10% of countries for which vaccination data is available at most 80% of infants receive the recommended three doses of pertussis vaccine (World Health Organization Department of Immunization, Vaccines And Biologicals, 2013). Even in nations with high overall vaccination rates, communities with high concentrations of unvaccinated or undervaccinated children may be at risk of outbreaks, as seems to be the case in some regions of the US (Omer et al., 2009).

Scenario II: Low efficacy. Primary vaccine failure, where a vaccine sometimes fails to provide protection when administered, is another mechanism by which a disease might persist despite infant vaccination efforts. For example, the outbreaks of pertussis in Canada beginning in the late 1980s have largely been attributed to the poor efficacy of the Connought Laboratories adsorbed vaccine that was used between 1985 and 1998 (Halperin et al., 1989; Ntezayabo et al., 2003). More generally, the lack of serological measures of immunity to pertussis makes it very difficult to determine the rates of primary vaccine failure, much less distinguish such failures in “take” from other phenomena such as waning immunity.

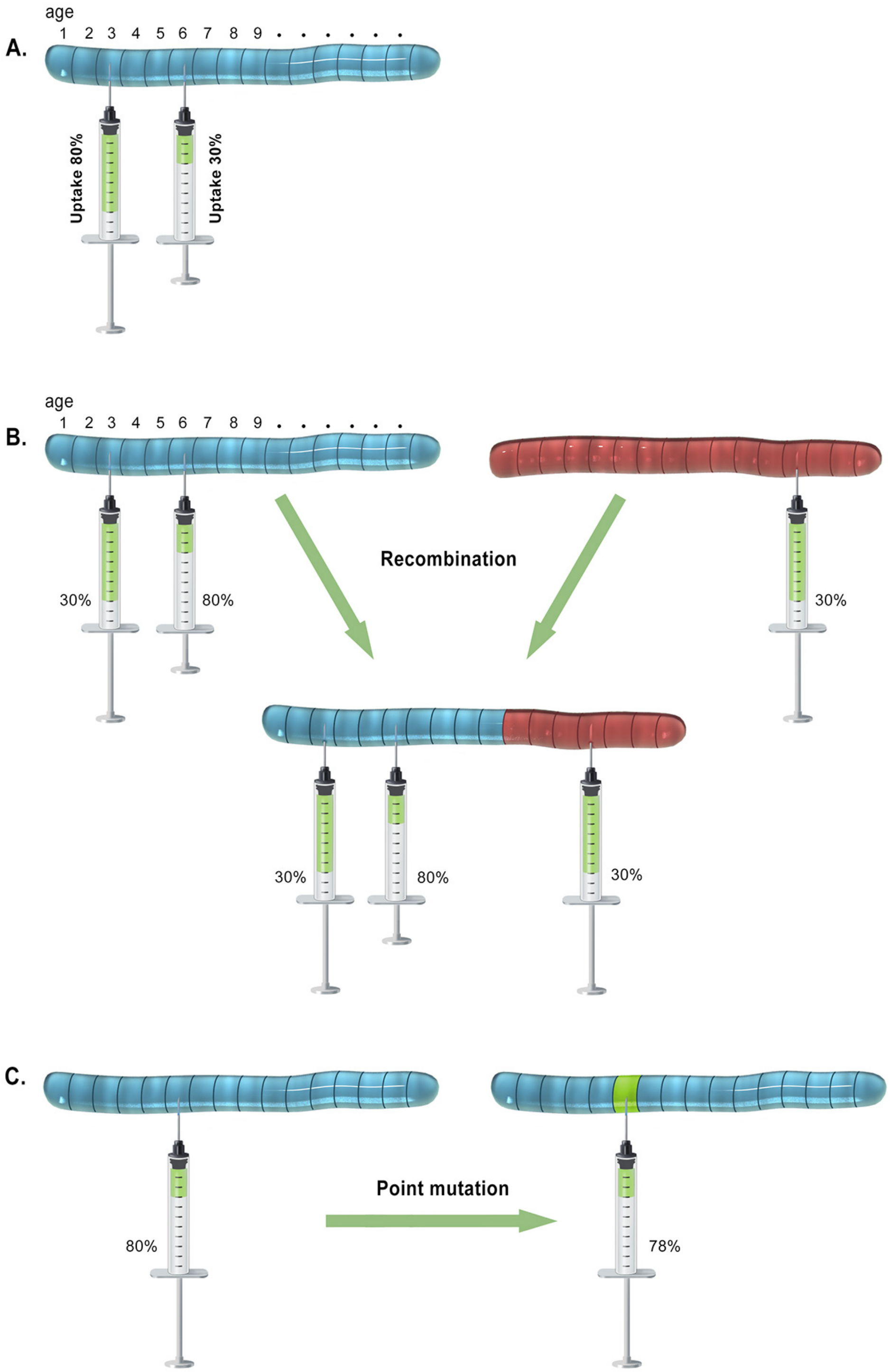
Scenario III: Waning immunity. Even if vaccination is widespread and initially efficacious, infant vaccination alone may fail to achieve elimination if vaccine derived immunity is only temporary. Some waning of vaccine derived immunity to pertussis is supported by data at both the population and individual scales, but while population level patterns of incidence generally suggest long-lasting immunity (Wearing and Rohani, 2009), individual level data on reinfections is better explained by a much shorter duration of protection (Jenkinson, 1988).

Scenario IV: leaky immunity. One potential mechanism for the changing epidemiology of pertussis is the evolution of *B. pertussis* (Mooi et al., 2001). In the case of multiple circulating strains of pertussis, for which vaccination induces at most limited cross-immunity, the probability of developing a transmissible, symptomatic infection after contact with an infectious individual may be reduced (but not eliminated) by vaccination. Such "leakiness" (Halloran et al., 1992) in vaccine derived immunity is another mechanism with the potential to facilitate the persistence of pertussis despite infant vaccination.

For simplicity, our current study considers only the extreme cases where one of these problems alone is responsible for ongoing pertussis transmission. However, finding successful, cost-effective booster schedules remains a non-trivial task. Because the space of possible schedules is so high-dimensional and the dynamics of infectious diseases are so nonlinear, many traditional optimization tools are ill-suited for the problem.

Here, we use a genetic algorithm (GA) to evolve cost-effective booster schedules by simulating the operation of natural selection on a "population" of schedules for booster vaccination (Holland, 1992). Each strategy in the population has a genotype that encodes its prescribed schedule of booster vaccinations (Fig. 3.1A). The choice of how to encode strategies can have important effects on the performance of a genetic algorithm, affecting both linkage and epistasis between genes as well as the topographical features of the fitness space (Davidor, 1990; Jones and Forrest, 1995). In the current study, we encode booster schedules on a single chromosome containing

Figure 3.1: Schematic showing the operation of the genetic algorithm on booster strategies. Panel A illustrates a single chromosome representation of a booster strategy, with each site on the chromosome determining the vaccine uptake of an age group. Panel B illustrates a recombination of two parent schedules. Panel C illustrates a point mutation changing the scheduled vaccine uptake in a single age group.



the probability of vaccination for each age cohort, to be carried out at the start of the each school term.

To evaluate the fitness of a strategy, we simulate transmission dynamics following the introduction of the booster schedule using an age-structured SEIR model. Individuals are born susceptible, with routine infant vaccination occurring at five months of age. Susceptible individuals become exposed through contact with infectious individuals and, after a latent period, become infectious. After recovering, individuals are immune to further infection. In the current study, we consider infection derived immunity to be lifelong, based on previous studies of pertussis in Sweden and Thailand, which found that previously infected individuals contributed very little to transmission (Rohani et al., 2010; Blackwood et al., 2013). The number of individuals of each age in each state (susceptible, exposed, infectious, recovered, and vaccinated) is updated stochastically with a fixed time step of one week.

Because the past history of pertussis incidence and infant vaccination are likely to have long-lasting effects on ongoing transmission dynamics (Riolo et al., 2013), we simulated pre-booster conditions with the same history of vaccination and chose parameters of coverage, efficacy, duration, and leakiness to give the same overall disease burden (measured in DALYs) in the two decades preceding the booster campaign (Fig. 3.2A-D).

We can think of the simulated transmission dynamics under a booster schedule as the phenotype of that strategy, with the combined cost of infections and vaccinations incurred during the simulation determining the strategy's fitness. We assessed the cost of infections based on the expected disability-adjusted life-years (DALYs) lost, calculated as the expected duration of illness plus, in the case of a death, the expected years of life lost (average lifespan minus the age at death). For simplicity, we assume that the costs of vaccination scale linearly with the number of doses given, rather than considering aspects of vaccination effort which may be nonlinear.

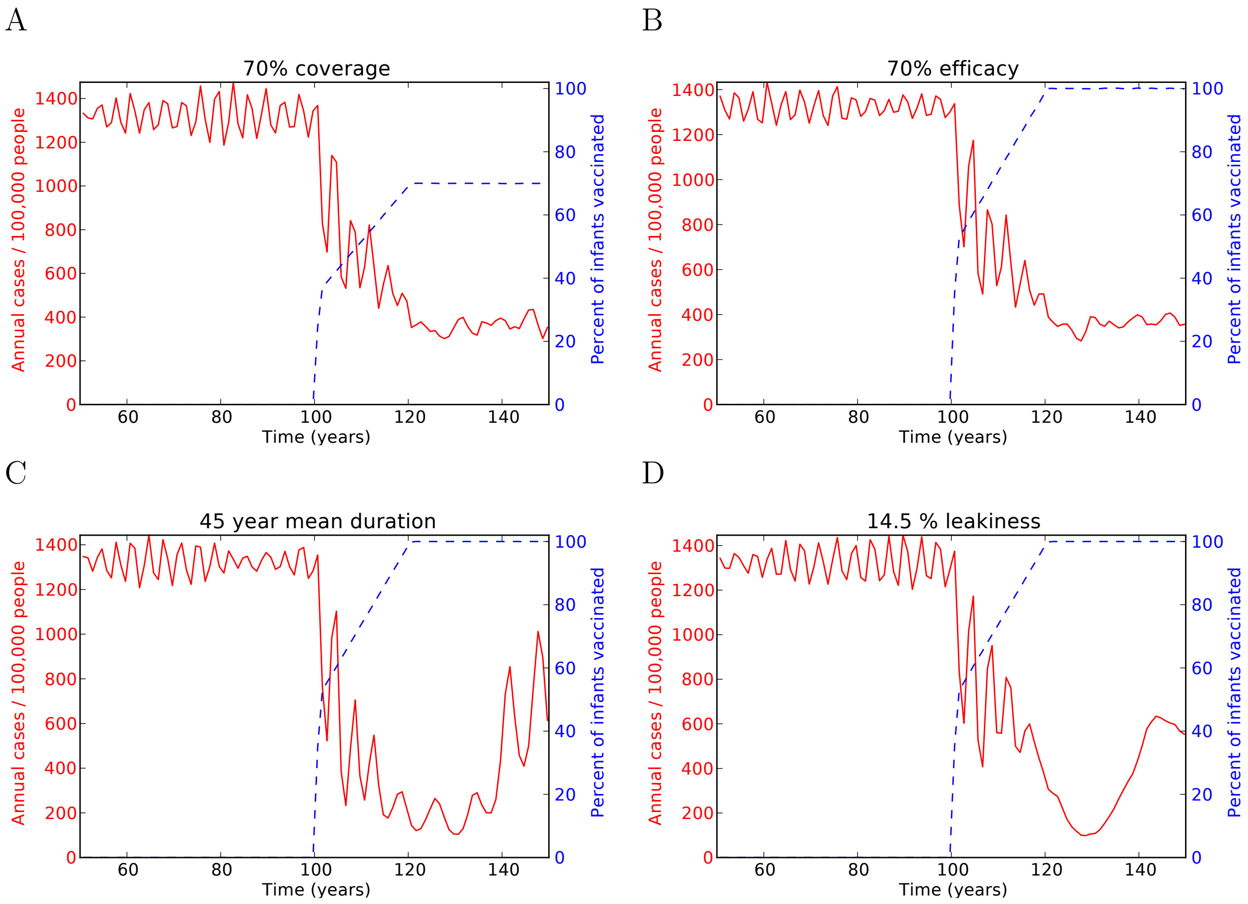


Figure 3.2: Pre-boosting incidence (solid red line) and infant vaccine uptake (dashed blue line) with 70% coverage (A), 70% efficacy (B), a 45 year mean duration of immunity (C), leaky immunity preventing 85.5% of infections (D). In each case, we simulated 100 years pre-vaccine (plotted starting at year 50). In year 100, vaccination began at 60% of its eventual coverage level and increased linearly over 20 years.

To generate a new population of booster schedules, parent schedules are chosen using "tournament selection"(Goldberg and Deb, 1991), in which pairs of schedules are selected uniformly at random with replacement to compete for the opportunity to reproduce. In each tournament, more fit schedule has a higher probability (in the current study, 90%) of winning. Carrying out N such tournaments, where N is the population size (in the current study, 2000 individuals), yields a pool of N parents. Note that while more fit strategies are likely to appear more often as parents than less fit strategies, a strategy's reproductive opportunities are only determined by its rank. This has the benefit of being less sensitive to the particular fitness function used than a selection method relying on the actual fitness values of strategies would be.

Once N parents have been selected, these parents are divided into pairs and each (not necessarily unique) pair of parents produces a pair of offspring. The parent chromosomes undergo crossover at a randomly chosen point along the chromosome (Fig. 3.1B). In each child chromosome, point mutations occur at a fixed rate μ (set to 0.05 per site in our experiments, which yields an average of 1.25 mutations per chromosome (Fig. 3.1C). This process leaves us with a new generation of booster strategies, ready to be evaluated through simulation.

Visualizing the evolving population of booster schedules presents its own challenges. While an average strategy can provide a rough idea of a genetic algorithm's behavior, what we are often interested in is more akin to a "typical good strategy," particularly if our algorithm may have discovered multiple, comparably fit strategies. Continuing the analogy of booster schedules as genotypes, we are interested in something like the consensus sequence of the most successful species-like clusters of schedules. To this end, we constructed a nearest-neighbor network of the most successful schedules found by the GA and applied a spectral community detection algorithm to find clusters in the network (Riolo and Newman, 2012). This allows us to visualize the distribution of alleles within each cluster of schedules in a more intuitive way, such as looking at the interquartile range of booster coverage in each age group, without losing as much information about the correlation between traits as we would by looking at the same summary statistics over the whole population.

In all four scenarios, the GA rapidly converges on a population of booster schedules with lower combined vaccination and disease costs than the initial, randomly generated population of strategies (Fig. B.1). When vaccine coverage of infants was insufficient (Scenario I), the successful booster schedules found by our GA had, on average, very little vaccination in older age groups, focusing most of their vaccination effort on young children (3.3A), which achieved more reduction of disease for less vaccination effort than the initial random population of strategies (Fig. B.2 & B.3).

For an infant vaccination program with low coverage (Scenario II), we found five distinct clusters of strategies, each corresponding to adding a single booster at 1, 2, 3, 4, or 5 years of age. Our results for a low efficacy vaccine were very similar (Fig. B.4B, Fig. 3.3B), with the exception that a higher vaccination effort was necessary to achieve the same effective coverage from the booster vaccine. In both cases, the most effective strategies came in the form of a single pre-school booster, the timing of which made little difference to fitness as long as it came before children entered school, with the high contact-rates and strong age-assortative mixing that come with it (Fig. 3.5A,B).

Booster schedules for a waning vaccine (Scenario III) differ greatly from those evolved to deal with low coverage or low efficacy (Fig. 3.3C). The most successful schedules call for one booster in the late teens and a second booster anytime between 35 and 45 years of age (with the exception of the strategy group shown in purple which does not include a single teenage booster but would, on average, be expected to include one booster between the ages of 15 and 35). These differences can be understood as a difference in the age-distribution of susceptible individuals. When

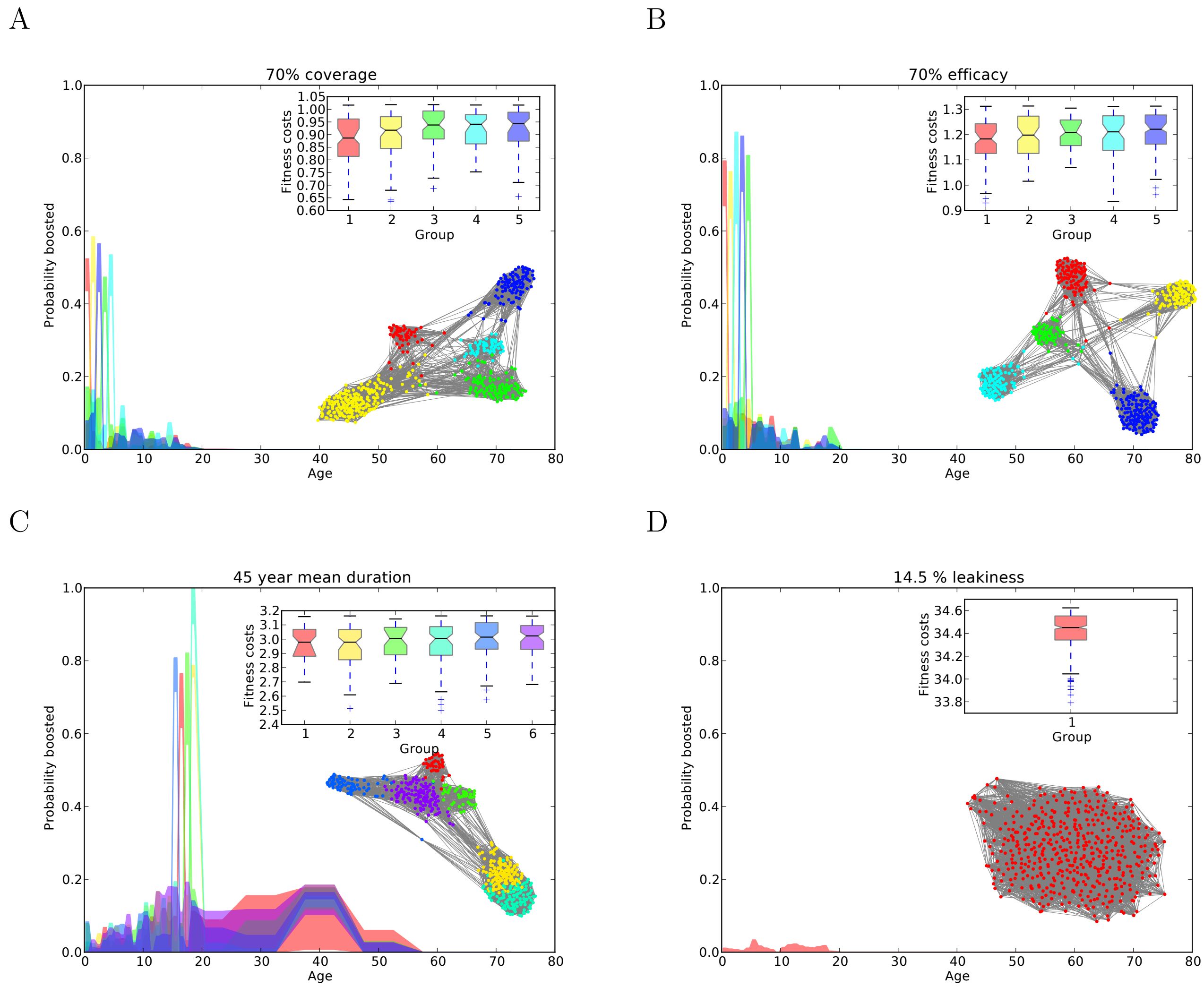


Figure 3.3: Best booster schedules found for 70% infant coverage (A), 70% efficacy (B), a 45 year mean duration of immunity (C), leaky immunity preventing 85.5% of infections (D). The best 500 booster schedules found in the last algorithm generation of each scenario are arranged into a network where each schedule is connected to the nearest 25 (under L_1 distance between genomes) strategies. The strategies are the clustered using a spectral partitioning algorithm (cluster indicated in the figure by color). The large plot shows the interquartile radius within each cluster of coverage at each age group (colors in plot correspond to colors in network). The inset plot shows the fitness costs (in DALYs) of strategies in each family, ordered from most fit (red, left) to least fit (purple, right).

we compare booster schedules evolved for vaccines with other durations of immunity (Fig. 3.4), we see that, although the number and timing of vaccinations depend on the rates of waning, successful schedules maintain vaccine-derived protection in large fraction of the population aged 5-45 years, particularly the school-aged children.

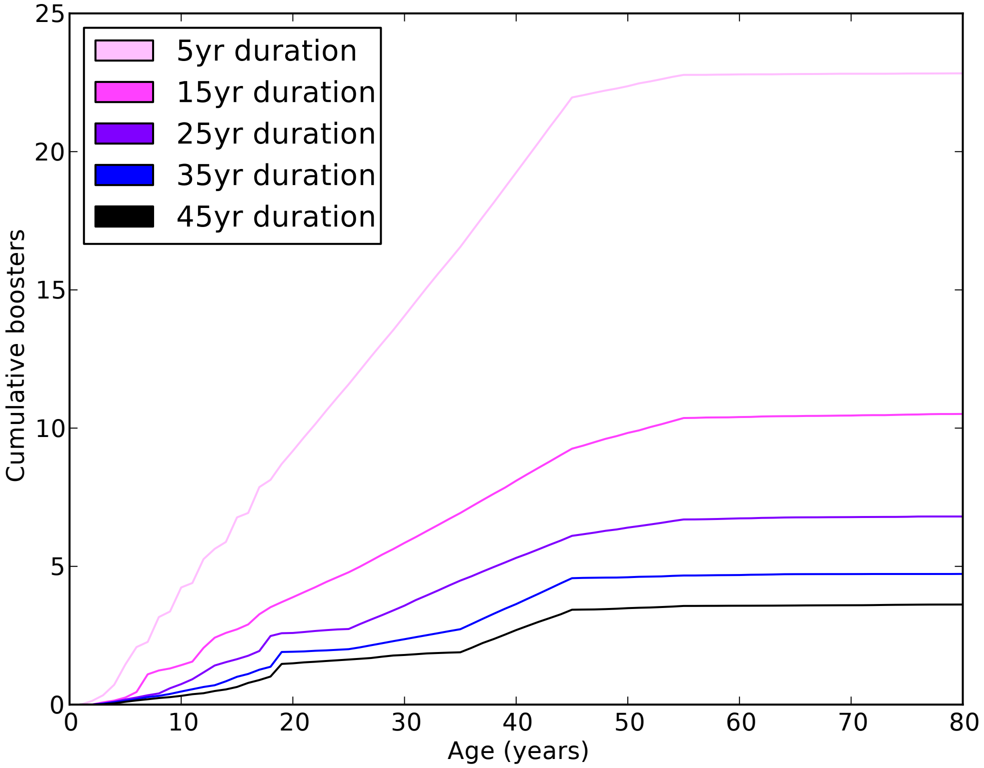


Figure 3.4: The cumulative number of booster vaccinations received over a lifetime, plotted for varying durations of vaccine derived immunity. In each case, the mean number of doses is plotted for the top quartile of booster schedules in the last generation of the genetic algorithm.

Leaky immunity (Scenario IV) presents yet another picture. The GA finds no booster schedules that effectively reduce the disease burden (Fig. B.2D). Looking at the nearest-neighbor network of the most successful strategies, there is no discernible community structure and the best strategies eschew booster vaccination (Fig. 3.3D). Again, the explanation lies in the patterns of susceptibility in the population. With infant vaccination already granting leaky protection to everyone, any remaining transmission already occurs among vaccinated individuals and cannot be disrupted by additional booster vaccinations (Fig. 3.5D). (Results for the scenario where booster vaccinations cause immunity to become less leaky are presented in SOM and resemble the results for a low level of primary vaccine failure.)

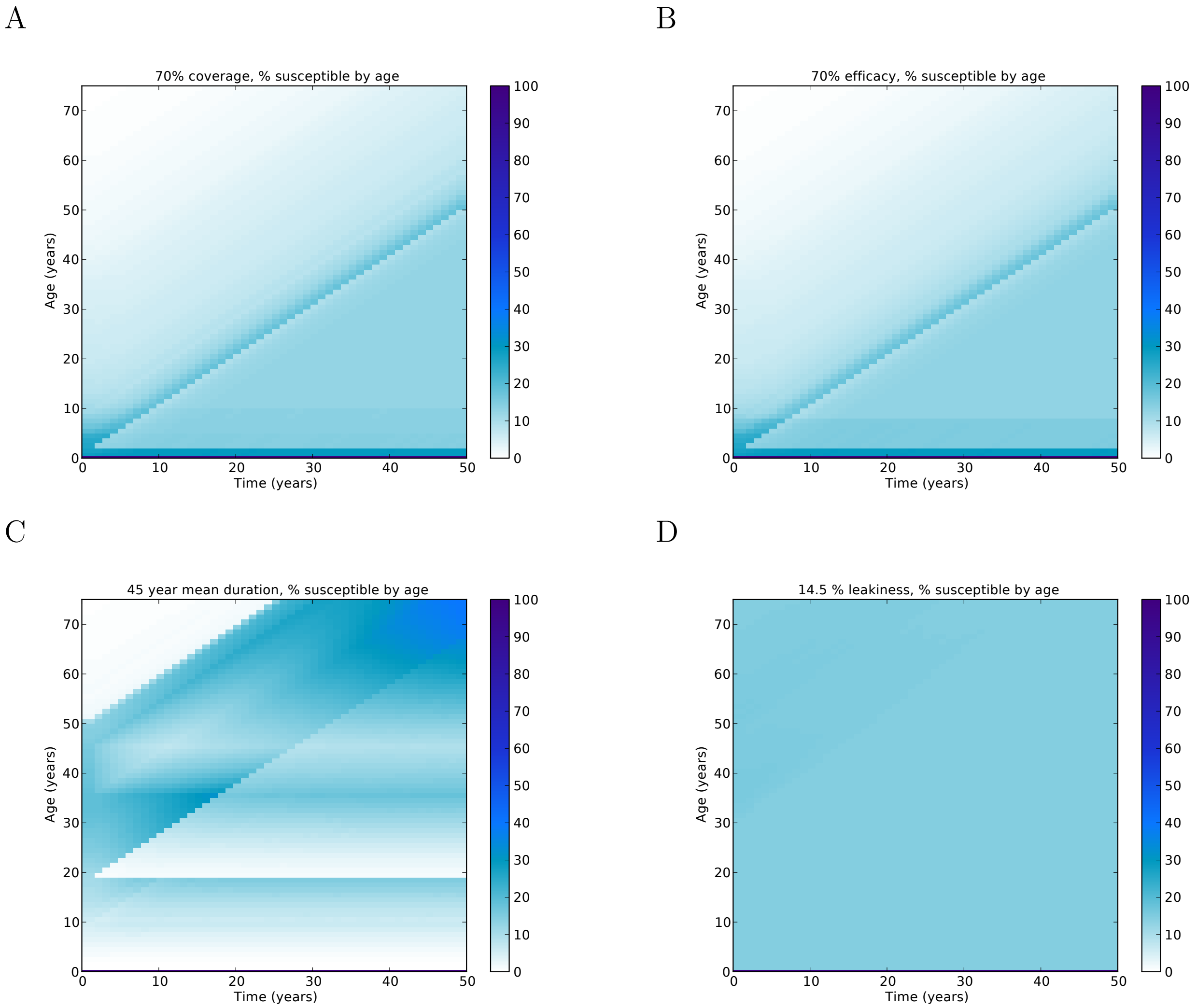


Figure 3.5: Susceptibilty by age under the most successful booster schedules for 70% infant coverage (A), 70% efficacy (B), a 45 year mean duration of immunity (C), leaky immunity preventing 85.5% of infections (D), plotted during the first fifty years following the introduction of booster vaccinations. Color intensity indicates the percentage of individuals in each age group who are susceptible at each time. In the leaky immunity case, 14.5% of vaccinated individuals are included in this calculation.

Our model is relatively simple and leaves out many mechanisms that may play important roles in pertussis transmission and control, including the spatial or social aggregation of unvaccinated individuals (Omer et al., 2009), the occurrence of asymptomatic cases of pertussis (de Greeff et al., 2010), and household structure (de Greeff et al., 2010). Perhaps equally importantly, our model does not attempt to mimic the real vaccination history and demographics of any place, instead exploring the relatively simple case of a linear increase in vaccine uptake followed by sustained, high levels of infant vaccination.

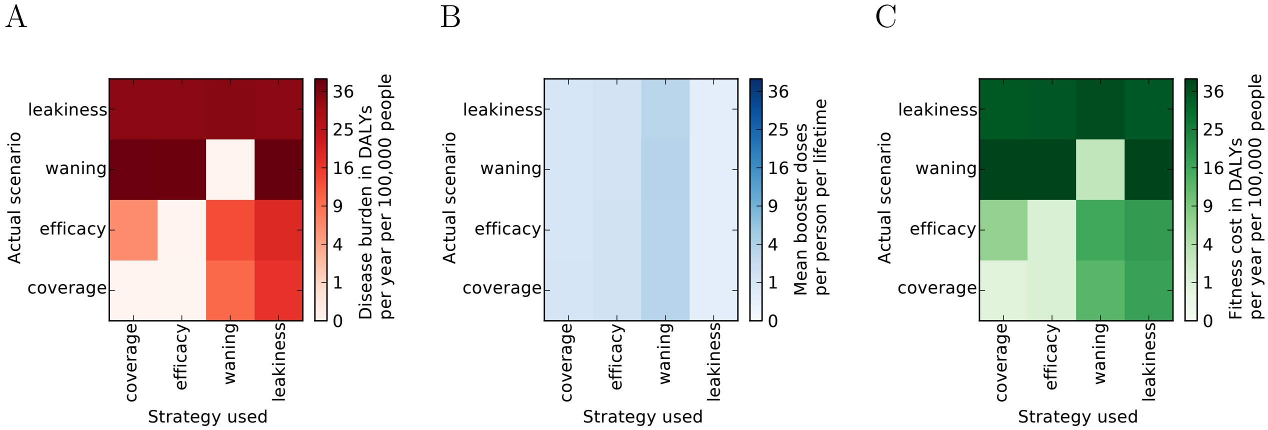


Figure 3.6: The disease burden (A), vaccine doses used (B), and total fitness cost (C) of applying a boosting schedule evolved for each scenario in each other scenario. The representative boosting schedules are constructed by taking the mean coverage at each age in the most fit strategy families. These representative schedules (Strategy used) are applied to each scenario (Actual scenario) and the mean fitness cost among 120 replicant runs is plotted.

While our results do not predict the quantitative impact of any particular vaccine schedule on any particular population, they nevertheless have implications for vaccine policy. We find that optimal booster schedules for controlling pertussis differ greatly for differing mechanisms of vaccine failure and that, for example, a booster vaccine schedule optimized to compensate for primary vaccine failure may be extremely ineffective in controlling a pertussis resurgence caused by waning immunity and *vice versa* (Fig. 3.6). Our results suggest that understanding pertussis immunity is critical to developing cost-effective control strategies.

CHAPTER IV

First-principles multiway spectral partitioning of graphs

4.1 Introduction

In host-pathogen systems, the pattern of contacts between hosts can have important consequences for disease dynamics and control (Chapter II&III, Newman, 2002). One problem of ongoing interest is, given a network of contacts between hosts, how to optimally target interventions against a disease (Eubank et al., 2004; Meyers et al., 2005; Salathé et al., 2010).

Let us consider the simple, extreme case of targeted vaccination with a highly effective vaccine against a highly transmissible disease. If we can expect each infected node to spread the disease to all of its neighbors, then the problem is to vaccinate individuals (*i.e.* remove vertices) in the network of susceptible hosts in such a way as to partition the network into smaller groups, each isolated from the others. In particular, for n is the total number of susceptibles and n_i the number in group i, we want to minimize the expected outbreak size

$$\sum_i \frac{n_i}{n} n_i$$

where $\frac{n_i}{n}$ is the probability that an outbreak starts in group i and n_i is the number of individuals who will be infected if it does, subject to whatever constraints are applied to the vaccination effort.

In the current study we will consider a related but more basic graph partitioning problem, where one is given an undirected, unweighted graph of n vertices and asked to divide the vertices into k nonoverlapping groups of given sizes, such that the number of edges running between groups—the so-called *cut size*—is minimized. This is known to be a computationally hard problem. Even for the simplest case where $k = 2$ it is NP-hard to find the division with minimum cut size (Garey and Johnson, 1979). Good approximations to the minimum can, however, be found using a variety of heuristic methods, including local greedy algorithms, genetic algorithms, tabu search, and multilevel algorithms. One particularly elegant and effective approach, which is the subject of this paper, is *spectral partitioning* (Von Luxburg, 2007), which makes use of the spectral properties of any of several matrix representations of the graph, most commonly the graph Laplacian.

The first Laplacian spectral partitioning algorithms date back to the work of Fiedler in the 1970s (Fiedler, 1973; Pothen et al., 1990) and were aimed at solving the graph bisection problem, i.e., the problem of partitioning a graph into just two parts. For this problem the underlying theory of the spectral method is well understood and the algorithms work well. One calculates the eigenvector corresponding to the second lowest eigenvalue of the graph Laplacian and then divides the graph according to the values of the vector elements—the complete process is described in Section 4.2. More recently, attention has turned to the general multiway partitioning problem with arbitrary k, which is harder. One elementary approach is to repeatedly bisect the graph into smaller and smaller parts using Fiedler's method or one of its variants (Simon and Teng, 1997), but this can give rise to poor solutions in some commonly occurring situations. A better approach, and the one in widest current use, is to construct the $n \times (k - 1)$ matrix whose columns are the eigenvectors for the second- to kth-lowest eigenvalues of the Laplacian (the $n \times n$ matrix containing the degree of each vertex on the diagonal, a -1 in each position i, j where there is an edge between vertex i and vertex j, and zeros elsewhere). The rows of this matrix define n vectors of $k-1$ elements each which are regarded as points in a $(k-1)$-dimensional space. One clusters these points into k groups using any of a variety of heuristics, the most common being the k-means method, and the resulting clusters define the division of the graph.

This method works well in practice, giving good results on a wide range of test graphs, and if one is concerned only with finding an algorithm that works, one need look no further. Formally, however, it does have some drawbacks. First, it is not a true generalization of the method for $k = 2$. If one were to apply this algorithm to a $k = 2$ problem, one would end up performing k-means on the second eigenvector of the graph Laplacian, which is a different procedure from the standard $k = 2$ algorithm. Second, the algorithm is not normally even derived directly from the minimum-cut problem. Instead, the algorithm is typically proposed without justification, and justified after the fact by demonstrating that it performs well on particular partitioning tasks. This approach is perfectly correct but somewhat unsatisfactory in that it does not give us much understanding of why the calculation works. For that, it would be better to derive the algorithm from first principles. The purpose of this paper is to give such a derivation for multiway spectral partitioning. Our main goal in doing so is to gain an understanding of *why* spectral partitioning works, by contrast with the traditional presentation which demonstrates only that it does work. However, as we will see, the algorithm that we derive in the process is different in significant ways from previous spectral partitioning algorithms and in Section 4.4 we present results that indicate that our algorithm can outperform more conventional approaches for certain classes of graphs.

The previous literature on spectral partitioning is extensive—this has been an active area of research, especially in the last few years. There is also a large literature on the related problem of spectral clustering of high-dimensional data sets, which can be mapped onto a weighted partitioning problem on a complete graph using an affinity matrix. The 1995 paper of Alpert and Yao (Alpert and Yao, 1995) provides an early example of an explicit derivation of a general multiway partitioning algorithm. Their algorithm is substantially different from the most commonly used variants, involving a vector partitioning step, and also contains one arbitrary parameter which affects the performance of the algorithm but whose optimal value is unknown. The algorithms of Shi and Malik (Shi and Malik, 2000) and Meilă and Shi (Meilă and Shi, 1995) are good examples of the standard multiway partitioning using k-means, although applied to the slightly different problem of normalized-cut partitioning. A number of subsequent papers have analyzed these algorithms or variants of them (Ng et al., 2001; Meilă and Xu, 2003; Kannan et al., 2004; Lee et al., 2012). Summaries are given by von Luxburg (Von Luxburg, 2007), Verma and Meilă (Verma and Meilă, 2003), and Bach and Jordan (Bach and Jordan, 2006), although the discussions are in the language of data clustering, not graph partitioning. Tu, Shieh, and Cheng (Tu et al., 2000) derive a k-way partitioning algorithm based on the spectrum of the graph's adjacency matrix which has some features that seem in some sense analogous to those of our graph Laplacian based algorithm, such as stretching the eigenvectors used according to the sizes of the different groups. Perhaps the work that comes closest to our own is that of Zhang and Jordan (Zhang et al., 2008), again on data clustering, in which partitions are indexed using a set of $(k-1)$-dimensional "margin vectors," which are oriented using a Procrustes technique. We also use Procrustes analysis in one version of the method we describe, although other details of our approach are different from the method of Zhang and Jordan.

The outline of this paper is as follows. In Section 4.2 we review the derivation of the standard spectral bisection algorithm and then in Section 4.3 present in detail the generalization of that derivation to the multiway partitioning problem, leading to an algorithm for multiway partitioning of an arbitrary undirected graph into any number of groups of specified sizes. In Section 4.4 we give example applications of this algorithm to a number of test graphs, and demonstrate that its performance is similar to, or in some cases slightly better than, approaches based on k-means. In Section 4.5 we give our conclusions and discuss directions for future research.

4.2 Spectral bisection

The term *spectral bisection* refers to the special case in which we partition a graph into exactly $k = 2$ parts. For this case there is a well-established first-principles derivation of the standard partitioning algorithm, which we review in this section. Our goal in subsequent sections will be to find a generalization to the case of arbitrary k.

Suppose we are given an undirected, unweighted graph on n vertices, which we will assume to be connected (i.e., to have only one component), and we wish to divide its vertices into two groups which, for the sake of simplicity, we will take to be of equal size $\frac{1}{2}n$ (with n even). We define an index variable s_i for each vertex $i = 1 \ldots n$ such that $s_i = 1$ if vertex i belongs to group 1 and $s_i = -1$ if i belongs to group 2. We note that

$$\tfrac{1}{2}(s_i s_j + 1) = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are in the same group,} \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

Thus the number of edges within groups is given by $\frac{1}{2}\sum_{ij}\frac{1}{2}(s_i s_j + 1)A_{ij}$, where A_{ij} is an element of the adjacency matrix (having value 1 if there is an edge between i and j, and zero otherwise) and the extra factor of $\frac{1}{2}$ compensates for double counting of vertex pairs in the sum. The total number of edges in the entire graph is $\frac{1}{2}\sum_{ij} A_{ij}$

and hence the number of edges between groups—which is the cut size R—is given by

$$\begin{aligned} R &= \tfrac{1}{2}\sum_{ij} A_{ij} - \tfrac{1}{4}\sum_{ij}(s_i s_j + 1)A_{ij} \\ &= \tfrac{1}{4}\sum_{ij}(1 - s_i s_j)A_{ij} = \tfrac{1}{4}\sum_{i} d_i - \tfrac{1}{4}\sum_{ij} s_i s_j A_{ij}, \end{aligned} \tag{4.2}$$

where $d_i = \sum_j A_{ij}$ is the degree of vertex i. Noting that $s_i^2 = 1$ for all i, this equation can be rewritten as

$$R = \tfrac{1}{4}\sum_{ij} d_i \delta_{ij} s_i s_j - \tfrac{1}{4}\sum_{ij} A_{ij} s_i s_j = \tfrac{1}{4}\sum_{ij} L_{ij} s_i s_j, \tag{4.3}$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. $L_{ij} = d_i \delta_{ij} - A_{ij}$ is the ijth element of the graph Laplacian matrix $\mathbf{L}$, which contains the degrees of vertices on the diagonal and -1s denoting edges. Alternatively, we can write R in matrix notation as

$$R = \tfrac{1}{4}\mathbf{s}^T \mathbf{L} \mathbf{s}, \tag{4.4}$$

where $\mathbf{s}$ is the n-component vector with elements s_i.

Our goal, for a given graph and hence for given $\mathbf{L}$, is to minimize the cut size R over possible bisections of the graph, represented by $\mathbf{s}$, subject to the constraint that the two groups are the same size, which is equivalent to saying that $\sum_i s_i = 0$ or

$$\mathbf{1}^T \mathbf{s} = 0, \tag{4.5}$$

where $\mathbf{1}$ is the uniform vector $(1, 1, 1, \ldots)$. (Note that, for division into two groups of n_1 and n_1 vertices, respectively, this constraint would be replaced by $\sum_i s_i = n_1 - n_2$.) Unfortunately, as mentioned in the introduction, this is a hard computational problem. But one can in many cases find good approximate solutions in polynomial time by using a *relaxation method.* We generalize the discrete variables $s_i = \pm 1$ to

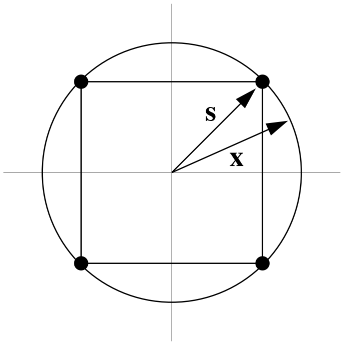


Figure 4.1: The possible values of the vector **s** lie at the corners of a hypercube in n-dimensional space, while the relaxed vector **x** can lie at any point on the circumscribing hypersphere.

continuous real variables x_i and solve the relaxed minimization with respect to the vector $\mathbf{x} = (x_1, x_2, \ldots)$ of

$$R_x = \tfrac{1}{4}\mathbf{x}^T\mathbf{L}\mathbf{x}, \tag{4.6}$$

subject to the constraint

$$\mathbf{1}^T\mathbf{x} = 0, \tag{4.7}$$

which is the equivalent of Eq. (4.5). One must however also apply an additional constraint to prevent **x** from becoming zero, which is normally taken to have the form

$$\mathbf{x}^T\mathbf{x} = n. \tag{4.8}$$

Choices of **x** satisfying this second constraint include all allowed values of the original unrelaxed vector **s**, since $\mathbf{s}^T\mathbf{s} = \sum_i s_i^2 = \sum_i 1 = n$, but also include many other values in addition. Geometrically, one can think of **s** as defining a point in an n-dimensional space, with the allowed values $s_i = \pm 1$ restricting the point to fall at one of the corners of a hypercube. The value of **x** falls on the circumscribing hypersphere, since $\mathbf{x}^T\mathbf{x} = \sum_i x_i^2 = n$ implies that **x** has constant length $\sqrt{n}$. The hypersphere coincides with the values of **s** at the corners of the hypercube, but includes other values in between as well—see Fig. 4.1.

The relaxed problem is straightforward to solve by differentiation. We enforce the two conditions (4.7) and (4.8) with Lagrange multipliers λ and μ so that

$$\frac{\partial}{\partial x_k}\left[\sum_{ij} L_{ij}x_ix_j - \lambda\sum_i x_i^2 - \mu\sum_i x_i\right] = 0. \tag{4.9}$$

Performing the derivatives we find that $2\sum_j L_{kj}x_j - 2\lambda x_k - \mu = 0$ or in matrix notation $\mathbf{Lx} = \lambda\mathbf{x} + \frac{1}{2}\mu\mathbf{1}$. Multiplying on the left by $\mathbf{1}$ we get $\mathbf{1}^T\mathbf{Lx} = \lambda\mathbf{1}^T\mathbf{x} + \frac{1}{2}n\mu$, and employing Eq. (4.7) and noting that $\mathbf{1}$ is an eigenvector of $\mathbf{L}$ with eigenvalue zero, we find that $\mu = 0$. Thus we have

$$\mathbf{Lx} = \lambda\mathbf{x}. \tag{4.10}$$

In other words $\mathbf{x}$ is an eigenvector of the graph Laplacian satisfying the two conditions (4.7) and (4.8).

Our solution is completed by noting that the cut size R_x within the relaxed approximation, evaluated at the solution of (4.10), is

$$R_x = \tfrac{1}{4}\mathbf{x}^T\mathbf{Lx} = \tfrac{1}{4}\lambda\mathbf{x}^T\mathbf{x} = \frac{n\lambda}{4}. \tag{4.11}$$

This is minimized by choosing λ as small as possible, in other words by choosing $\mathbf{x}$ to be the eigenvector corresponding to the lowest possible eigenvalue. The lowest eigenvalue of $\mathbf{L}$ is always zero, with corresponding eigenvector proportional to $\mathbf{1}$, but we cannot choose this eigenvector because it is forbidden by the condition $\mathbf{1}^T\mathbf{x} = 0$, which requires that the solution vector $\mathbf{x}$ be orthogonal to $\mathbf{1}$. (This is equivalent to saying that we are not allowed to put all vertices in the same one group, which would certainly ensure a small cut size, but wouldn't give a bisection of the graph.) Our next best choice is to choose $\mathbf{x}$ proportional to the eigenvector for the second-lowest eigenvalue, the so-called *Fiedler vector*.

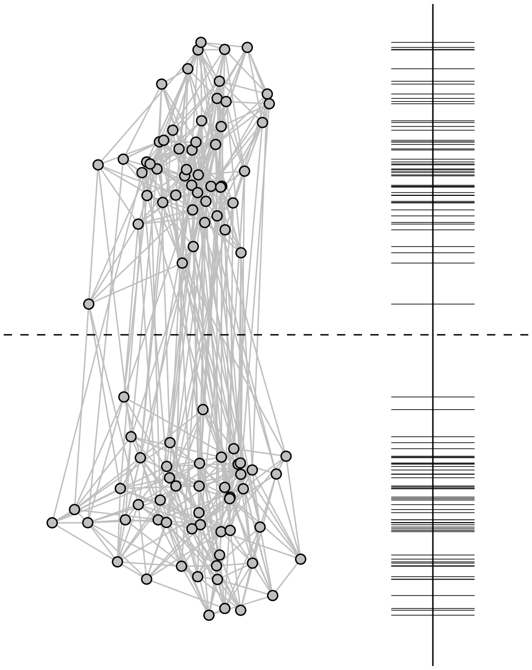


Figure 4.2: Left: a small, computer-generated graph with two equally sized groups of vertices. Right: the elements of the Fiedler vector—the second eigenvector of the graph Laplacian—plotted on an arbitrary scale. A division of the vertices into two groups according to the signs of these elements (the dashed line indicates zero) recovers the groups.

This solves the relaxed optimization problem exactly. The final step in the process is to “unrelax” back to the original variables $s_i = \pm 1$, which we do by rounding the x_i to the nearest value ± 1, which means that positive values of x_i are rounded to $+1$ and negative values to -1. Thus our final algorithm is a straightforward one: we calculate the eigenvector of the graph Laplacian corresponding to the second-lowest eigenvalue and then divide the vertices into two groups according to the signs of the elements of this vector. Although the solution of the relaxed optimization is exact, the unrelaxation process is only an approximation—there is no guarantee that rounding to ± 1 gives the correct optimum for the unrelaxed problem—and hence the overall algorithm only gives an approximate solution to the partitioning problem. In practice, however, it appears to work well. Figure 4.2 shows an example.

As we have described it, the algorithm above also does not guarantee that the two final groups of vertices will be of equal size. The relaxed optimization guarantees that $\sum_i x_i = 0$ because of Eq. (4.7), but we are not guaranteed that $\sum_i s_i = 0$ after the rounding procedure. Normally $\sum_i s_i$ will be close to zero, and hence the groups will be of nearly equal sizes, but there may be some imbalance. In typical usage, however, this is not a problem. In many applications one is willing to put up with a small imbalance anyway, but if one is not then a post-processing step can be performed that moves a small number of vertices between groups in order to restore balance.

4.3 Generalization to more than two groups

Our primary purpose in this paper is to give a generalization of the derivation of the previous section to spectral partitioning into more than two groups. The method we present allows the groups to be of any sizes we choose—they need not be of equal size as in our two-way example. As we will see, the algorithm we derive differs in significant ways from previous multiway partitioning algorithms.

4.3.1 Cut size for multiway partitioning

To generalize the spectral bisection algorithm to the case of more than two groups we need first to find an appropriate generalization of the quantities s_i used in Section 4.2 to denote membership of the different communities. For a partitioning into k groups, we propose using $(k-1)$-dimensional vectors to denote the groups, vectors that in the simplest case point to the k vertices of a $(k-1)$-dimensional regular simplex.

Let us denote by $\mathbf{w}_r$ with $r = 1 \ldots k$ a set of vectors pointing to the vertices of a regular $(k-1)$-dimensional simplex centered on the origin. For $k = 3$, for example, the three vectors would point to the corners of an equilateral triangle; for $k = 4$ the vectors would point to the corners of a regular tetrahedron, and so forth—see Fig. 4.3. Such simplex vectors are not orthogonal. Rather they satisfy a relation of the form

$$\mathbf{w}_r^T \mathbf{w}_s = \delta_{rs} - \frac{1}{k}. \tag{4.12}$$

(Note that the individual vectors are normalized so that $\mathbf{w}_r^T \mathbf{w}_r = 1 - 1/k$. One could normalize them to have unit length, as presented in section on spectral bisection, but subsequent formulas work out less neatly that way.)

We will use these simplex vectors as labels for the k groups in our partitioning problem, assigning one vector to represent each of the groups. All assignments are equivalent and any choice will work equally well. We label each vertex i with a vector variable $\mathbf{s}_i$ equal to the simplex vector $\mathbf{w}_r$ for the group r that it belongs to. Then Eq. (4.12) implies that

$$\mathbf{s}_i^T \mathbf{s}_j + \frac{1}{k} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are in the same group,} \\ 0 & \text{otherwise,} \end{cases} \tag{4.13}$$

which is the equivalent of Eq. (4.1) (up to the scaling of $\mathbf{s}_i$ and $\mathbf{s}_j$), and the derivation

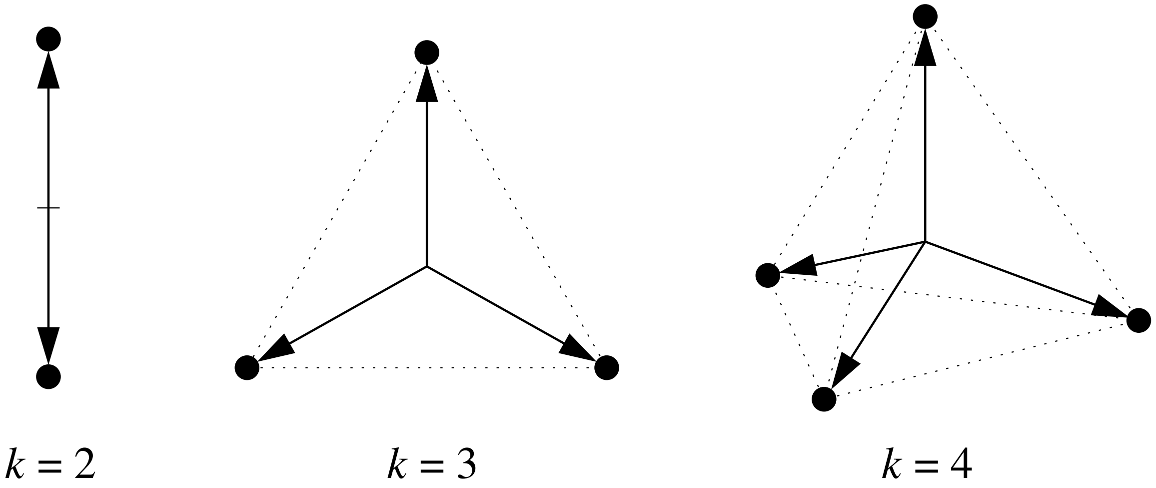


Figure 4.3: As group labels we use vectors $\mathbf{w}_r$ pointing from the center to the corners of a regular $(k-1)$-dimensional simplex. For $k = 2$ the simplex consists of just two points on a line, consistent with the indices $s_i = \pm 1$ used in Section 4.2. For $k = 3$ the simplex is an equilateral triangle; for $k = 4$ it is a regular tetrahedron. In higher dimensions it takes the form of the appropriate generalization of a tetrahedron to four or more dimensions, which would be difficult to draw on this two-dimensional page.

of the cut size follows through as before, giving

$$R = \tfrac{1}{2}\sum_{ij}(d_i\delta_{ij} - A_{ij})\mathbf{s}_i^T\mathbf{s}_j = \tfrac{1}{2}\sum_{ij} L_{ij}\mathbf{s}_i^T\mathbf{s}_j, \tag{4.14}$$

where L_{ij} is once again an element of the graph Laplacian matrix $\mathbf{L}$. Alternatively, we can introduce an $n \times (k-1)$ indicator matrix $\tilde{\mathbf{S}}$ whose ith row is equal to $\mathbf{s}_i$ and write the cut size in matrix notation as

$$R = \tfrac{1}{2}\,\mathrm{Tr}(\tilde{\mathbf{S}}^T\mathbf{L}\tilde{\mathbf{S}}). \tag{4.15}$$

These simplex vectors are not, however, the only possible choice for the vectors $\mathbf{s}_i$. In fact, we have a lot of latitude about our choice. The vectors can be translated, rotated, reflected, stretched or shrunk in a variety of ways and still give a simple expression for the cut size. If the groups into which our graph is to be partitioned are

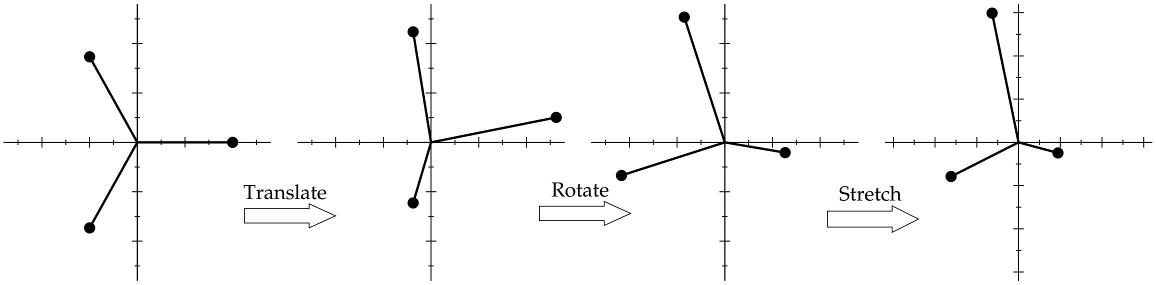


Figure 4.4: As labels for the groups we use vectors derived from the regular simplex vectors of Fig. 4.3 by a uniform translation, followed by a rotation and/or reflection, followed in turn by stretching or shrinking along each axis independently.

of equal size, then plain simplex vectors as above are a good choice, but for unequal groups it will be useful to consider group vectors of the more general form $\mathbf{DQ}(\mathbf{w}_r - \mathbf{t})$, where $\mathbf{D}$ is a $(k-1)\times(k-1)$ diagonal matrix, $\mathbf{Q}$ is a $(k-1)\times(k-1)$ orthogonal matrix, and $\mathbf{t}$ is an arbitrary vector. This choice takes the original simplex vectors and does three things, as illustrated in Fig. 4.4: first it translates them an arbitrary distance given by $\mathbf{t}$, then it rotates and/or reflects them according to the orthogonal transformation $\mathbf{Q}$, and finally it shrinks (or stretches) them independently along each axis by factors given by the diagonal elements of $\mathbf{D}$. The result is that the matrix $\tilde{\mathbf{S}}$ describing the division of the network into groups is transformed into a new matrix $\mathbf{S}$:

$$\mathbf{S} = (\tilde{\mathbf{S}} - \mathbf{1}\mathbf{t}^T)\mathbf{Q}^T\mathbf{D}. \tag{4.16}$$

Inverting this transformation, we get $\tilde{\mathbf{S}} = \mathbf{S}\mathbf{D}^{-1}\mathbf{Q} + \mathbf{1}\mathbf{t}^T$ and substituting this expression into Eq. (4.15), the cutsize can be written in terms of the new matrix as

$$\begin{aligned} R &= \tfrac{1}{2}\,\mathrm{Tr}\big[(\mathbf{Q}^T\mathbf{D}^{-1}\mathbf{S}^T + \mathbf{t}\mathbf{1}^T)\mathbf{L}(\mathbf{S}\mathbf{D}^{-1}\mathbf{Q} + \mathbf{1}\mathbf{t}^T)\big] \\ &= \tfrac{1}{2}\,\mathrm{Tr}(\mathbf{Q}^T\mathbf{D}^{-1}\mathbf{S}^T\mathbf{L}\mathbf{S}\mathbf{D}^{-1}\mathbf{Q}) = \tfrac{1}{2}\,\mathrm{Tr}(\mathbf{S}^T\mathbf{L}\mathbf{S}\mathbf{D}^{-2}), \end{aligned} \tag{4.17}$$

where in the second equality we have made use of $\mathbf{L1} = 0$.

The freedom of choice of the vector $\mathbf{t}$ and the matrices $\mathbf{Q}$ and $\mathbf{D}$ allows us to simplify our problem as follows. First, we will require that $\mathbf{S}^T\mathbf{1} = 0$. Taking the transpose of Eq. (4.16) and multiplying by $\mathbf{1}$, this implies that $(\tilde{\mathbf{S}}^T - \mathbf{t}\mathbf{1}^T)\mathbf{1} = 0$ and hence fixes the value of $\mathbf{t}$:

$$\mathbf{t} = \frac{1}{n}\tilde{\mathbf{S}}^T\mathbf{1} = \sum_r \frac{n_r}{n}\mathbf{w}_r, \tag{4.18}$$

where n_r is the number of vertices in group r. The condition $\mathbf{S}^T\mathbf{1} = 0$ is the equivalent of Eq. (4.5) in the two-group case, and, just as (4.5) does, it fixes the sizes of the groups, since the sum $\sum_r n_r\mathbf{w}_r$ occupies a unique point in the space of the simplex vectors for every choice of group sizes.

We would also like our matrix $\mathbf{S}$ to satisfy a condition equivalent to $\mathbf{s}^T\mathbf{s} = n$ in the two-group case, which will take the form $\mathbf{S}^T\mathbf{S} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. The freedom to choose $\mathbf{Q}$ and $\mathbf{D}$ allows us to do this. We note that

$$\mathbf{S}^T\mathbf{S} = \mathbf{D}\mathbf{Q}(\tilde{\mathbf{S}}^T - \mathbf{t}\mathbf{1}^T)(\tilde{\mathbf{S}} - \mathbf{1}\mathbf{t}^T)\mathbf{Q}^T\mathbf{D}, \tag{4.19}$$

and that the central product in this expression is a $(k-1)\times(k-1)$ symmetric matrix that expands as

$$\begin{aligned}(\tilde{\mathbf{S}}^T - \mathbf{t}\mathbf{1}^T)(\tilde{\mathbf{S}} - \mathbf{1}\mathbf{t}^T) &= \tilde{\mathbf{S}}^T\tilde{\mathbf{S}} - \tilde{\mathbf{S}}^T\mathbf{1}\mathbf{t}^T - \mathbf{t}\mathbf{1}^T\tilde{\mathbf{S}} + \mathbf{t}\mathbf{1}^T\mathbf{1}\mathbf{t}^T \\ &= \tilde{\mathbf{S}}^T\tilde{\mathbf{S}} - n\mathbf{t}\mathbf{t}^T = \sum_r n_r\mathbf{w}_r\mathbf{w}_r^T - \sum_{rs}\frac{n_r n_s}{n}\mathbf{w}_r\mathbf{w}_s^T,\end{aligned} \tag{4.20}$$

where we have used (4.18).

We perform an eigenvector decomposition of this matrix in the form $\mathbf{U}\boldsymbol{\Delta}\mathbf{U}^T$, where where $\mathbf{U}$ is an orthogonal matrix and $\boldsymbol{\Delta}$ is the diagonal matrix of eigenvalues, which are all nonnegative since the original matrix is a perfect square. Then we let

$$\mathbf{Q} = \mathbf{U}^T, \qquad \mathbf{D} = \boldsymbol{\Delta}^{-1/2}, \tag{4.21}$$

and we have

$$\begin{aligned}\mathbf{S}^T\mathbf{S} &= \mathbf{D}\mathbf{Q}(\tilde{\mathbf{S}}^T - \mathbf{t}\mathbf{1}^T)(\tilde{\mathbf{S}} - \mathbf{1}\mathbf{t}^T)\mathbf{Q}^T\mathbf{D} \\ &= \boldsymbol{\Delta}^{-1/2}\mathbf{U}^T\mathbf{U}\boldsymbol{\Delta}\mathbf{U}^T\mathbf{U}\boldsymbol{\Delta}^{-1/2} \\ &= \boldsymbol{\Delta}^{-1/2}\boldsymbol{\Delta}\boldsymbol{\Delta}^{-1/2} = \mathbf{I},\end{aligned} \tag{4.22}$$

as required.

To summarize, we take simplex vectors centered on the origin and transform them according to

$$\mathbf{w}_r \rightarrow \mathbf{D}\mathbf{Q}(\mathbf{w}_r - \mathbf{t}), \tag{4.23}$$

where $\mathbf{t}$, $\mathbf{Q}$, and $\mathbf{D}$ are chosen according to Eqs. (4.18) and (4.21). We use the transformed vectors to form the rows of the matrix $\mathbf{S}$, then the cut size for the partition of the graph indicated by $\mathbf{S}$ is given by

$$R = \tfrac{1}{2}\,\mathrm{Tr}(\mathbf{S}^T\mathbf{L}\mathbf{S}\mathbf{D}^{-2}), \tag{4.24}$$

while $\mathbf{S}$ obeys

$$\mathbf{S}^T\mathbf{1} = 0, \tag{4.25}$$

and

$$\mathbf{S}^T\mathbf{S} = \mathbf{I}. \tag{4.26}$$

In the two-group case, where $\mathbf{D}$ is a scalar (since $(k-1) = 1$) and $\mathbf{S}$ is an $n \times 1$ vector on the the corner of a hypercube, we are left with the same optimization problem as in the traditional spectral bisection.

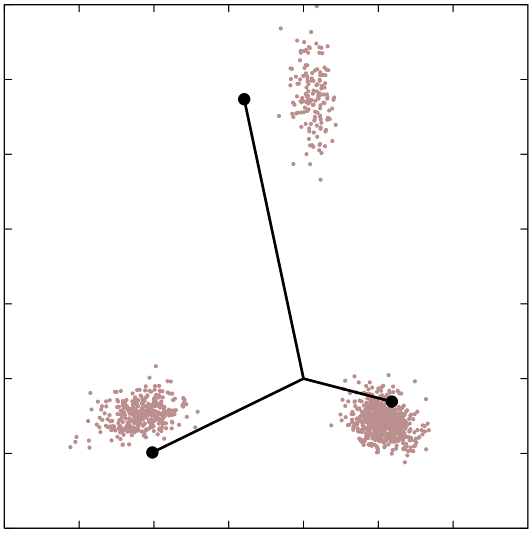


Figure 4.5: The eigenvectors of the Laplacian define n points in a $(k-1)$-dimensional space and we round each point to the nearest simplex vector to get an approximate solution to the partitioning problem. The particular graph used here was, for the purposes of illustration, deliberately created with three groups in it of varying sizes, using a simple planted partition model (Condon and Karp, 2001) in which edges are placed between vertices independently at random, but with higher probability between vertices in the same group than between those in different groups.

4.3.2 Minimization of the cut size

The remaining steps in the derivation are now straightforward, following lines closely analogous to those for the two-group case. Our goal is to minimize the cut size (4.24) subject to the condition that the group sizes take the desired value, which is equivalent to the constraint (4.25). Once again this is a hard computational problem, but, by analogy with the two-group case, we can render it tractable by relaxing the requirement that each row of $\mathbf{S}$ be equal to one of the discrete vectors (4.23), solving this relaxed problem exactly, then rounding to the nearest vector again to get an approximate solution to the original unrelaxed problem. The process is illustrated in Fig. 4.5.

We replace $\mathbf{S}$ with a matrix $\mathbf{X}$ of continuous-valued elements to give a relaxed cut size

$$R_x = \tfrac{1}{2}\,\mathrm{Tr}(\mathbf{X}^T\mathbf{L}\mathbf{X}\mathbf{D}^{-2}), \tag{4.27}$$

where the elements of $\mathbf{X}$ will be allowed to take any real values subject to the constraint

$$\mathbf{X}^T\mathbf{1} = 0, \tag{4.28}$$

equivalent to Eq. (4.25). Again, however, we also need an additional constraint, equivalent to Eq. (4.8), to prevent all elements of $\mathbf{X}$ from becoming zero (which would certainly minimize R_x, but would not give a useful partition of the graph), and the natural choice is the generalization of (4.26):

$$\mathbf{X}^T\mathbf{X} = \mathbf{I}. \tag{4.29}$$

As in the two-group case, choices of $\mathbf{X}$ satisfying this condition necessarily include as a subset the original group vectors that satisfy (4.26), but also include many other choices as well. Between them, the two conditions (4.28) and (4.29) imply that the columns of $\mathbf{X}$ should be orthogonal to one another, orthogonal to the vector $\mathbf{1}$, and normalized to have unit length. As we now show, the correct choice that satisfies all of these conditions and minimizes R_x is to make the columns proportional to the eigenvectors of the graph Laplacian corresponding to the second- to kth-lowest eigenvalues.

The relaxed cut size (4.27) can be minimized, as before, by differentiating, applying the conditions (4.28) and (4.29) with Lagrange multipliers:

$$\begin{split}\frac{\partial}{\partial X_{kl}}\biggl[\sum_{ijm} L_{ij}X_{im}X_{jm}D_{mm}^{-2}\\ -\sum_{imn}\lambda_{mn}X_{im}X_{in} - \sum_{im}\mu_m X_{im}\biggr] = 0,\end{split} \tag{4.30}$$

so that $2\sum_j L_{kj}X_{jl}D_{ll}^{-2} - 2\sum_m X_{km}\lambda_{ml} - \mu_l = 0$, or in matrix notation $\mathbf{LXD}^{-2} = \mathbf{X\Lambda} + \frac{1}{2}\mathbf{1}\boldsymbol{\mu}^T$, where $\boldsymbol{\Lambda}$ is a $(k-1)\times(k-1)$ symmetric matrix of Lagrange multipliers and $\boldsymbol{\mu}$ is a $(k-1)$-dimensional vector. As before, we can multiply on the left by $\mathbf{1}^T$ to show that $\boldsymbol{\mu} = 0$, and hence we find that

$$\mathbf{LX} = \mathbf{X\Lambda D}^2. \tag{4.31}$$

Now, making use of the fact that $\mathbf{X}^T\mathbf{X} = \mathbf{I}$ (Eq. (4.29)), we have

$$\mathbf{\Lambda D}^2 = \mathbf{X}^T\mathbf{X\Lambda D}^2 = \mathbf{X}^T\mathbf{LX} = \mathbf{D}^2\mathbf{\Lambda X}^T\mathbf{X} = \mathbf{D}^2\mathbf{\Lambda}. \tag{4.32}$$

In other words, $\mathbf{D}^2$ and $\mathbf{\Lambda}$ commute. Since $\mathbf{D}$ is diagonal this impiles that $\mathbf{\Lambda}$ is also diagonal, in which case Eq. (4.31) implies that each column of $\mathbf{X}$ is an eigenvector of the graph Laplacian with the diagonal elements of $\mathbf{\Lambda D}^2$ being the eigenvalues. In other words the eigenvalues are

$$\lambda_i = \Lambda_{ii}D_{ii}^2. \tag{4.33}$$

The conditions (4.28) and (4.29) tell us that the eigenvectors must be distinct (because they are orthogonal to each other), normalized to unity, and orthogonal to the vector $\mathbf{1}$.

This still leaves us considerable latitude about which eigenvectors we use. We can resolve the uncertainty by considering the cut size R_x, Eq. (4.27), which is given by

$$\begin{aligned} R_x &= \tfrac{1}{2}\operatorname{Tr}(\mathbf{X}^T\mathbf{LXD}^{-2}) = \tfrac{1}{2}\operatorname{Tr}(\mathbf{X}^T\mathbf{X\Lambda}) = \tfrac{1}{2}\operatorname{Tr}\mathbf{\Lambda} \\ &= \tfrac{1}{2}\sum_i \Lambda_{ii} = \tfrac{1}{2}\sum_i \frac{\lambda_i}{D_{ii}^2}. \end{aligned} \tag{4.34}$$

Our goal is to minimize this quantity and, since both D_{ii}^2 and the eigenvalues of the Laplacian λ_i are nonnegative, the minimum is achieved by choosing the smallest allowed eigenvalues of the Laplacian. We are forbidden by Eq. (4.28) from choosing the lowest (zero) eigenvalue, because its eigenvector is the vector $\mathbf{1}$, so our best allowed choice is to choose the columns of $\mathbf{X}$ to be the eigenvectors corresponding to the second- to kth-lowest eigenvalues of the Laplacian. Which column is which depends on the values of the D_{ii}. The minimum of R_x is achieved by pairing the largest λ_i with the largest D_{ii}, the second largest λ_i with the second largest D_{ii}, and so on.

This now specifies the value of the matrix $\mathbf{X}$ completely and hence consitutes a complete solution of the relaxed minimization problem. The correct choice of $\mathbf{X}$ is one in which the $k-1$ columns of the matrix are equal to the normalized eigenvectors corresponding to the second- to kth-lowest eigenvalues of the graph Laplacian, with the columns arranged so that their eigenvalues increase in the same order as the diagonal elements of the matrix $\mathbf{D}$.

The only remaining step in the algorithm is to reverse the relaxation, which means rounding the rows of the matrix $\mathbf{X}$ to the nearest of the group vectors—see Fig. 4.5. As in the two-group case, this introduces an approximation. Although our solution of the relaxed problem is exact, when we round to the nearest group vector there is no guarantee that the result will be a true minimum of the unrelaxed problem. Furthermore, as in the two-group case, we are not guaranteed that the groups found using this method will be of exactly the required sizes n_r. The relaxed optimization must satisfy Eq. (4.28), but the corresponding condition, Eq. (4.25), for the unrelaxed division of the graph is normally only satisfied approximately and hence the groups will only be approximately the correct size. As in the two-group case, however, this is typically not a problem. Often we are content with an approximate solution to the problem, but if not the groups can be balanced using a post-processing step. For example, the rounding of the relaxed solution to the group vectors that preserves precisely the desired group sizes can be calculated exactly in polynomial time using the so-called Hungarian algorithm (Papadimitriou and Steiglitz, 1998), or approximately using a variety of vertex moving heuristics.

4.3.3 Practical considerations

The method described in the previous section in principle constitutes a complete algorithm for the approximate spectral solution of the minimum-cut partitioning problem. In practice, however, there are some additional issues that arise in implementation.

First, note that the sign of the eigenvectors of the Laplacian is arbitrary, and hence our matrix $\mathbf{X}$ is only specified up to a change of sign of any column, meaning there are 2^{k-1} choices of the matrix that give equally good solutions to the relaxed optimization of the cut size. These 2^{k-1} solutions are reflections of one another in the axes of the space occupied by the group vectors, and in practice the quality of the solutions to the unrelaxed problem obtained by rounding each of these reflections to the nearest group vector varies somewhat. If we want the best possible solution we need to look through all 2^{k-1} possibilities to find which one is the best, and this could take a long time if k is large.

A second and more serious issue arises when the group sizes are equal to one another, or nearly equal. When the group sizes are equal the conditions $\mathbf{S}^T\mathbf{1} = 0$ and $\mathbf{S}^T\mathbf{S} = \mathbf{I}$ are satisfied by the original, symmetric simplex vectors of Fig. 4.3 in any orientation. This means that the group vectors are not fully specified in this case—their orientation is arbitrary. When rounding the rows of the matrix $\mathbf{X}$ to the nearest simplex vector, therefore, an additional rotation may be required to find the best solution.

The situation is depicted in Fig. 4.6 for the case $k = 3$. The rows $\mathbf{x}_i$ are two-dimensional vectors in this case and form a scatter of points in the plane of the plot as shown. The points do indeed approximate reasonably well to the corners of a

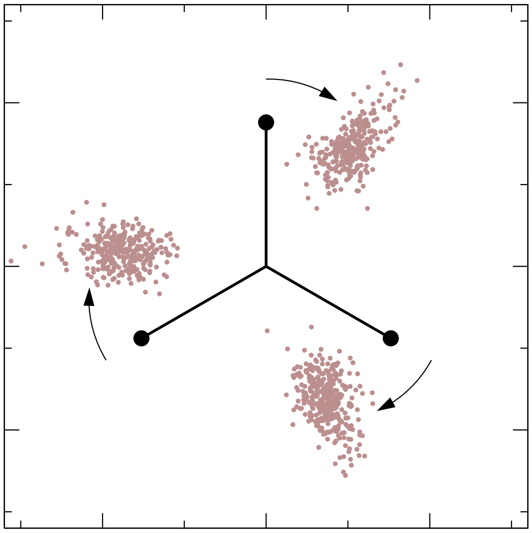


Figure 4.6: The points in this plot represent the elements of the second and third eigenvectors of the Laplacian for a small graph of about a thousand vertices. The graph used was, like that of Fig. 4.5, created using a planted partition model, but with equally sized groups in this case. The resulting points fall, roughly speaking, at the corners of a two-dimensional regular simplex, i.e., an equilateral triangle. To determine the division of the graph into groups, we need to round these points to the nearest simplex vector, but the simplex must first be rotated to match the orientation of the points.

regular simplex (an equilateral triangle in this case), so in principle we should be able to round them off and get a good solution to the partitioning problem. But we do not know *a priori* what the correct orientation of the simplex is, and in this case our first guess, as shown in the figure, is off and a rotation is required. We can rotate either the points or the simplex, but we recommend rotating the simplex because it requires less work.

Given an assignment of vertices to groups, we can write down the matrix $\mathbf{S}$ of (unrotated) simplex vectors. If we rotate the vectors, this matrix becomes $\mathbf{SR}$, where $\mathbf{R}$ is a $(k-1)\times(k-1)$ orthogonal matrix. The sum of the squares of the Euclidean distances from each point to the corresponding simplex vector is given by

$$\begin{aligned}\sum_{ij}[\mathbf{SR}-\mathbf{X}]_{ij}^2 &= \operatorname{Tr}\bigl[(\mathbf{R}^T\mathbf{S}^T-\mathbf{X}^T)(\mathbf{SR}-\mathbf{X})\bigr]\\ &= \operatorname{Tr}\mathbf{S}^T\mathbf{S} - 2\operatorname{Tr}\mathbf{R}^T\mathbf{S}^T\mathbf{X} + \operatorname{Tr}\mathbf{X}^T\mathbf{X}.\end{aligned} \tag{4.35}$$

The first and last terms in this expression are independent of $\mathbf{R}$ and hence, for the purposes of choosing the rotation $\mathbf{R}$ that minimizes the whole expression, we need only minimize the middle term, or equivalently maximize $\operatorname{Tr}\mathbf{R}^T\mathbf{S}^T\mathbf{X}$. The maximization of this quantity over orthogonal matrices $\mathbf{R}$ is a standard problem in so-called Procrustes analysis (Gower and Dijksterhuis, 2004), which seeks to compare the shapes of objects by optimally superimposing them (often minimizing some distance metric) under some set of allowed transformations. It can be solved by performing a singular value decomposition of the matrix $\mathbf{S}^T\mathbf{X}$ (Schönemann, 1966):

$$\mathbf{S}^T\mathbf{X} = \mathbf{U\Sigma V}^T, \tag{4.36}$$

where $\mathbf{\Sigma}$ is the diagonal matrix of singular values and $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices. Then

$$\begin{aligned}\mathrm{Tr}(\mathbf{R}^T\mathbf{S}^T\mathbf{X}) &= \mathrm{Tr}(\mathbf{R}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T) = \mathrm{Tr}(\mathbf{V}^T\mathbf{R}^T\mathbf{U}\mathbf{\Sigma}) \\ &\le \mathrm{Tr}\,\mathbf{\Sigma}, \end{aligned} \tag{4.37}$$

the inequality following because $\mathbf{V}^T\mathbf{R}^T\mathbf{U}$, being a product of orthogonal matrices, is also itself orthogonal, and all elements of an orthogonal matrix are less than or equal to 1. It is now trivial to see that the exact equality—which is, by definition, the maximum of $\mathrm{Tr}(\mathbf{R}^T\mathbf{S}^T\mathbf{X})$ with respect to $\mathbf{R}$—is achieved when $\mathbf{R}^T = \mathbf{V}\mathbf{U}^T$ or equivalently when

$$\mathbf{R} = \mathbf{U}\mathbf{V}^T. \tag{4.38}$$

The product $\mathbf{U}\mathbf{V}^T$ is the orthogonal part of the *polar decomposition* of $\mathbf{S}^T\mathbf{X}$. Calculating it in practice involves calculating first the singular value decomposition, Eq. (4.36), and then discarding the diagonal matrix $\mathbf{\Sigma}$. Note that $\mathbf{S}^T\mathbf{X}$ is only a $(k-1)\times(k-1)$ matrix (not an $n \times n$ matrix), and hence its singular value decomposition can be calculated rapidly provided k is small, in $\mathrm{O}(k^3)$ time.

These developments assume that we know the assignment of the vertices to the groups. In practice, however, we don't. (If we did, we wouldn't need to partition the graph in the first place.) So in the algorithm we propose we start with a random guess at the orientation of the simplex. We round the rows of $\mathbf{X}$ to the simplex vectors to determine group memberships and then rotate the simplex vectors to fit the resulting groups according to Eq. (4.38). We repeat this procedure until the groups no longer change. In a clear-cut case like that of Fig. 4.6, only one or two iterations would be needed for convergence, but in more ambiguous cases we have found that as many as half a dozen or more iterations may be necessary. Though we have not extensively explored the conditions leading to problems with convergence, our preliminary results suggest that convergence tends to be slower when cutting a network into many groups or when the best available division is not very good (e.g. cutting a network with a natural four group division into five groups).

This algorithm works well for the case of exactly equal group sizes, while the algorithm described in Section 4.3.2 works well for very unbalanced groups, when the group vectors are specified completely, without need for any rotation or Procrustes analysis. A trickier scenario is when the groups are almost but not exactly equal in size. In such cases the algorithm of Section 4.3.2 is correct in principle, but in practice tends not to work very well—the particular orientation of the group vectors picked by the algorithm may not agree well with the scatter of points described by the rows of the matrix $\mathbf{X}$. In such cases, we find that an additional Procrustes step to line up the points with the group vectors usually helps.

But this raises the question of when the groups can be considered sufficiently balanced in size that a possible rotation of the group vectors may be needed. Rather than try to answer this difficult question, we recommend simply performing a Procrustes analysis and rotation for all partitioning problems, whether one is needed or not. In practice it does not take long to do, and if it is not needed—if the points described by the elements of $\mathbf{X}$ are already well lined up with the group vectors, as they are in Fig. 4.5 for instance—then the Procrustes analysis will simply do nothing. It will leave the group vectors unrotated (or rotate them only very slightly).

This approach has the added advantage of offering a solution to our other problem as well, the problem of undetermined signs in the eigenvectors of the Laplacian. Since the orthogonal matrix $\mathbf{R}$ in the Procrustes analysis can embody a reflection as well as a rotation, the Procrustes analysis will also determine which reflection gives the best fit of the group vectors to the points, so we do not require an additional step to deal with reflections.

Since the Procrustes analysis is an iterative method we do, in practice, find that it can converge to the wrong minimum of the mean-square distance. In the calculations presented in the remained of this paper, we run the analysis several times with randomized starting conditions, taking the best result over all runs, in order to mitigate this problem.

4.3.4 Summary of the algorithm and running time

Although the derivation of the previous sections is moderately lengthy, the final algorithm is straightforward. In summary the algorithm for partitioning a given graph into k groups of specified sizes is as follows.

1. Generate a set of vectors k pointing to the vertices of a regular simplex centered at the origin and assign one vector as the label for each of the k groups. Any orientation of the simplex can be used at this stage and any assignment of vectors to groups.

2. Define $\mathbf{t}$, $\mathbf{Q}$, and $\mathbf{D}$ according to Eqs. (4.18) to (4.21), then transform the simplex vectors according to

$$\mathbf{w}_r \to \mathbf{D}^{-1}\mathbf{Q}(\mathbf{w}_r - \mathbf{t}). \tag{4.39}$$

3. Find the second- to kth-smallest eigenvalues of the graph Laplacian, and the corresponding normalized eigenvectors. Pair the largest of these eigenvalues with the largest diagonal element of $\mathbf{D}$, the second largest eigenvalue with the second largest diagonal element, and so forth. Then form the matrix $\mathbf{X}$, whose columns are the eigenvectors arranged in the same order as the diagonal elements of $\mathbf{D}$ with which they are paired.

4. Rotate the group vectors $\mathbf{w}_r$ into a random initial orientation.

5. Round each of the rows of $\mathbf{X}$ to the nearest group vector and construct the corresponding group matrix $\mathbf{S}$ whose ith row is the group vector for the group

that vertex i now belongs to.

6. Form the singular value decomposition $\mathbf{S}^T\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ and from it calculate the rotation matrix $\mathbf{R} = \mathbf{U}\mathbf{V}^T$.
7. Rotate the group vectors $\mathbf{w}_r \to \mathbf{w}_r\mathbf{R}$.
8. Repeat from step 5 until group memberships no longer change.

Most often we are interested in sparse graphs in which the number of edges is proportional to the number of vertices, so that the mean degree of a vertex tends to a constant as the graph becomes large. In this situation the eigenvectors of the Laplacian can be calculated using sparse iterative methods such as the Lanczos algorithm. The Lanczos algorithm takes time $\mathrm{O}(k^2n)$ per iteration, and although there are no formal results for the number of iterations required for convergence, the number in practice seems to be small. The other steps of the algorithm all also take time $\mathrm{O}(k^2n)$ or less, and hence the algorithm has leading-order worst-case running time $\mathrm{O}(k^2n)$ times the number of Lanczos iterations, making it about as good as traditional approaches based on k-means, and well suited for large graphs. (Formal results for the number of iterations k-means takes to converge are also not available, so a precise comparison of the complexity of the two methods is not possible.)

4.3.5 Weighted graphs and data clustering

The methods described in the previous sections can be extended in a straightforward manner to weighted graphs—graphs with edges of varying strength represented by varying elements in the adjacency matrix. For such graphs the goal of partitioning is to divide the vertices into groups such that the sum of the weights of the edges running between groups is minimized. To achieve this we generalize the degree d_i of vertex i in the obvious fashion $d_i = \sum_j A_{ij}$ and the elements of the Laplacian accordingly $L_{ij} = d_i\delta_{ij} - A_{ij}$. Then the cut size once again satisfies Eq. (4.15), and the rest of the algorithm follows as before. We have not experimented extensively with applications to weighted graphs, but in preliminary tests the results look promising.

One can also apply our methods to the problem of data clustering, the grouping of points within a multidimensional data space into clusters of similar values (Verma and Meilă, 2003; Von Luxburg, 2007). One standard approach to this problem makes use of an affinity matrix. Suppose one has a set of n points represented by vectors $\mathbf{r}_i$ in a d-dimensional data space. One then defines the affinity matrix $\mathbf{A}$ to have elements

$$A_{ij} = \mathrm{e}^{-|\mathbf{r}_i - \mathbf{r}_j|^2/2\sigma^2}, \tag{4.40}$$

where σ is a free parameter chosen by the user. If σ is roughly of order the distance between intra-cluster points, then A_{ij} will approximately take the form of the adjacency matrix of a weighted graph in which vertices are connected by strong edges if the corresponding data points are near neighbors in the data space. (For values of σ much larger or smaller than this clustering methods based on the affinity matrix will not work well, so some care in choosing σ is necessary to get good results. Automated methods have been proposed for choosing a good value (Ng et al., 2001).)

Given the affinity matrix, we can now apply the method described above for weighted graphs to this matrix and derive a clustering of the data points. We will not pursue this idea further in the present paper, but in preliminary experiments on standard benchmark data sets we have found that the algorithm gives results comparable with, and in some cases better than, other simple spectral clustering methods. However, data clustering applications present additional difficulties because they need not have any predefined target for the number of groups or the sizes of groups. While our algorithm often performs well under the assumption of equally sized groups, even when the actual groups are quite unbalanced, choosing the *number* of groups efficiently for large data sets remains a challenge.

4.4 Results

Our primary purpose in this paper is to provide a first-principles derivation of a multiway spectral partitioning algorithm. However, given that the algorithm we have derived differs from standard algorithms, it is also of interest to examine how well it performs in practice. In this section we give example applications of the algorithm to graphs from a variety of sources. Our tests do not amount to an exhaustive characterization of performance, but they give a good idea of the basic behavior of the algorithm. Overall, we find that the algorithm has performance comparable to that of the k-means algorithm on Laplacian eigenvectors, but there exist classes of graphs for which our algorithm does measurably better. In particular, the algorithm appears to perform better than some competitors in cases where the partitioning task is particularly difficult.

While there are many sophisticated methods of k-way partitioning, including some spectral methods (see Introduction), we use the k-means algorithm as a point of comparison because it is a simple, commonly used method that is in many ways similar to our algorithm. Both algorithms partition vertices according to the lowest-lying Laplacian eigenvectors and both use alternating steps of assigning vertices to the nearest "group center" and then adjusting the positions of the centers based on the new assignment. However, while the group centers in the k-means algorithm are unconstrained, our algorithm restricts their shape up to orthogonal transformations. The simplex extension of spectral bisection's "spins" is at the heart of our derivation of our algorithm, so it is a useful comparison to consider k-means as a simplex-less analogue.

As a first example, Fig. 4.7 shows the result of applying our algorithm to a graph from the University of Florida Sparse Matrix Collection. This graph is a two-dimensional mesh network drawn from a NASA structural engineering computation, and is typical of finite-element meshes used in such calculations (which are a primary

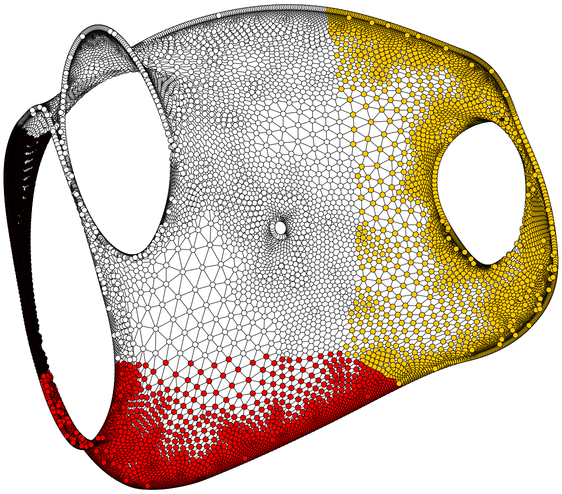


Figure 4.7: Division of a structural engineering mesh network of 15 606 vertices into four parts—represented by the four colors—using the algorithm described in this paper. The sizes of the parts in this case were 1548, 2745, 4979, and 6334. The complete graph has 45 878 edges; this division cuts just 351 of them, less than 1%. Graph data courtesy of the University of Florida Sparse Matrix Collection.

application of partitioning methods). Figure 4.7 shows a split of the graph into four parts of widely varying sizes. The split is closely similar to that found by conventional spectral partitioning using k-means.

Figure 4.8 shows an application to a graph representing a power grid, specifically the Western States Power Grid, which is the network of high-voltage electricity transmission lines that serves the western part of the United States (Watts and Strogatz, 1998). The figure shows the result of splitting the graph into four parts and the split is an intuitively sensible one and again comparable to that found using more traditional methods.

There are, however, some graphs for which our method gives results that are significantly better than those given by previous methods, particularly when the target group sizes are significantly unbalanced. As a controlled test of the performance of the algorithm we have applied it to artificial graphs generated using a planted partition model (Condon and Karp, 2001) (also called a stochastic block model in the

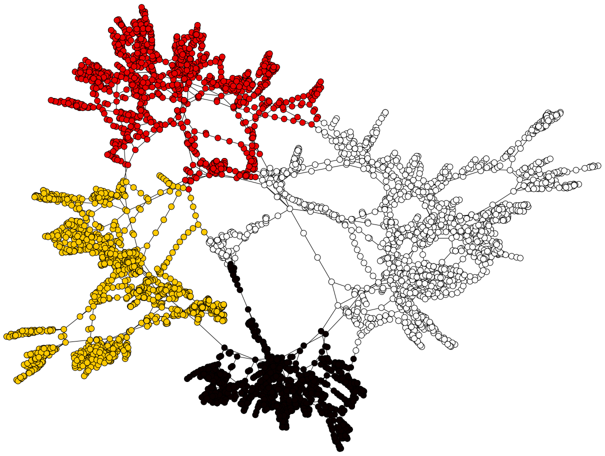


Figure 4.8: Division into four parts of a 4941-vertex graph representing the Western States Power Grid of the United States. The sizes of the parts were 898, 1066, 1240, and 1737. The complete graph contains 6594 edges, of which 25 are cut in this division. Graph data courtesy of Duncan Watts.

statistical literature (Snijders and Nowicki, 1997)). In this model one creates graphs with known partitions by dividing a specified number of vertices into groups and then placing edges within and between those groups independently with given probabilities. In our tests we generated graphs of 3600 vertices with three groups. Edges were placed between vertices with two different probabilities, one for vertices in the same group and one for vertices in different groups, chosen so that the average degree of a vertex remained constant at 40. We then varied the fraction of edges placed within groups to test the performance of the algorithm.

Figure 4.9 shows the results of applying both our algorithm and a standard k-means spectral algorithm to a large set of graphs generated using this method. We compare the divisions found by each algorithm to the known correct divisions and calculate the fraction of vertices classified into the correct groups as a function of the fraction of in-group edges. When the latter fraction is large the group structure in the network should be clear and we expect any partitioning algorithm to do a good job of finding the best cut. As the fraction of in-group edges is lowered, however, the task gets harder and the fraction of correct vertices declines for both algorithms,

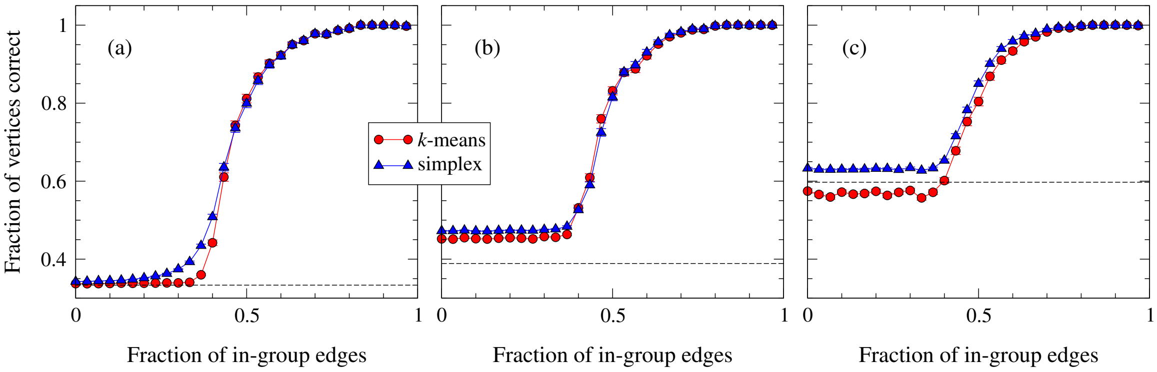


Figure 4.9: Fraction of vertices classified into the correct groups by a standard spectral algorithm based on k-means (red circles) and by the algorithm described in this paper (blue triangles), when applied to graphs of 3600 vertices, artificially generated using a planted partition model with three groups. In (a) the groups are of equal sizes. In (b) the sizes are 1800, 1200, and 600. In (c) they are 2400, 900, and 300. The dashed horizontal line in each frame represents the point at which the algorithms do no better than chance. Each data point is an average over 500 networks and the calculation for each network is repeated with random initial conditions as described in the text; the results shown here are the best out of ten such repeats.

eventually approaching the value represented by the dashed horizontal lines in the figure, which is the point at which the classification is no better than chance—we would expect a random division of vertices to get about this many vertices right just by luck. (If group i occupies a fraction ν_i of the network, then a random division into groups of the given sizes will on average get $\sum_i \nu_i^2$ vertices correct.)

The first panel in the figure shows results for groups of equal size and for this case the performance of the two algorithms is similar. Both do little better than random for low values of the fraction of in-group edges. The simplex algorithm of this paper performs slightly better in the hard regime, but the difference is small. When the group sizes are different, however, our algorithm outperforms the k-means algorithm, as shown in the second and third panels. In the third panel in particular, where the group sizes are strongly unbalanced, our algorithm performs substantially better than k-means for all parameter values, but particularly in the hard regime where the fraction of in-group edges is small. In this regime the k-means algorithm does no better than a random guess, but our simplex-based algorithm does significantly better.

To be fair, we should also point out that there are some cases in which the k-means algorithm outperforms the algorithm of this paper. In particular, we find that in tests using the planted partition model with three groups of equal sizes, but where the between-group connections are asymmetric and one pair of groups is more weakly connected than the other two pairs, the k-means algorithm does better in certain parameter regimes. The explanation for this phenomenon appears to be that our algorithm has difficulty finding the best orientation of the simplex to perform the partitioning. It is possible that one could achieve better results using a different method for finding the orientation other than the Procrustes method used here. The k-means partitioning algorithm, which does not use an orientation step, has no corresponding issues.

4.5 Conclusions

In this paper, we have derived a multiway spectral partitioning algorithm from first principles as a relaxation approximation to a well-defined minimum-cut problem. This contrasts with more traditional presentations in which an algorithm is proposed *ex nihilo* and then proved after the fact to give good results. While both approaches have merit, ours offers an alternative viewpoint that helps explain why spectral algorithms work—because the spectral algorithm is, in a specific sense, an approximation to the problem of minimizing the cut size over divisions of the graph.

Our approach not only offers a new derivation, however; the end product, the algorithm itself, is also different from previous algorithms, involving a vector representation of the partition with the geometry of an irregular simplex. In practice, the algorithm appears to give results that are comparable with those of similar algorithms and in some cases better. The algorithm is also comparably efficient to similar, simple spectral partitioning algorithms such as k-means. For graphs of n vertices divided into k groups, the running time is dictated by the calculation of the eigenvectors of the graph Laplacian matrix, which for a sparse graph can be done using the Lanczos method in time $\mathrm{O}(k^2 n)$ times the number of Lanczos iterations (which is typically small), so overall running time is roughly linear in n for given k.

The developments described here leave some questions unanswered. In particular, our method fixes the group sizes within the relaxed approximation to the minimization problem, but in the true problem the sizes are only fixed approximately. A common variant of the minimum-cut problem arises when the group sizes are not exactly equal but are allowed to vary within certain limits. Our method could be used to tackle this problem as well, but one would need a scheme for preventing the size variation from passing outside the allowed bounds. These and related ideas we leave for future work.

CHAPTER V

Local variation in plant quality influences large-scale population dynamics

5.1 Introduction

Spatial variation in ecological systems can arise both as a consequence of underlying variation in the quality and availability of resources (Denno and McClure, 1983; Pulliam, 1988) and as an emergent property of spatially structured trophic interactions (Hassell, 2000). Although there have been many exceptions (Roughgarden, 1974; Cantrell and Cosner, 1991; Oksanen et al., 1992; Bjørnstad and Hansen, 1994; Helms and Hunter, 2005; Holt and Barfield, 2003; Underwood, 2004), mathematical models of spatial heterogeneity have largely focused on spatial and temporal variation in organism abundance caused by top-down mechanisms rather than resource heterogeneity. For example, effects of spatially structured host-parasitoid interactions on spatial and temporal host dynamics have been studied for some years (Hassell et al., 1991; Comins et al., 1992). Such interactions can stabilize otherwise unstable population systems and generate spatial patterns that include spirals and traveling waves (Bjørnstad et al., 2002). At the same time, migration can also destabilize dynamics which would, in isolated populations, lead to stable persistence (Reeve, 1988).

In models of this type, spatial variation in abundance emerges because global populations are not well-mixed; there is always some degree of asynchrony between growth and decline in local but weakly coupled populations. These top-down mechanisms can induce spatial heterogeneity even among populations living in a network of identical habitat patches.

However, in the real world, habitat patches are rarely identical. Predator-prey interactions are superimposed upon landscapes that vary dramatically in their quality for herbivores (Hunter and Price, 1992). For example, insect herbivore populations are affected by very local, bottom-up forces such as the chemistry (Hunter et al., 1996) and genotype (Underwood and Rausher, 2000; McIntyre and Whitham, 2003; Evans et al., 2012*b*) of their host plants. At larger spatial scales, the local plant community (Pimentel, 1961; Murdoch et al., 1972; Andow, 1991) and surrounding landscape (Cappuccino et al., 1998) affect both the diversity and abundance of herbivores and of their natural enemies (Root, 1973).

Given that both trophic interactions and underlying resource heterogeneity are important drivers of spatial and temporal dynamics, there is growing interest in understanding how they interact to influence the dynamics of species (Hunter et al., 1997). The impact of resource heterogeneity can depend on the scale of variation in resource quality (Oksanen et al., 1992; Roland and Taylor, 1997; Thies et al., 2003), and may affect population dynamics at different spatial scales (Murdoch et al., 1972). Further, the effects of resource heterogeneity have been found to interact with the top-down forces acting on populations both in field (Batch, 1984) and modeling studies (Foster et al., 1992).

While linking studies of spatial variation in resource quality and trophic interactions is of significant theoretical interest, such studies may also inform management, particularly in agricultural ecosystems where pest organisms attack crops that are planted in well defined spatial arrays (Andow, 1991). Managing the species and genotypes of crops and their arrangements in space provide opportunities to minimize pest attack and maximize biological control by the natural enemies of pests (Belyea, 1923; Pimentel, 1961; Root, 1973). Intercropping, planting multiple crops together in fields, has shown some success as a method of enhancing top-down control of insect herbivores by their natural enemies (Bickerton, 2011; Chen et al., 2011). At larger spatial scales, increased predation and parasitism of insect herbivores in agricultural systems have also been associated with landscape features such as close proximity to uncultivated land, higher proportions of non-crop area in the surrounding region, and more diversity in nearby habitat patches (Thies and Tscharntke, 1999; Cronin and Reeve, 2005).

In field studies, the effects of variation in plant quality on herbivore populations depends not only on mean plant quality but on the distribution of plant qualities (Underwood, 2004; Helms and Hunter, 2005; Underwood, 2009) as well as the spatial arrangement of high and low quality plants (Evans et al., 2012*b*; Thies et al., 2003). Consequently, we explored the effects of variability around a constant mean plant quality on the outcome of spatially structured host-parasitoid interactions.

We modified an existing spatially-explicit model of host-parasitoid dynamics (Rohani and Miramontes, 1995) to simulate host and parasitoid populations on landscapes with different variance and spatial layout of plant quality. We aimed to determine the extent to which a combination of local variation in resource quality and simple parasite-host interactions can give rise to the kind of complex spatial and temporal dynamics observed in field studies.

Specifically, we tested the following two hypotheses:

1. Mean patch hypothesis — Large scale population dynamics in landscapes of heterogeneous plant quality differ from those in homogeneous landscapes with the same mean plant quality.
2. Patch architecture hypothesis — Effects of variation in plant quality on dynamics are contingent on the arrangement of high and low quality plants in

space.

5.2 Mathematical model and simulation methods

Our model of host-parasitoid dynamics is similar to previous models (Hassell et al., 1991; Rohani and Miramontes, 1995) with the addition of spatial heterogeneity in plant quality. In this framework, herbivore hosts (hereafter, "hosts") and parasitoids with synchronized generations grow and disperse in alternating steps according to a coupled map lattice. Each position on the lattice represents an individual plant of known quality for hosts. During the growth phase, populations on each plant grow according to a simple Nicholson-Bailey model (Nicholson and Bailey, 1935). In the absence of parasitoids, the host population grows at a fixed rate λ. Throughout our study, we will use λ as a proxy for plant quality. Because homogeneous landscapes where host fecundity $\lambda = 2.0$ have been studied previously (Hassell et al., 1991; Rohani and Miramontes, 1995), we will use $\overline{\lambda} = 2.0$ as the mean plant quality in our heterogeneous fields.

For simplicity, we assume each host attacked results in an average of one adult parasitoid in the next generation. Thus, given the densities of hosts (H_t) and parasitoids (P_t) during generation t, the population densities after reproduction are given by

$$H'_t = \lambda H_t f(H_t, P_t) \tag{5.1}$$

$$P'_t = H_t(1 - f(H_t, P_t)) \tag{5.2}$$

where $f(H_t, P_t)$ is the fraction of hosts escaping parasitoid attack given the current densities of hosts and parasitoids.

Our model uses a linear parasitoid functional response with attacks distributed at random among hosts. Assuming that populations within plants are well–mixed and parasitoids search over some area a for hosts to oviposit in, we expect a total of aH_tP_t encounters between hosts and parasitoids. With parasitoid attacks distributed at random among hosts, the number of attacks on each host follows a Poisson distribution with mean aP_t attacks per host. Thus, the expected fraction of hosts to escape parasitism is $f(H_t, P_t) = e^{-aP_t}$. Because a does not qualitatively affect the dynamics of the model, but rather acts as a scaling parameter (Hassell et al., 1991), we arbitrarily fix a at 0.2 throughout our simulations. Given host and parasitoid populations H_t and P_t during generation t, their populations after reproduction will thus be given by

$$H'_t = \lambda H_t e^{-aP_t} \tag{5.3}$$

$$P'_t = H_t(1 - e^{-aP_t}) \tag{5.4}$$

On a single plant or a small lattice of plants coupled by dispersal, these growth dynamics result in rapidly growing oscillations of both host and parasitoid populations (Nicholson and Bailey, 1935). However, if the size of the total arena of coupled sites (hereafter, “landscape”) is large compared to the distance hosts and parasitoids are able to disperse, long–term coexistence can occur (Comins et al., 1992; Hassell et al., 1991). Our simulations all take place on 60 plant by 60 plant square lattices (landscapes) with absorbing boundaries, which was a sufficiently large arena that neither hosts nor parasitoids experienced global extinctions in any of our runs.

Hosts and parasitoids disperse to nearby plants with probabilities m_H and m_P respectively. Previous studies have found that the spatial patterns of host and parasitoid populations depend on λ, m_H, m_P, and the size of the landscape (Comins et al., 1992; Hassell et al., 1991). Our preliminary sensitivity analyses found that the dispersal traits of herbivores and parasitoids strongly modulated the effects of spatial variation in resource quality. To cope with the nonlinear, interacting effects of these parameters on the interplay between plant heterogeneity and host-parasitoid interactions in the absence of an existing body of theory, we have chosen to focus our current study on a single, biologically plausible region of parameter space, where parasitoids disperse more frequently than their hosts (Briggs and Latto, 2000; Taylor, 1991). Our main body of results fixes m_H and m_P at 0.2 and 0.8, respectively, with other values explored in our sensitivity analyses.

We set the dispersal range of both hosts and parasitoids to be their home plant and the eight surrounding plants in the landscape, which has been the dispersal range typically used in studies of hosts and parasitoids in homogeneous landscapes (Hassell et al., 1991; Comins et al., 1992; Rohani and Miramontes, 1995). We denote this neighborhood of plant i by $N(i)$.

We model hosts as dispersing uniformly among plants in range, so that after dispersal the host density at plant i becomes

$$H_{t+1}(i) = (1 - m_H)H'_t(i) + m_H\overline{H}(i) \tag{5.5}$$

where $\overline{H}(i)$ is the mean density of hosts in the neighborhood of plant i. Instead of diffusing uniformly like hosts, we allow parasitoids to aggregate to plants with a higher density of hosts as modeled in Hassell and May (1973), so that the density of parasitoids on plant i after dispersal is

$$P_{t+1}(i) = (1 - m_P)P'_t(i) + m_P\Sigma_{i\in N(j)}\beta(j,i)P'_t(j) \tag{5.6}$$

The fraction of dispersing parasitoids from plant j that land in plant i, $\beta(j,i)$, is determined by nearby host populations according to

$$\beta(j,i) = \nu\left(\frac{H'_t(i)}{\Sigma_{k\in N(j)}H'_t(k)}\right)^{\mu} \tag{5.7}$$

where ν is a normalization constant so that $\Sigma_{i \in N(j)} \beta(j,i) = 1$. The parameter μ controls the strength of the aggregation, with $\mu = 0$ resulting in uniform diffusion like the host population, $\mu = 1$ resulting in parasitoid dispersal proportional to the relative host populations of nearby plants, and parasitoids increasingly ignoring all but the most host rich plant in range as μ approaches infinity. In our main body of simulations, μ is fixed at 1.0, so as to qualitatively approximate the preferential dispersal of parasitoids to nearby plants with high host density as observed in the field (Fischbein et al., 2012). Our sensitivity analyses included values of μ from 0.0 (pure diffusion) to 2.0 (strong aggregation).

Within this framework, we can model spatial variation in host plant quality by varying the intrinsic growth rate of the local host population λ on each plant. For example, to construct an environment with source–sink dynamics, we could give plants in our desired source patches $\lambda > 1$ and plants in our sink patches $\lambda \leq 1$. Hirzel *et al.* (2007) explored an extreme case of a similar host-parasitoid model in which patches were either habitable (with fixed $\lambda > 1$) or completely hostile ($\lambda = 0$) (Hirzel et al., 2007).

For simplicity, we will consider landscapes where half of the plants are high quality, with host fecundity λ_+, and half are low quality, with host fecundity λ_-. This allows us to vary the difference in plant quality $(\lambda_+ - \lambda_-)$, while keeping the mean plant quality $\overline{\lambda} = \frac{1}{2}(\lambda_+ + \lambda_-)$ fixed. Even restricting our choices to the set of landscapes with equal numbers of high and low quality plants, we are still left with over 10^{1080} possible arrangements of plants.In our current study, we consider two arrangements of high and low quality plants: a landscape where one half contains all high fecundity plants and the other contains all low fecundity plants (hereafter, the "half-and-half landscape"), and a landscape of high and low quality plants alternating in a checkerboard pattern (hereafter, the "checkerboard landscape"). Figure 5.1 shows a schematic of these landscape types.

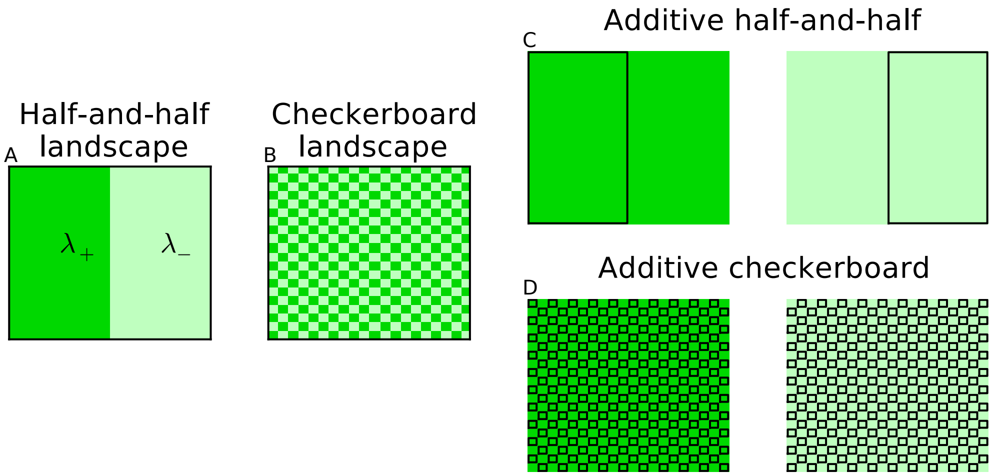


Figure 5.1: Plants of high (dark green) and low (light green) quality arranged in (A) a "half-and-half" landscape containing one large patch of high fecundity plants and one large patch of low quality plants, (B) a "checkerboard" landscape of alternating high and low quality plants, (C) an "additive" half-and-half landscape, and (D) an "additive" checkerboard landscape, using plants from two homogeneous landscapes (outlined by black rectangles).

These two arrangements both approximate plausible agricultural landscapes, where two crop species or genotypes are intercropped at a coarse (e.g. whole field) or fine (e.g. within-field) spatial scale. They also represent two extremes of spatial correlation in patch quality. In the half-and-half landscape, the vast majority of plants share the same quality as all of their neighbors, with the only exceptions occurring at the border between the high and low fecundity sides. At the opposite extreme, each plant in the checkerboard matches only half of its neighbors in host fecundity.

In order to test the extent to which a heterogeneous landscape's dynamics could be captured by simply aggregating the behavior of high and low quality plants, we constructed an "additive" base case to compare with each heterogenous landscape. Specifically, we independently simulated population dynamics on one homogeneous, high fecundity landscape and one homogeneous, low fecundity landscape. For the purposes of calculating population statistics, we then treated the left half of the high fecundity landscape and the right half of the low fecundity landscape as a single run, giving us a time series of host and parasitoid populations in a field with the appropriate distribution of plant qualities but none of the connections between high and low quality plants (see Figure 5.1C). We constructed checkerboard patterned "additive landscapes" in an analogous fashion, by first creating homogenous landscapes of high or low quality plants, and then sampling every second plant from each of those landscapes to give the same average plant quality (see Figure 5.1D).

We fixed mean quality $\overline{\lambda}$ at 2.0, and varied the difference in qualities so that $\lambda_- = 1.0, 1.05, 1.1, \ldots, 1.95, 2.0$ (homogeneous) for a total of 21 different levels of variation in plant quality, but all with the same global average. For each case, we constructed heterogeneous landscapes and associated pairs of high and low quality homogeneous landscapes. We ran 20 replicate simulations of each, with randomly generated initial populations, and simulated populations for 1000 generations. The initial populations on each plant were independently drawn from uniform distributions from 50% to 150%

of what would be (unstable) equilibrium populations of the Nicholson–Bailey equation on that plant alone. All temporal means and variances were calculated over the last 200 generations of the run, which we found to be much longer than the dominant periods at which populations in homogeneous landscapes oscillated. All populations in heterogeneous landscapes persisted until the end and global extinctions of hosts or parasitoids occurred only in homogeneous fields with $\lambda = 1$.

5.3 Results

Figure 5.2 shows snapshots of host and parasitoid density in homogeneous landscapes of high and low quality plants (panels A & C), a half-and-half landscape (panel E), and a checkerboard landscape (panel H). The spatial patterns of host and parasitoid populations in the high and low quality regions of the half-and-half landscape were generally similar to those in homogeneous fields of the same quality. However, the spatial patterns in the checkerboard landscape are organized over a larger spatial scale than the variation in plant quality, so that one traveling wave of hosts and parasitoids would pass through many plants of both high and low quality. This difference in spatial population dynamics between half-and-half and checkerboard landscapes can also be observed in the way local populations oscillate over time. While the populations on a high quality plant in the half-and-half landscape (panel D) oscillate with a higher frequency and amplitude than those on a low quality plant in the same landscape (panel F), the oscillations of populations on high and low quality plants at the same locations in a checkerboard landscape (panels G & H) are more similar in frequency.

In additive landscapes, where high and low quality plants were not connected by any dispersal, both host and parasitoid populations increased with increasing variance in plant quality. However, populations in half-and-half and checkerboard landscapes differed both from additive landscapes and from each other (Figure 5.3 A & B). Half-

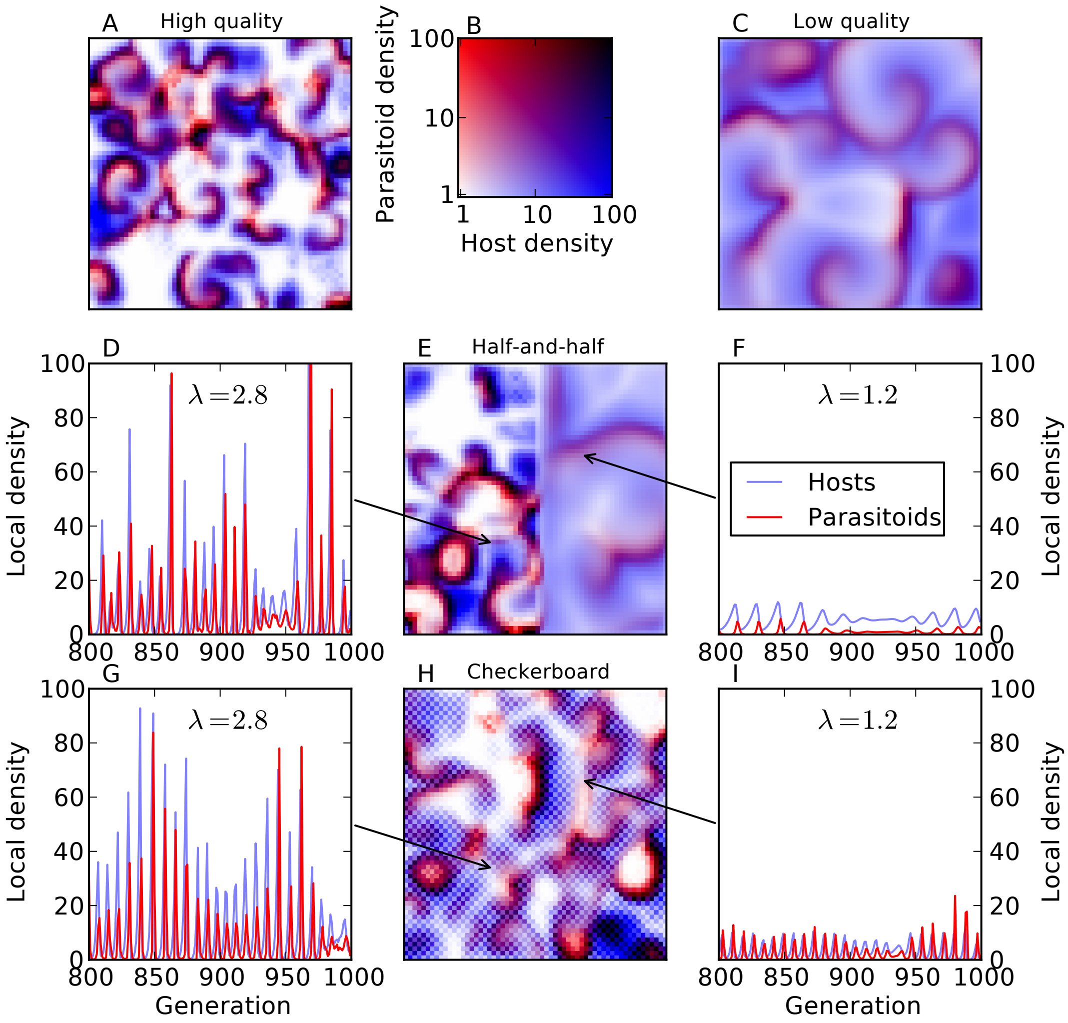


Figure 5.2: Snapshots of host and parasitoid densities in space after 1000 generations in single realizations of (A) a homogeneous high fecundity ($\lambda = 2.8$) landscape (C) a homogeneous low fecundity ($\lambda = 1.2$) landscape (E) a half-and-half landscape ($\lambda_+ = 2.8$, $\lambda_- = 1.2$) and (H) a checkerboard landscape ($\lambda_+ = 2.8$, $\lambda_- = 1.2$). Higher host densities are indicated by more blue and higher parasitoid densities by more red (see colormap in panel B). Local populations of hosts (in blue) and parasitoids (in red) are plotted over the last 200 of 1000 generations on (D) a high quality plant in the half-and-half landscape (F) a low quality plant in the half-and-half landscape (G) a high quality plant in the checkerboard landscape and (I) a low quality plant in the checkerboard landscape.

and-half landscapes consistently supported slightly larger populations than did the unconnected, additive landscapes. In contrast, herbivore populations in checkerboard landscapes were lowest at intermediate levels of variation in plant quality. Parasitoid populations in checkerboard landscapes, however, did not undergo a corresponding decrease as variance in plant quality increases, but rather stayed roughly constant until increasing along with host densities when variance in plant quality was large.

In fact, checkerboard landscapes had more parasitoids per host than other landscapes (Figure 5.4). While the ratio of parasitoids to hosts increased with the difference in plant quality in all landscapes, the most dramatic increase occurred in the checkerboard landscape. The half-and-half landscape also had more parasitoids to hosts than either additive landscape, although the difference was less extreme as long as low quality plants could support growth in host populations ($\lambda_{-} > 1$).

The temporal variance in population densities also differed with the spatial arrangement of high and low quality plants (Figure 5.3 C & D). In all cases, the variance in both host and parasitoid populations increased relative to the mean population as the variance in plant quality increased. However, each type of landscape showed a different pattern of increase. The coefficient of variation of host and parasitoid populations (calculated by dividing the standard deviation of the population in time by the mean population) increased gradually with variance in plant quality in half-and-half landscapes and both arrangements of additive landscapes, though variation in additive half-and-half landscapes was consistently lower. Meanwhile, the coefficient of variation of populations in checkerboard landscapes only began increasing when plants varied greatly in quality.

These differences between populations on checkerboard landscapes, on half-and-half landscapes, and on their additive counterparts were not evenly distributed among high and low quality plants (see Figure 5.5). As one might expect on a landscape where every plant has some neighbors of the opposite quality and there is plenty of

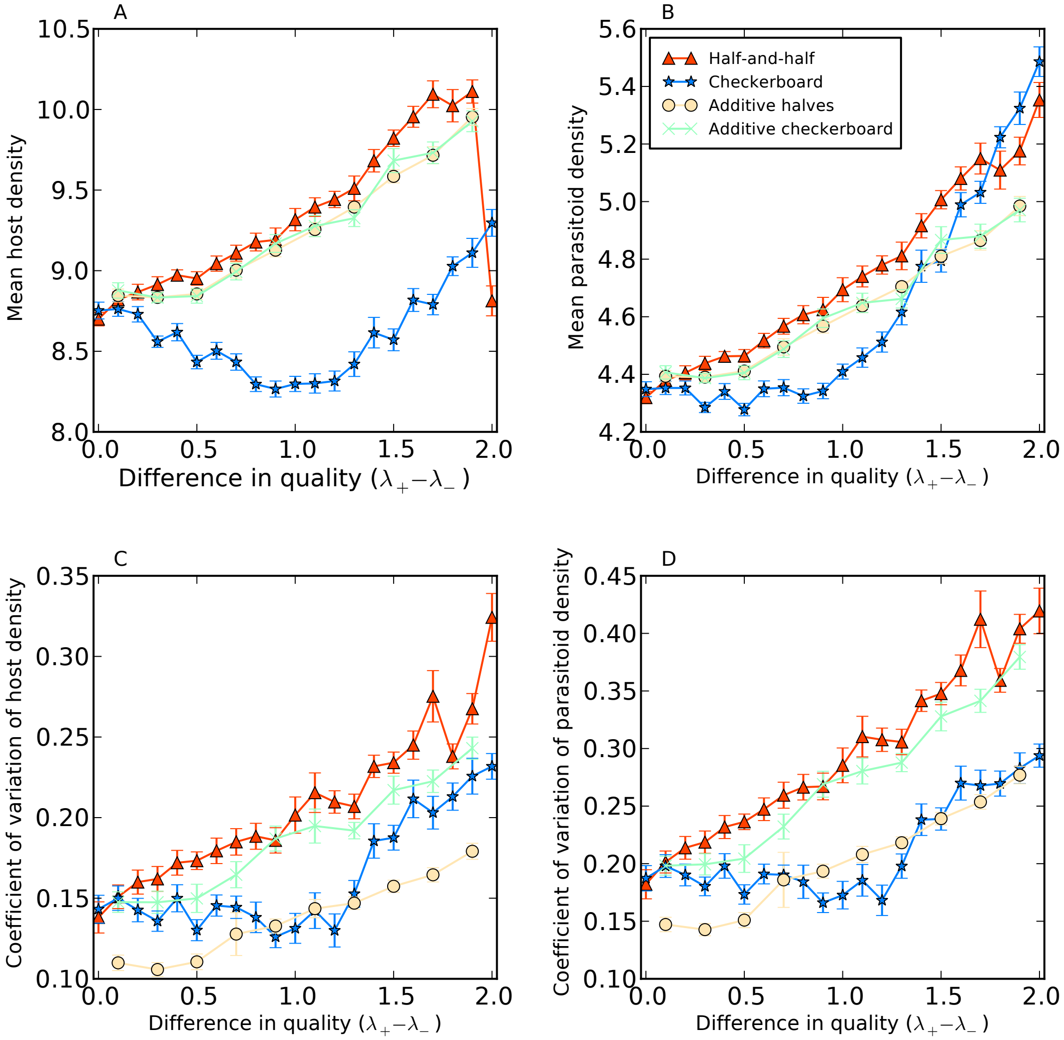


Figure 5.3: Plots of (A) mean host density, (B) mean parasitoid density, (C) coefficient of variation (mean / standard deviation) of host density, (D) coefficient of variation (mean / standard deviation) of parasitoid density during the last 200 of 1000 generations plotted against the difference in plant quality. Error bars show one standard error above and below the mean value for twenty replicate fields.

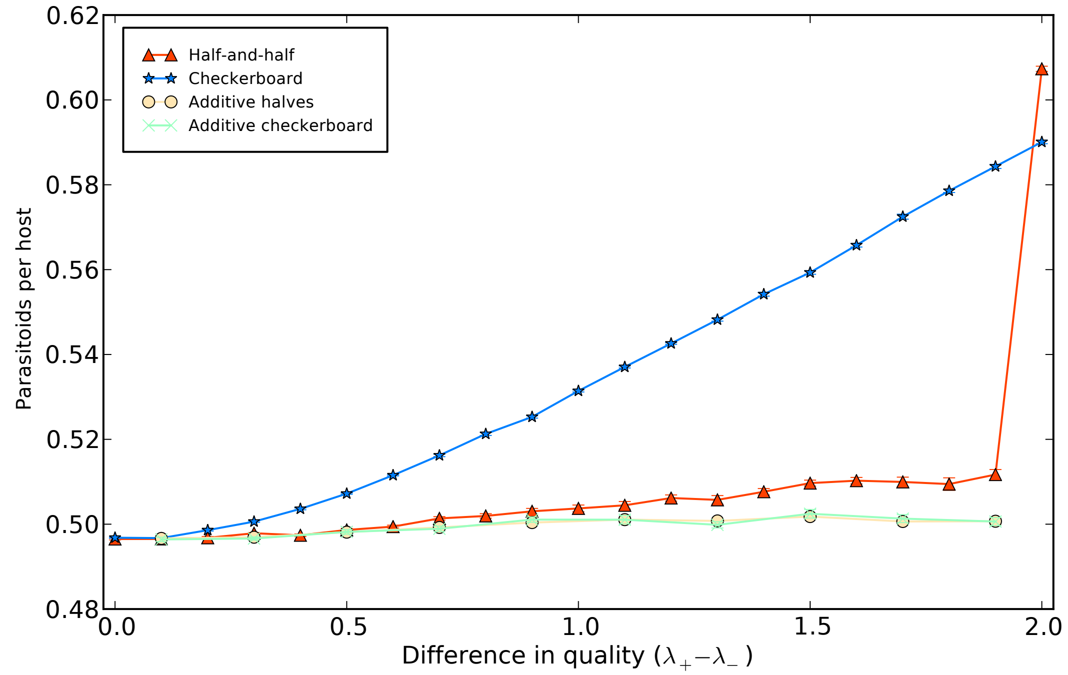


Figure 5.4: The mean parasitoids per host during the last during the last 200 of 1000 generations plotted against the difference in plant quality. Error bars show one standard error above and below the mean value for twenty replicate fields. However, the standard deviations in each case were consistently small enough that the error bars are not visible.

dispersal between high and low quality plants, high quality plants on checkerboard landscapes had lower densities of parasitoids than high quality plants on additive landscapes did (Figure 5.5B), while parasitoid densities were higher on low quality plants on checkerboard landscapes than on additive landscape (Figure 5.5D). These differences became more pronounced as the difference in host fecundity on high and low quality plants increased. However, host populations on checkerboard landscapes did not follow this same pattern. On low quality plants, host populations were even lower on checkerboard landscapes than on additive landscapes, likely due to the higher parasitoid density (Figure 5.5C). At the same time, high quality plants on checkerboard landscapes supported at most slightly larger host populations than high quality plants on additive landscapes did (Figure 5.5A), even with their substantially lower parasitoid densities.

Populations on half-and-half landscapes also differed from those on additive landscapes. On both high quality and low quality plants on half-and-half landscapes, parasitoid populations were slightly higher than on the corresponding additive landscapes (Figure 5.5B & D). However, populations of hosts on high quality plants were even higher (Figure 5.5A) and populations of hosts on low quality plants even lower (Figure 5.5C) than on high and low quality plants, respectively, in additive landscapes.

In order to understand the somewhat counterintuitive behavior of the half-and-half landscapes, we examined local populations along a transect on individual plants from the high fecundity side of the field to the low fecundity side (Figure 5.6 A & B). The difference between populations on the half-and-half field and those on its non–interacting, additive counterpart seems to stem from the plants near the boundary between high and low quality plants. The low quality plants that are situated near high quality plants have especially low host populations while those high quality plants, as well as those near the absorbing boundary of the landscape,

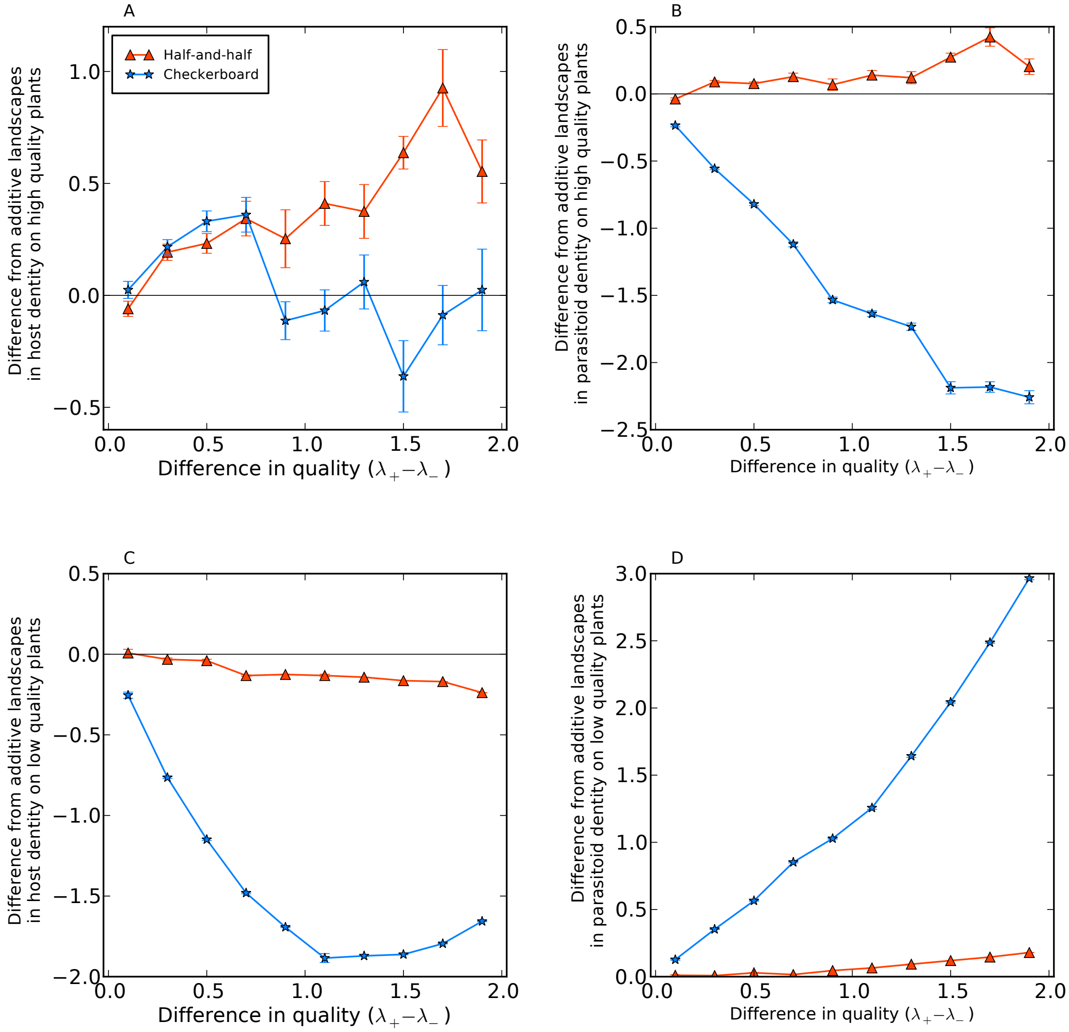


Figure 5.5: Plots of the difference between connected and additive landscapes in the (A) mean host density on high quality plants, (B) mean parasitoid density on high quality plants, (C) mean host density on low quality plants, and (D) mean parasitoid density on low quality plants during the last 200 of 1000 generations, plotted against the difference in plant quality. For clarity, the horizontal line where the difference equals zero is plotted in black.

have especially high populations. When we look at the frequency at which these local host populations oscillate, we see that the low quality border plants share the rapid oscillations of the high fecundity region, rather than the slower, long period outbreak cycle of the low fecundity side (Figure 5.6 C & D).

This pattern seems to be driven by fast, high density periodic traveling waves from the high fecundity side crossing the border to the low quality side (see video supplement C.1). At a lower per capita host fecundity, the high densities of hosts and parasitoids are no longer sustainable, so a rapid crash occurs. Because these crashes on the low fecundity side are driven by the same traveling waves dominating the dynamics on the high fecundity side, they share the same periodicity. This phenomenon can be seen in an example run in video supplement C.1).

5.4 Sensitivity

We varied the fraction of hosts and parasitoids dispersing each generation, $m_H = 0.1, 0.15, 0.2, 0.25.0.3$ and $m_P = 0.7, 0.75, 0.8, 0.85, 0.9$, and the aggregation of the parasitoids, $\mu = 0.0, 0.5, 1.0, 1.5, 2.0$. For each set of dispersal parameters, we fixed the mean plant quality at 2.0 and simulated 4 replicate landscapes each with checkerboard and half-and-half arrangements at $\lambda_+ - \lambda_- = 0.0(\text{homogeneous}), 0.5, 1.0, 1.5, 2.0(\text{source-sink})$.

Our results throughout this region of parameter space were qualitatively similar (Figure 5.8 & Figure 5.7), with the exception of the case where parasitoids did not aggregate to plants with more hosts ($\mu = 0$). When parasitoids preferentially dispersed to plants with more herbivore hosts ($\mu > 0$), herbivore populations in checkerboard landscapes decreased steadily (at $\mu = 2$) or decreased and then increased as variance in plant quality increased (Figure 5.7). In all half-and-half landscapes, herbivore populations increased or stayed constant as variance in plant quality increased, until decreasing sharply when the lower quality plants became unable to sustain growth in host populations ($\lambda_- \leq 1$).

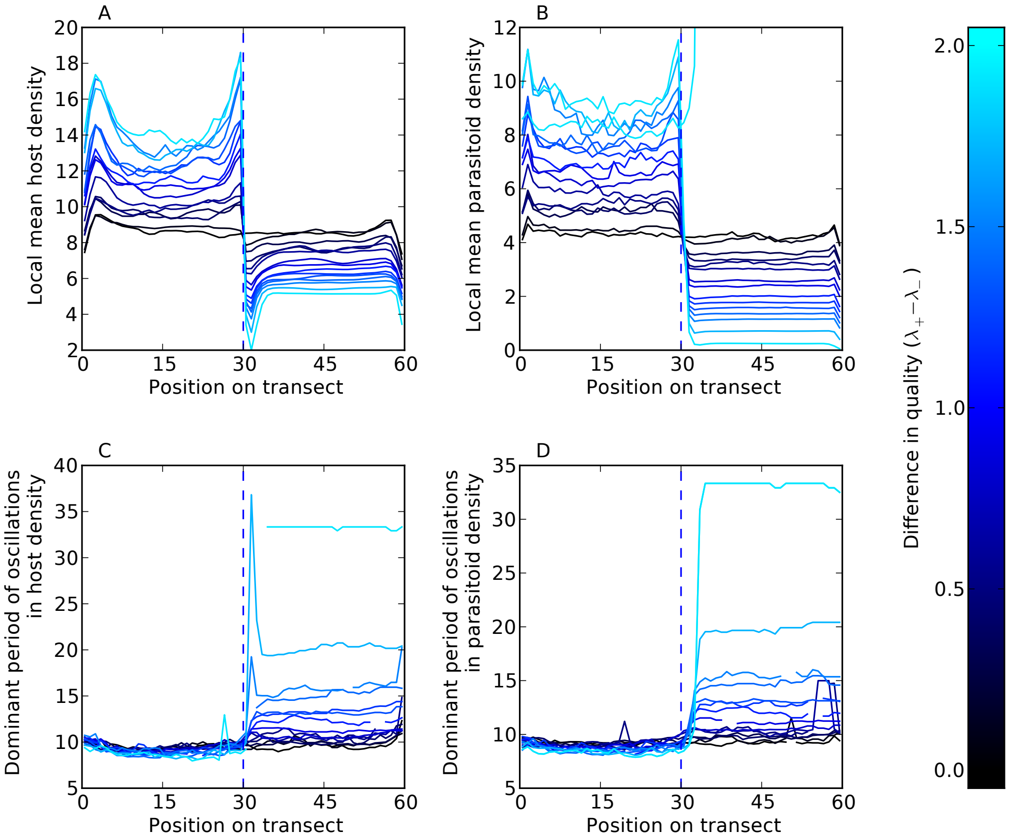


Figure 5.6: Plots of (A) mean host density, (B) mean parasitoid density, (C) dominant period of oscillations in host density, (D) dominant period of oscillations in parasitoid density during the last 200 of 1000 generations at each plant along a transect. Each line shows the average of 20 runs on identical fields starting from random initial conditions. Line color indicates the difference in host fecundity ($\lambda_+ - \lambda_-$) between the high and low quality plants. The border between high quality plants to the left of the boundary (host fecundity λ_+) and low quality plants to the right of the boundary (host fecundity λ_-) is indicated with a blue, dashed line.

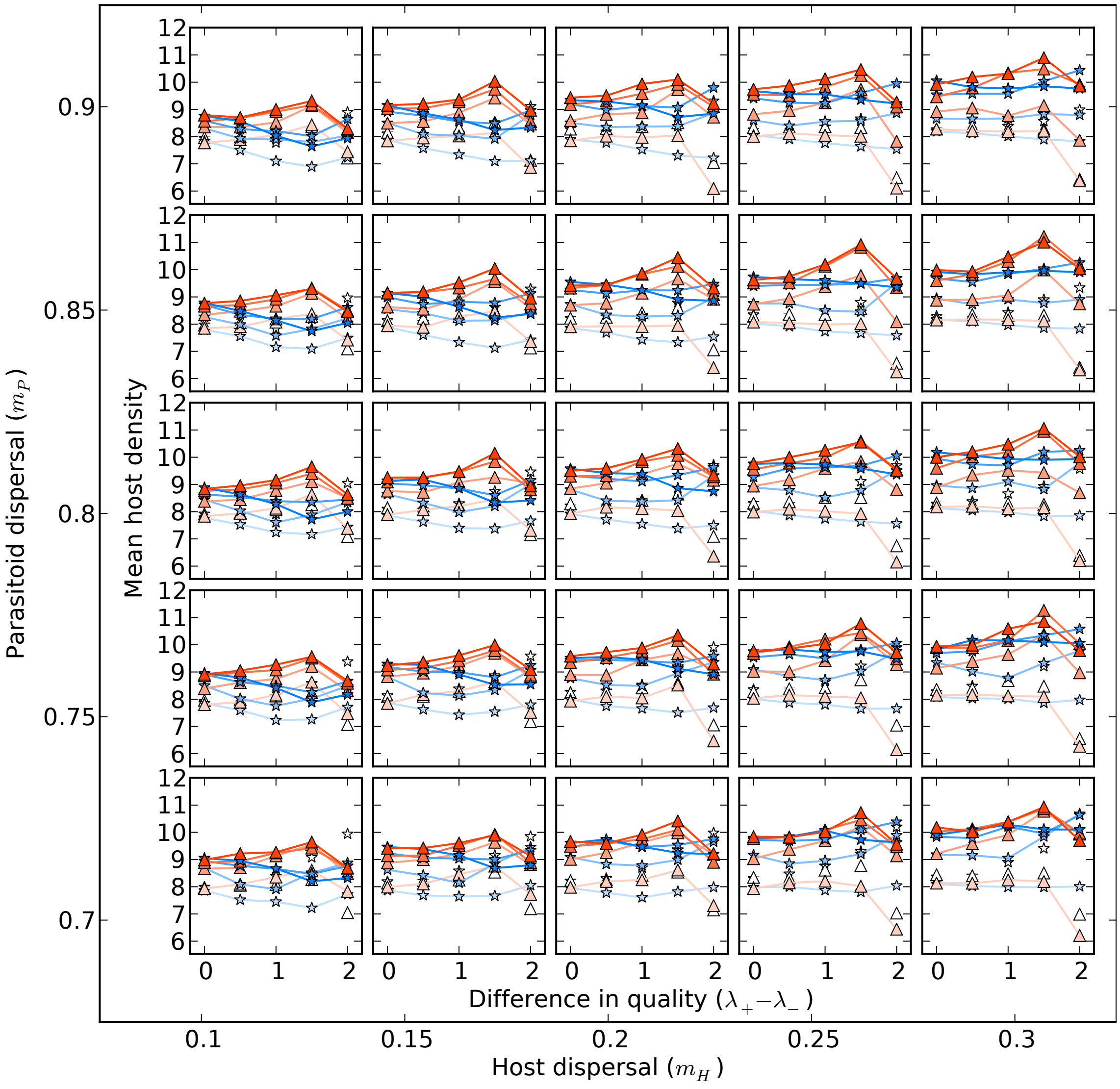


Figure 5.7: Effect of variation in plant quality on mean host density explored under varying dispersal parameters. Each line shows the mean host density during the last 200 of 1000 generations of four replicate landscapes plotted against the difference in plant quality. Standard error bars are omitted for clarity. Half-and-half landscapes are represented with red triangles, checkerboard landscapes with blue stars. The saturation of color varies with parasitoid aggregation (μ) from purely diffusing parasitoids (white, $\mu = 0$) to strong aggregation (darkest color, $\mu = 2$). The subplot column indicates the fraction of hosts dispersing each generation (m_H), subplot row indicates the fraction of parasitoids dispersing each generation (m_P).

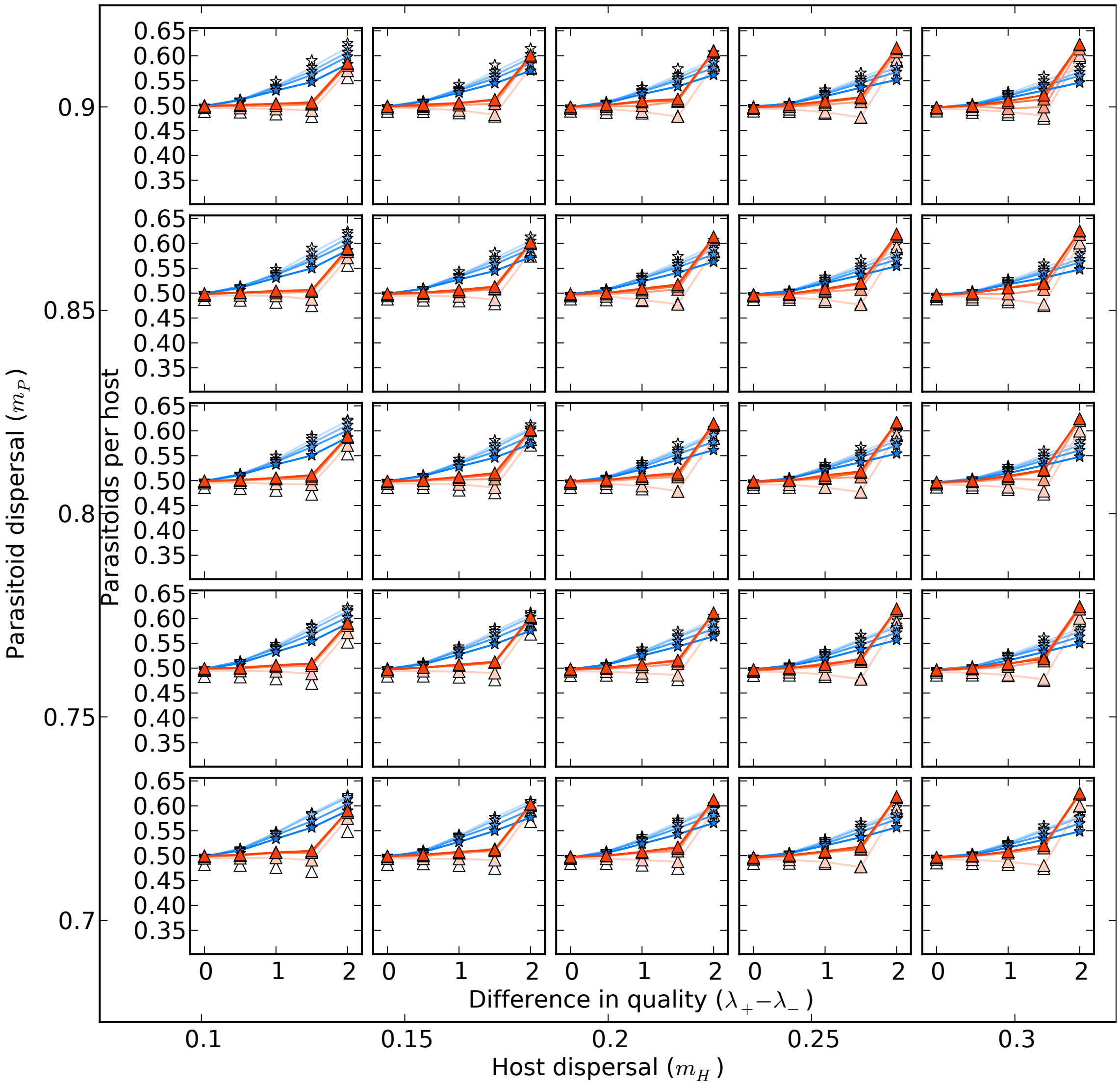


Figure 5.8: Effect of variation in plant quality on the ratio of parasitoids to hosts explored under varying dispersal parameters. Each line shows the parasite to host ratio during the last 200 of 1000 generations of four replicate landscapes plotted against the difference in plant quality. Standard error bars are omitted for clarity. Half-and-half landscapes are represented with red triangles, checkerboard landscapes with blue stars. The saturation of color varies with parasitoid aggregation (μ) from purely diffusing parasitoids (white, $\mu = 0$) to strong aggregation (darkest color, $\mu = 2$). The subplot column indicates the fraction of hosts dispersing each generation (m_H), subplot row indicates the fraction of parasitoids dispersing each generation (m_P).

When parasitoids were non-aggregating ($\mu = 0$), checkerboard landscapes had higher densities of herbivores than half-and-half landscapes, and the herbivore populations in checkerboard landscapes increased rapidly with variance in plant quality (Figure 5.7), suggesting that some degree of aggregation is necessary to produce the suppression of host populations on low-quality plants in the checkerboard landscape observed in Figure 5.5.

5.5 Discussion

We used a spatially explicit model of host-parasitoid dynamics to explore the interaction between variation in resource quality and spatially structured trophic interactions. Even in very simple landscapes, we found that local variation in resource quality influences global population dynamics dramatically.

When either the spatial scale of variation in plant quality or the magnitude of that variation was large, both the mean and variance of host and parasitoid populations in heterogeneous landscapes were substantially greater than those in homogeneous landscapes with the same mean plant quality. The ratio of parasitoids to hosts also increased with the variance in plant qualities present in the landscape. This confirms that the spatially structured interaction between herbivores and their natural enemies can combine with the effects of variable plant quality to generate spatial and temporal dynamics observed previously in field studies (Underwood, 2004; Helms and Hunter, 2005; Underwood, 2009). Moreover, field studies have found that the shape and spatial arrangement of high quality plant patches can also affect herbivory (Evans et al., 2012*a*). In our model, landscapes with different spatial arrangements of plants also gave rise to differences in population dynamics.

Most strikingly, we found that small variations in plant quality occurring on a small spatial scale led to decreased herbivore density. The frequent dispersal of parasitoids between high and low quality plants led to strong suppression of herbivore populations on the lower quality plants, without a corresponding increase in herbivore populations on higher quality plants, an effect that Oksanen *et al.* (1992) found for predator-prey interactions in habitats where the scale of heterogeneity is smaller than the home-range of an individual predator. This enhancement of parasitism by small-scale variation in plant quality may have particularly important consequences for productive plant populations. Because herbivore fecundity is, in many cases, affected by plant genotype (Underwood and Rausher, 2000; McIntyre and Whitham, 2003; Evans et al., 2012*b*), we might expect the genetic variation in natural plant populations to lead to more local variation in plant quality and higher parasitism than we would see in an agricultural planting of a single plant genotype. Indeed, adding genetic variation in plant quality back into agricultural ecosystems might serve to increase rates of parasitism on pest insects.

We see a similar effect in landscapes composed of two large, homogeneous patches of different quality, where host populations were suppressed on low-quality plants that were near the boundary between low- and high-quality regions, resembling the "spillover effect" observed at crop–noncrop boundaries (Rand and Louda, 2006). In their study of predatory coccinellid beetles in crops and grassland sites in Nebraska, Rand and Louda found that coccinellid densities were $2.6 - 9.0$ times higher in grassland sites in landscapes containing mostly crops than in grassland sites surrounded by more grassland (Rand and Louda, 2006). At the same time, they found that aphid density increased significantly when predators were excluded from experimental aphid colonies via a mesh sleeve, suggesting that the coccinellids could indeed suppress aphid populations and that the spillover of beetles from cultivated areas could lead to lower aphid populations in the surrounding grasslands.

The sharp delineation between high and low plant quality in our simulated landscapes may reasonably approximate an intercropped field, but natural systems often contain smoother gradients in plant quality and arthropod population density over space (Salmore and Hunter, 2001; Pennings et al., 2007). For example, Salmore and Hunter (2001) found that the concentration of defensive alkaloids in *Sanginaria canadensis* (bloodroot) along an elevational gradient from Georgia to North Carolina tended to decrease with elevation. Pennings *et al.* (2007) reported a latitudinal gradient of plant palatability in their study of plant-herbivore interactions in European salt-marshes. Leaves and plant litter from higher latitudes were more palatable than their low-latitude congenerics to several insect species. Several of the plant taxa studied also tended to suffer more herbivory at lower latitudes than higher, suggesting that the latitudinal gradient in plant palatability might result from differing selective pressure from herbivory.

Such resource gradients, coupled with herbivore-natural enemy interactions, have the potential to drive patterns in insect herbivore populations over large spatial and temporal scales. In their investigation of the periodic traveling waves of larch budmoths (*Zeiraphera diniana* Gn.) across the European Alps, Bjørnstad *et al.* (2002) simulated the populations of an herbivore host and its parasitoid along a resource gradient. They found that, under a wide range of parameter values, the addition of a spatial gradient in resource quality caused the emergence of periodic traveling waves of hosts and parasitoids similar to those observed in larch budmoth populations or to the winter moth (*Operphtera brumata*) outbreaks that periodically move across Europe (Tenow et al., 2013).

The magnitude and spatial scale of variation in plant quality can depend on both abiotic factors, such as variation in altitude and moisture, and life-history and dispersal traits of plants (Loveless and Hamrick, 1984). Different parasitoid species respond to variation in host habitat quality at different spatial scales, and it is likely that these differences stem at least in part from differences in the dispersal patterns of the parasitoids and their hosts (Roland and Taylor, 1997). Both in models (Comins et al., 1992; Rohani and Miramontes, 1995) and in the field (Cronin and Reeve, 2005), dispersal of hosts and parasitoids can have strong effects on the stability and spatiotemporal patterns of abundance. Not only do host and parasitoid dispersal patterns have profound effects on their population dynamics, but landscape structure and diversity has been found to affect both herbivore (Hill et al., 1996) and parasitoid dispersal (Cronin, 2003). Further, the structure of spatial heterogeneities in resource quality may influence the evolution of dispersal strategies, with different landscapes selecting for different dispersal strategies (Johst et al., 2002).

In previous studies on homogeneous landscapes, the choice of dispersal parameters have been found to strongly affect the stability of host-parasitoid interactions, the efficiency of parasitoid control of hosts, and the emergent spatial patterns of host and parasitoid populations (Hassell and May, 1973; Hassell et al., 1991; Comins et al., 1992; Rohani and Miramontes, 1995; Bjørnstad et al., 2002; Hirzel et al., 2007). Though we have found that our main results hold for a range of parameters where parasitoids disperse much more frequently than their hosts and parasitoids preferentially dispersed to plants with higher host populations, our preliminary exploration of a wider range of dispersal behaviors suggest that host and parasitoid populations with different dispersal may have qualitatively different responses to environmental heterogeneity. The interplay between host and parasitoid dispersal behaviors and spatial variation in plant quality remains a rich avenue for future study.

Models similar to the one used here have been used to simulate a variety of processes in homogenous landscapes. These include a wide variety of demographic processes acting on herbivores and their parasitoids, and the attack behavior of foraging enemies (Ruxton and Rohani, 1996; Hassell, 2000). We suggest that exploring such interactions on landscapes that differ in resource quality for herbivores could provide interesting insights into population dynamics of natural and agricultural systems. For example, the combined effect of demographic stochasticity and a strong spillover effect might make a low quality region effectively impermeable to hosts, even if a deterministic model allowed very small densities of hosts and parasitoids to colonize it.

The large scale consequences of other plant quality effects on hosts and parasitoids also remain open to exploration. For simplicity, we only modeled spatial heterogeneity in host fecundity, but there is strong evidence to suggest that the size and growth rate of hosts, which also vary with plant quality, can affect parasitoid reproductive success (Hunter, 2003). Additional factors contributing to parasitoid fecundity, such as availability of overwintering sites or alternate food sources, may also vary spatially (Cronin and Reeve, 2005), but need not correlate with plant quality. Further, plant quality is, in turn, influenced by host and parasitoid populations. In the short term, damage from host herbivory may reduce plant quality for the next generation of hosts. In the longer term, high levels of herbivory select for plants with stronger defenses against hosts. Another potential extension of our model would be to incorporate some of these tritrophic interactions between plants, herbivore hosts, and parasitoids.

Even without the additional complexity of these mechanisms, we find that large scale population dynamics of hosts and parasitoids depend both on the distribution of plant qualities and their behaviors in space. Our results suggest that fine-scale variation in plant quality may be particularly important for supporting populations of parasitoids and predators and that plant genetic variation in agricultural ecosystems could be managed to enhance biological control of herbivores.

CHAPTER VI

Diffusion induced spatial structure in the logistic map

6.1 Introduction

The role of spatial structure in shaping populations is a topic of longstanding interest in ecology (Huffaker, 1958; Elton, 1958; Levin, 1976; Denno and McClure, 1983; Hassell et al., 1991; Levin, 1992). The dispersal of organisms and the resulting spatial coupling of populations can, under some circumstances, destabilize a spatially homogeneous equilibrium population and drive the emergence of spatially structured heterogeneity (Turing, 1952; Kot, 1989; Neubert et al., 1995; Rohani and Ruxton, 1999). Away from equilibrium, heterogeneity among coupled populations can enable the long-term survival of populations that would not persist in isolation, whether through desynchronization of the timing of local extinction events (Levin, 1976) or through spatiotemporal heterogeneity in the local risk of predation or parasitism (Huffaker, 1958; Hassell et al., 1991; Comins et al., 1992).

One important factor in the emergence and persistence of spatial heterogeneity is the scale at which dispersal occurs. For example, Kot (1989) found that diffusion driven instability in a simple integrodifference model of predator-prey dynamics depended on the widths of the predator's and prey's dispersal kernels. The distance scales of dispersal influenced not only the occurrence of diffusion driven instability but also the range of spatial wavelengths over which the homogeneous equilibrium was unstable to perturbations.

However, such analytical results relating dispersal patterns to spatial heterogeneity in population density have been largely concerned with the stability of spatially homogeneous equilibrium solutions, while our understanding of non-equilibrium dynamics comes primarily from numerical experiments. While conditions under which the stable equilibrium of a well-mixed system becomes unstable with spatial structure have been established in several discrete-time population models (Kot, 1989; Neubert et al., 1995; Rohani and Ruxton, 1999), less is known about the emergence of spatial heterogeneity when the dynamics of the well-mixed system do not tend towards a stable fixed point.

I aim to investigate analytically the relationship between spatial scales of dispersal and the emergence of spatial heterogeneity in the absence of a stable equilibrium of the well-mixed system. I use a simple integrodifference model of a spatially structured population of organisms which reproduce and disperse in different life stages (Kot and Schaffer, 1986; Hardin et al., 1988). Such simple models of this form have been known to exhibit a wide range of phenomena, including diffusion driven instability and pattern formation (Kot, 1989; Andersen, 1991; Neubert et al., 1995), traveling waves (Wang et al., 2002; Lutscher and Van Minh, 2013), and long-lasting transient dynamics (Hastings and Higgins, 1994).

6.2 Population Model

Let $N_t(x)$ be the density of sessile adults at position x during generation t. I model the local density of juveniles produced as the logistic map,

$$g(N_t(x)) = rN_t(x)(1 - N_t(x)) \tag{6.1}$$

with intrinsic growth rate r. The juveniles disperse and settle in new positions, so that

$$N_{t+1}(x) = \int\limits_{\Omega} rN_t(y)(1 - N_t(y))\kappa(x, y)\mathrm{d}y \tag{6.2}$$

where κ is a dispersal kernel. In the current study, I consider a Gaussian dispersal kernel, so that $N_{t+1}(x)$ is the result of allowing $rN_t(x)(1 - N_t(x))$ to diffuse for some fixed time under some fixed diffusion constant.

Let us consider the diffusion operator mapping $f(x)$ to $\int_{\Omega} f(y)\kappa(x, y)\mathrm{d}y$, in terms of its damping effects on spatial heterogeneity. Because the dispersal kernel κ is Gaussian, it is useful to note that

$$\int\limits_{\Omega} \cos(\omega y)\kappa(x, y)\mathrm{d}y = e^{-\alpha\omega^2} \cos(\omega x).$$

where α is a dimensionless parameter depending on the width of the kernel. For convenience, I will call the diffusion operator D^{α}, such that

$$D^{\alpha} \cos(\omega x) = e^{-\alpha\omega^2} \cos(\omega x).$$

This allows us to write the integrodifference equation for the next generation's population as

$$N_{t+1}(x) = D^{\alpha} rN_t(x)(1 - N_t(x)) \tag{6.3}$$

with a single growth parameter r and a single dispersal parameter α.

6.3 Linear stability of the spatially homogeneous equilibrium

Consider a small perturbation about the spatially homogeneous equilibrium population,

$$N_t(x) = N^* + \epsilon\zeta_t(x) \tag{6.4}$$

where $N^* = \frac{r-1}{r}$ is the nonzero fixed point of the logistic map for $r > 1$, which is stable for $1 < r < 3$. By calculating the next generation's population density

$$N_{t+1}(x) = N^* + \epsilon r(1 - 2N^*)D^\alpha \zeta_t(x) + O(\epsilon^2) \tag{6.5}$$

and noting that $r(1 - 2N^*) = (r - 2)$, one can see that

$$\zeta_{t+1}(x) = (r - 2)D^\alpha \zeta_t(x). \tag{6.6}$$

Looking at the Fourier transform of the perturbation, one can see that

$$\hat{\zeta}_{t+1}(\omega) = (2 - r)e^{-\alpha\omega^2}\hat{\zeta}_t(\omega). \tag{6.7}$$

Thus the component of ζ_t at wavenumber ω is amplified (or damped) by a factor of $(r - 2)e^{-\alpha\omega^2}$. For $1 < r < 3$, this factor is always between zero and one, so the spatially homogeneous equilibrium $N \equiv N^*$ is linearly stable to perturbations in any spatial frequency. However, for $r > 3$, the homogeneous equilibrium loses stability to perturbations at wavenumbers less than

$$\omega_c = \left(\frac{\log(r-2)}{\alpha}\right)^{\frac{1}{2}},$$

where the growth away from the unstable equilibrium is able to overcome the damping from diffusion.

The magnitude of amplification of linear perturbations of the spatially homogeneous equilibrium is plotted as a function of wavenumber in Figure 6.1 (for an intrinsic growth rate of $r = 3.2$), with sample simulations of selected values of ω. One can see that, indeed, perturbations at wavenumbers below w_c (insets A, B, and C) are initially amplified, perturbations at wavenumbers above w_c (insets E and F) are

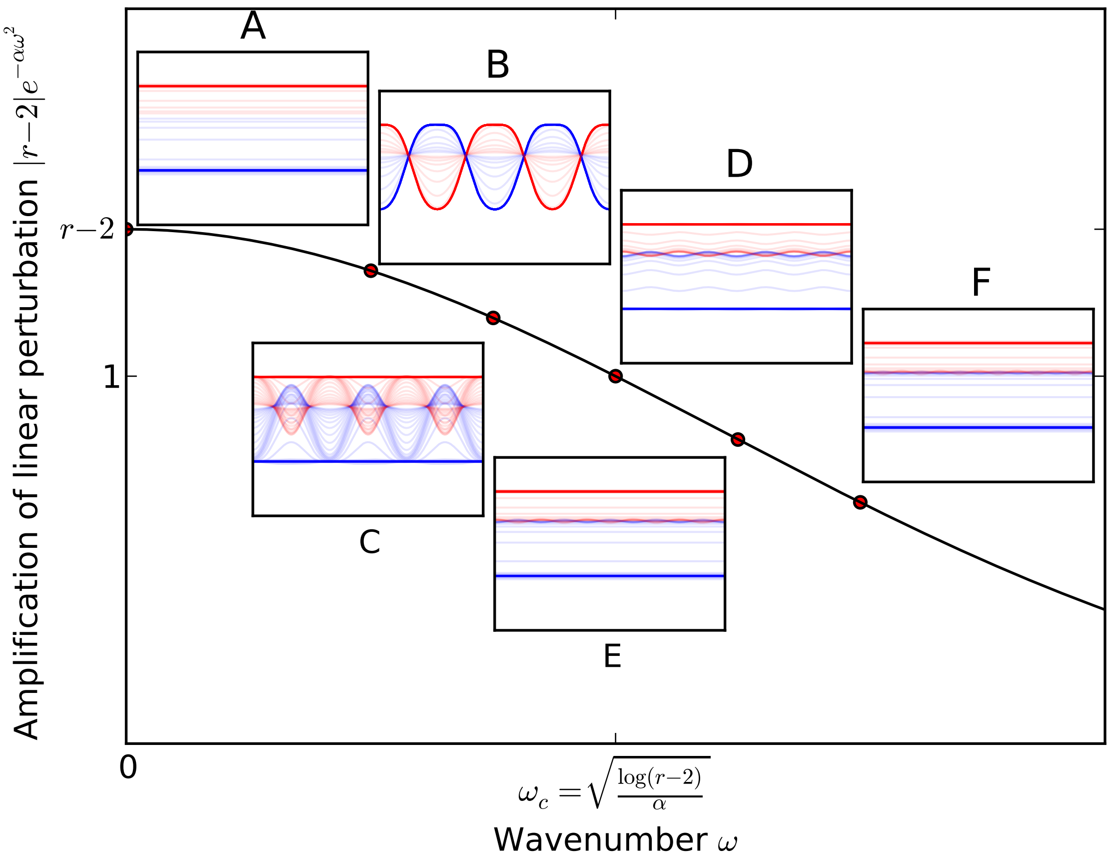


Figure 6.1: The amplification of perturbations about the spatially homogeneous equilibrium plotted by wavenumber. Insets A-F show the simulated populations for 100 generations starting from a small (amplitude $= 0.01r$) sinusoidal perturbation with wavenumbers (A) 0, (B) $\frac{w_c}{2}$, (C) $\frac{3w_c}{4}$, (D) w_c, (E) $\frac{5w_c}{4}$, and (F) $\frac{3w_c}{2}$ (marked with red circles on the line plot). Even and odd numbered generations are plotted in red and blue, respectively, so that the temporal oscillations are apparent. Deeper colors indicate more recent generations. The parameters used in the simulation were $r = 3.4$ (for which the logistic map has a stable 2-cycle) and $\omega_c = 8$ in a 1024 point discretized space. Results with populations simulated in frequency space, preserving the first 64 spatial frequencies, are qualitatively unchanged.

rapidly damped, and a sinusoidal perturbation at wavenumber w_c (inset D) more-or-less maintains its amplitude until the population has moved some distance away from the equilibrium.

However, spatially structured populations (such as the population in inset B) seem to occur in an even narrower band of wavenumbers within the range of perturbations which are initially amplified (such as in inset C). To understand why this might be, let us examine what happens to a small sinusoidal perturbation about the spatially homogeneous equilibrium,

$$N_0(x) = N^* + \epsilon \cos(\omega x). \tag{6.8}$$

The resulting larval density is

$$g(N_0(x)) = N^* + \epsilon(2 - r)\cos(\omega x) - \epsilon^2 \cos^2(\omega x),$$

which can be written by means of the identity $\cos^2(\omega x) = \frac{1}{2}(1 - \cos(2\omega x))$ as

$$N_1(x) = N^* - \frac{r\epsilon^2}{2} + e^{-\omega^2\alpha}\epsilon(2 - r)\cos(\omega x) + \frac{r\epsilon^2}{2}e^{-4\omega^2\alpha}\cos(2\omega x). \tag{6.9}$$

Written in this form, it becomes clear that neglecting the $O(\epsilon^2)$ perturbation induced in the mean in favor of $O(e^{-\omega^2\alpha}\epsilon)$ terms will only be reasonable for a narrow band of wavenumbers (in particular, a band that grows as $O\big(\sqrt{-\log(\epsilon)}\big)$ as ϵ shrinks).

More generally, if the population density varies periodically in space with some period L, so that its Fouier series can be written as

$$N_t(x) = \sum_{m=-\infty}^{\infty} a_m(t) e^{\frac{2\pi m i}{L}x},$$

then the coefficients in the next generation $a_m(t+1)$ will have the form

$$a_m(t+1) = re^{-\alpha m^2}\left(a_m(t) + \sum_{j=-\infty}^{\infty} a_{m-j}(t)a_j(t)\right). \tag{6.10}$$

Since $N_t(x)$ is real, the coefficients $a_{m-j}(t)$ are the complex conjugates of $a_{j-m}(t)$, which are the coefficients of $e^{\frac{2\pi mi}{L}}N_t(x)$. Thus, by Parseval's theorem,

$$\sum_{j=-\infty}^{\infty} a_{m-j}(t)a_j(t) = \frac{1}{2\pi L}\int_0^L e^{-\frac{2\pi mi}{L}}N_t^2(x)\mathrm{d}x.$$

Note that, in any generation, $N_t(x)$ is between zero and one for all x and t. It follows that

$$\Big|\sum_{j=-\infty}^{\infty} a_{m-j}(t)a_j(t)\Big| \leq 1$$

and therefore

$$|a_m(t+1)| \leq 2re^{-\alpha m^2}. \tag{6.11}$$

This strongly constrains the amplitude of variation in density that can be sustained at fine spatial scales.

6.4 Single spatial frequency approximation

With our previous results in mind, let us consider a population whose density varies sinusoidally in space,

$$N_t(x) = M + A\cos(\omega x)$$

for $\omega > 0$. The population density in the next generation will be

$$N_{t+1}(x) = rM(1-M) - \frac{r}{2}A^2 + e^{-\omega^2\alpha}r(1-2M)A\cos(\omega x) + O(e^{-4\omega^2\alpha}).$$

Neglecting the higher frequency terms, this yields a pair of coupled difference equations for the mean density M and the coefficient A of the spatial oscillation of the density in each generation:

$$M_{t+1} = rM_t(1 - M_t) - \frac{r}{2}A_t^2 \tag{6.12}$$

$$A_{t+1} = rd(1 - 2M_t)A_t \tag{6.13}$$

where $d = e^{-\alpha\omega^2}$ is the damping effect of diffusion on the focal spatial frequency. One can quickly see that $d = 0$ (the limit of high frequency or high α) will always yield a spatially homogeneous population, where $d = 1$ (α equal to zero) corresponds to a population with no dispersal. Further, a spatially homogeneous ($A = 0$) orbit that exists for one value of d must exist for all values of d.

The extent to which the behavior of the single spatial frequency model reflects that of of the full system will be investigated in Section 6.5. For now, let us explore the dynamics of this simpler model.

The fixed points of this single spatial frequency system are the spatially homogeneous fixed points ($A = 0$ and $M = 0$ or $\frac{r-1}{r}$) and an additional pair of fixed points where

$$1 = rd(1 - 2M)$$

and

$$M = rM(1 - M) - \frac{r}{2}A^2.$$

Simplifying these conditions, one can see that any heterogeneous fixed points must satisfy

$$M = \frac{1}{2}\Big(1 - \frac{1}{rd}\Big)$$

and

$$A^2 = 2M\Big(1 - \frac{1}{r} - M\Big).$$

To investigate the linear stability of these fixed points, one can calculate the Jacobian of the system

$$J|_{(M,A)} = \begin{pmatrix} d(1-2M) & -rA \\ -2rdA & rd(1-2M) \end{pmatrix}. \tag{6.14}$$

At the spatially heterogeneous fixed points, this is equal to

$$J = \begin{pmatrix} \frac{1}{d} & -rA \\ -2rdA & 1 \end{pmatrix}.$$

Using the Jury test (Jury, 1964, 1974), all eigenvalues of J will have magnitude less than one (*i.e.* the fixed point will be stable) if and only if the following three conditions hold

$$1 + \operatorname{Tr} J + \det J > 0 \tag{6.15}$$

$$1 - \operatorname{Tr} J + \det J > 0 \tag{6.16}$$

$$1 - \det J > 0. \tag{6.17}$$

At the spatially heterogeneous fixed points,

$$\operatorname{Tr} J = 1 + \frac{1}{d},$$

$$\det J = \frac{1}{r} - 2dr^2A^2,$$

and

$$1 - \operatorname{Tr} J + \det J = -2dr^2A^2 < 0.$$

Thus, any spatially heterogeneous fixed points will violate (6.16) and are therefore never stable.

One can similarly look for regions of parameter space with stable 2-cycles. A natural special case to investigate is 2-cycles of the form observed in Fig. 6.1B, where the mean population density (M) and amplitude ($|A|$) of the spatial oscillations remain fixed, but phase of the oscillation (the sign of A) alternates. In such 2-cycles, the mean density and the amplitude at the focal spatial frequency must obey

$$-1 = rd(1 - 2M)$$

$$A^2 = 2M(1 - \frac{1}{r} - M)$$

and the matrix relevant for linear stability would be the product of the Jacobian at each step around the cycle,

$$J = \begin{pmatrix} \frac{1}{d} & -rA \\ -2rdA & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{d} & rA \\ 2rdA & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{d^2} - 2dr^2A^2 & \frac{1}{d}rA - rA \\ -2rA + 2rdA & 1 - 2dr^2A^2 \end{pmatrix}.$$

Again applying the Jury test, one can find regions where these spatially structured stable 2-cycles are stable for all growth rates high enough that homogeneous 2-cycles exist, even for large growth rates which would generate chaotic dynamics in the spatially homogeneous system (see Fig. 6.2).

It is natural to wonder whether similar regions of spatially structured stability exist for cycles of other lengths. When a stable n-cycle exists in the uniform diffusion, spatially homogeneous case ($d = 0$), spatially structured cycles with periodicities belonging to the set of n and its factors must exist in the no-diffusion ($d = 1$) case, simply by virtue of each local population following the stable n-cycle in its own phase. In the single spatial frequency approximation, one would expect to see n such cycles (including the homogeneous cycle), each corresponding to a different phase lag. For a cycle $N_1, \ldots, N_n$, the no-diffusion cycle at lag ρ would have mean populations equal to $\frac{1}{2}(N_i + N_{i+\rho})$ and amplitudes equal to $\frac{1}{2}(N_i - N_{i+\rho})$ for $i = 1, \ldots, n$. However, deter-

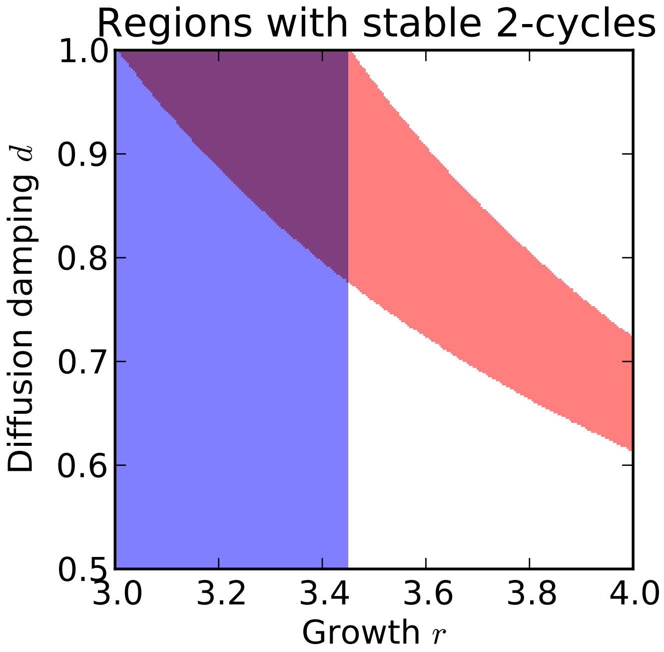


Figure 6.2: Regions with theoretically stable 2-cycles. Red shading indicates that our single spatial frequency model predicts stable spatially structured 2-cycles with a fixed mean and amplitudes of alternating sign. Blue indicates regions with spatially homogeneous 2-cycles.

mining the conditions for the existence and stability of analogous spatially structured cycles in diffusion populations is more challenging. In the following section, I will explore the bifurcation structure of both this single spatial frequency approximation and the full spatial model numerically.

6.5 Numerical results

In order to characterize the long-term behavior at each level of growth and diffusion, I simulated populations in the single spatial frequency system (Fig. 6.3) and the full spatial system (Fig. 6.5) under a range of growth and diffusion values. For each set of parameter values, I ran 1024 replicates with random initial conditions and plotted the number of values of the mean population density and amplitude of spatial oscillations in density observed after in the fairly long term (1000 generations). In the full spatial case, the damping from diffusion d and the amplitude observed were both defined relative to a focal wavenumber (Fig. 6.5 used focal wavenumber 1). Note that in cases where longer term transient dynamics occur, one would expect to observe

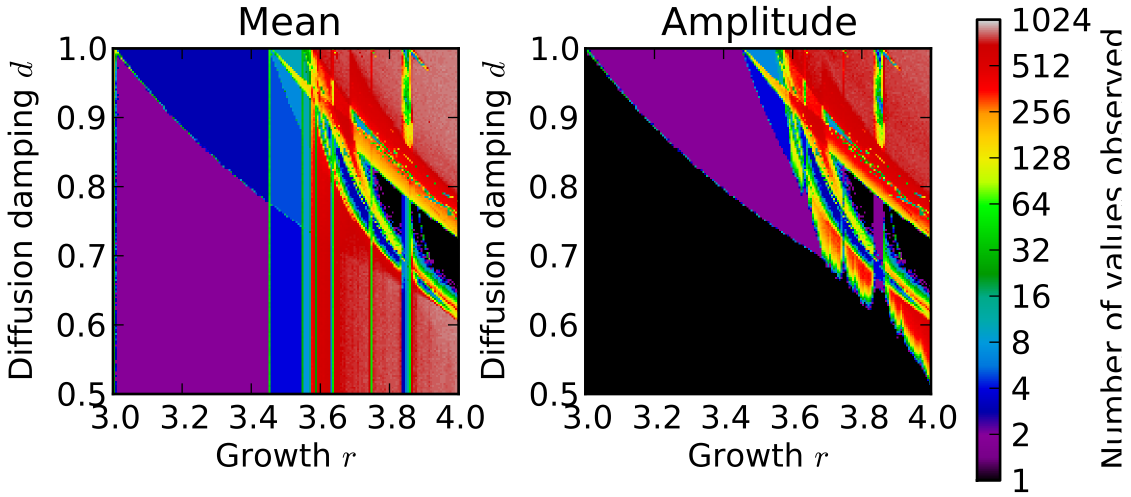


Figure 6.3: The number of observed values of the mean (left panel) and amplitude (right panel) after 1000 generations in 1024 runs of the model at each damping d, growth g value. Initial conditions were generated by choosing M_0 uniformly at random from $[0, 1]$ and choosing A_0 uniformly at random from $[-\min(M_0, 1-M_0), \min(M_0, 1-M_0)]$ to keep the population density at each point in space in $[0, 1]$.

more values than actually appear in the set of stable orbits.

In the single spatial frequency system (Fig. 6.3), the boundaries of the region containing a stable spatially structured two-cycle are visible. For growth rates with stable spatially homogeneous 2-cycles ($3 < r < 1 + \sqrt{6}$), both spatially homogeneous and spatially structured 2-cycles are present when diffusion is weak enough (*i.e.* when d is large enough). However, for larger growth rates where the spatially homogeneous system displays chaotic dynamics, the spatially structured 2-cycle appears to be the only pattern observed in the long-term for some levels of diffusion (see Fig. 6.4A for an example).

Intriguingly, one can see a similar band of structure that appears to run from the range of growth rates with stable homogeneous 4-cycles in the no diffusion case down to the lower edge of the stable structured 2-cycle region as the growth rate increases.

One might conjecture that this is an analogous band of stable spatially structured 4-cycles and, indeed, this is supported by our numerical results (see Fig. 6.4B).

While by no means identical, the bifurcation structure of the full spatial system (plotted in Fig. 6.5) has many features in common with the single spatial frequency approximation. In particular, there are similar regions of stable spatially heterogeneous 2-cycles and 4-cycles (see Fig. 6.6A&B for examples). Above the region of stable spatially heterogeneous 2-cycles (that is, with slightly weaker diffusion), one can see see an additional region of what seems to be stable spatially heterogeneous 4-cycles that do not occur in the single spatial frequency model (compare panel C in Figs. 6.4&6.6).

The other regions of order visible in Fig. 6.5 that are not captured in the single spatial frequency model appear to correspond to stable spatially heterogeneous 2-cycles in higher wavenumbers (see Fig. 6.7). I found that populations tended to take longer to converge to spatially heterogeneous orbits in higher wavenumbers, which may be one reason that the ordered in parameter space corresponding to 2-cycles in higher spatial frequencies were not as well-characterized as their wavenumber 1 counterpart. Another reason may be the presence of other stable states. For example, in the band where the single spatial frequency approximation predicts spatially structured 2-cycles with wavenumber 6, it is more common for 4-cycles with wavenumber 4 to occcur, though only after a very long transient period (see Fig. 6.8).

6.6 Discussion

I analyzed a discrete time, continuous space model of population growth, with a spatially structured population alternating local growth according to the well-studied logistic map and dispersal using a Gaussian kernel. This simple model produced a wide range of behaviors at varying levels of growth and dispersal. I found that many of these complexities were captured in a further simplified model, looking at only the

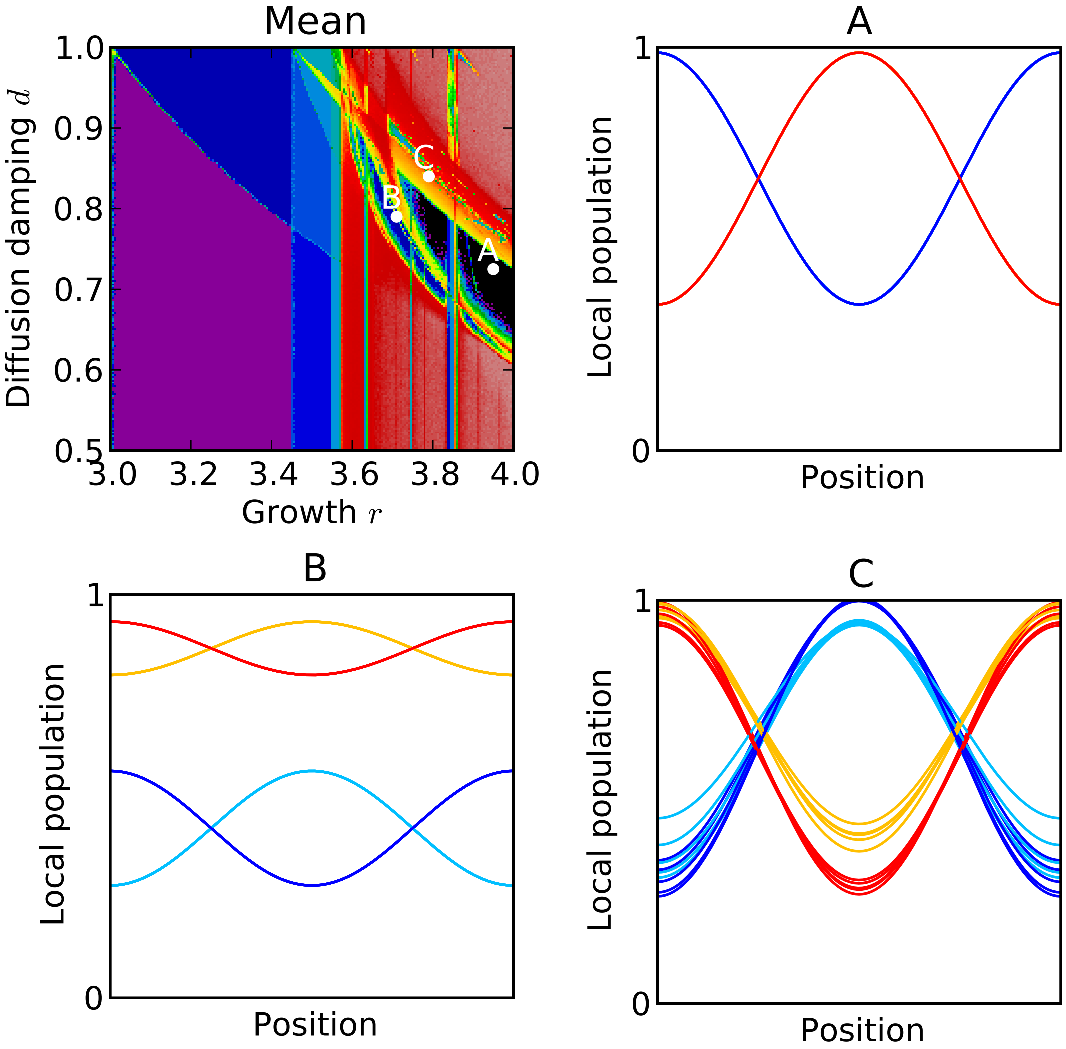


Figure 6.4: Examples of orbits observed in the single spatial frequency model. The points in parameter space corresponding to panels A, B, and C are labeled in the top left panel, which is a reproduction of the left panel of Fig. 6.3. Orbits are plotted over the last 20 generations of a 1000 generation run, in 4 alternating colors so that the periodicity is visible.

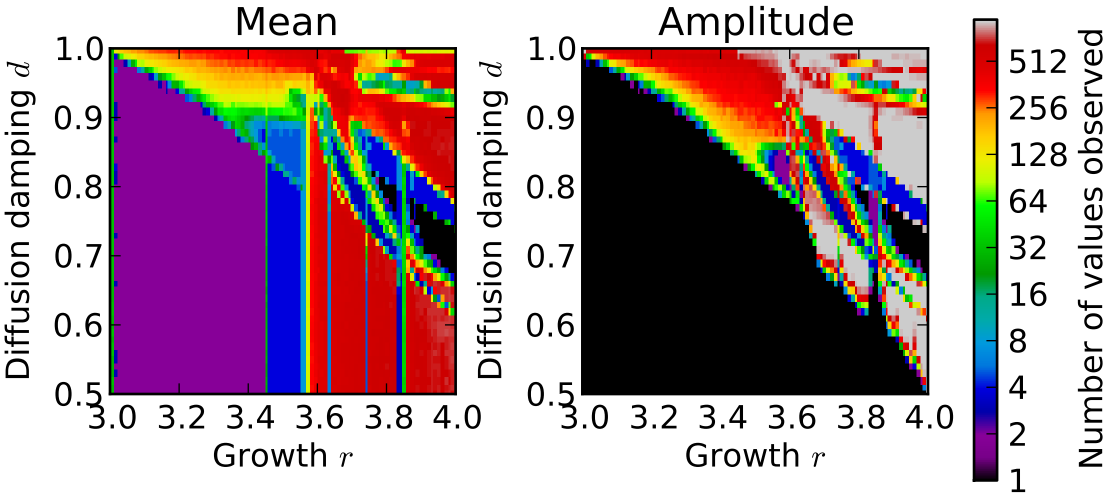


Figure 6.5: The number of observed values of the mean (left panel) and the amplitude of the focal spatial frequency (right panel) after 1000 generations in 1024 runs of the model at each damping d, growth g value. Initial conditions were generated by choosing an initial sinusoidal (wavenumber 1) population, mean M_0 uniformly at random from $[0, 1]$ and amplitude A_0 uniformly at random from $[-\min(M_0, 1 - M_0), \min(M_0, 1 - M_0)]$ to keep the population density at each point in space in $[0, 1]$. A 1024 point discretized space was used for these simulations.

mean population and the magnitude of spatial variation at a single focal spatial scale.

I found that spatially heterogeneous populations with temporally periodic dynamics occurred even in parameter regimes where the spatially synchronous dynamics were chaotic. The range of spatial scales on which spatial heterogeneities could persist was limited by dispersal. As in other reaction-diffusion systems (Turing, 1952), continuous time models of density dependent populations display pattern formation at some characteristic spatial scale that depends of dispersal parameters. Such a relationship between spatial scales of correlated abundance and of movement has also been observed in the field (Roos, 1991). In our discrete time model, I was able to quantify the extent to which spatial variation at scales that are "too small" will be smoothed out by dispersal (as in eq. 6.11).

Because of the strong constraints on the amplitude of variation at fine spatial scales, a single spatial frequency approximation was able to capture many features of our original model for a wide range of dispersal scales. The bifurcation structure of our full spatial model shared many recognizable structures with the single spatial frequency model and, especially in ordered regions, often behaved similarly (Figs. 6.6 and 6.4). In the single frequency approximation, I was able to determine analytically conditions under which certain classes of periodic solution would be stable (Fig. 6.2), which were borne out in our numerical simulations (Fig. 6.3). I also observed signatures of these regions in the full spatial model (Fig. 6.5).

When diffusion was local enough that variations at multiple spatial scales could sustain nontrivial dynamics, on the other hand, deviation from the single spatial frequency model became the norm. In this regime, transient population dynamics often persisted on very long time scales (as in Fig. 6.8), as observed in Hastings and Higgins's study using the Ricker model (Hastings and Higgins, 1994). It is possible that these long transient dynamics are due to the presence of multiple saddle nodes in solution space (Cushing et al., 1998; Hanski et al., 1995). One potential direction of future study could be to investigate the stability of the 2-periodic, spatial wavenumber 2 solutions examined in this study to perturbations at different spatial frequencies.

Even with longer-distance diffusion, there remained some differences between behavior of the single spatial frequency model and that of the full spatial model. For example, compare the solutions shown in Fig. 6.4C and Fig. 6.6C, which are observed at the same parameter values in the single spatial frequency approximation and full spatial model, respectively. The potential of nonlinear interactions to sustain high-frequency 'harmonics' despite the smoothing effects of diffusion has long been recognized (Steele, 1974). In the present model, I find that such 'harmonics,' even at relatively small amplitudes, are critical for the existence of certain solutions, such as the 2-periodic solution in Fig. 6.6C, which exists in the full spatial model but not in the single spatial frequency approximation (see Fig. 6.4C).

The spatial scale of dispersal and the emergent patterns of population density are tightly linked in our simple model. I find that many characteristics of the integrodifference equation model can be captured by approximation the heterogeneity to a small number of focal frequencies, and that some classes of solution can be investigated analytically in this finite approximation. This approximation breaks down when the use of different focal frequencies predicts different solutions. I conjecture that competition between these "overlapping" solutions at different spatial scales may play a role in the long-lasting transient dynamics observed in this and other spatially coupled discrete time population models.

Although several studies have found that dispersal induced instability of a homogeneous population at equilibrium requires the interaction of multiple species with sufficiently different dispersal rates (Kot, 1989; Neubert et al., 1995; Rohani and Ruxton, 1999), I find that the same does not hold for the diffusion induced destabilization of periodic orbits. Rather, when the well-mixed dynamics of a single population do not tend towards a stable equilibrium point, dispersal at intermediate length scales can lead to the emergence of a wide variety of spatially ordered population dynamics. Our results further suggest that self-organized spatiotemporal structure in non-equilibrium populations depends strongly on the relative scales of individual organism movement distances and the size of the available habitat, with spatial structure breaking down in the extreme cases of global, uniform dispersal or no dispersal, resulting in a single well-mixed population or a collection of isolated, well-mixed populations, respectively.

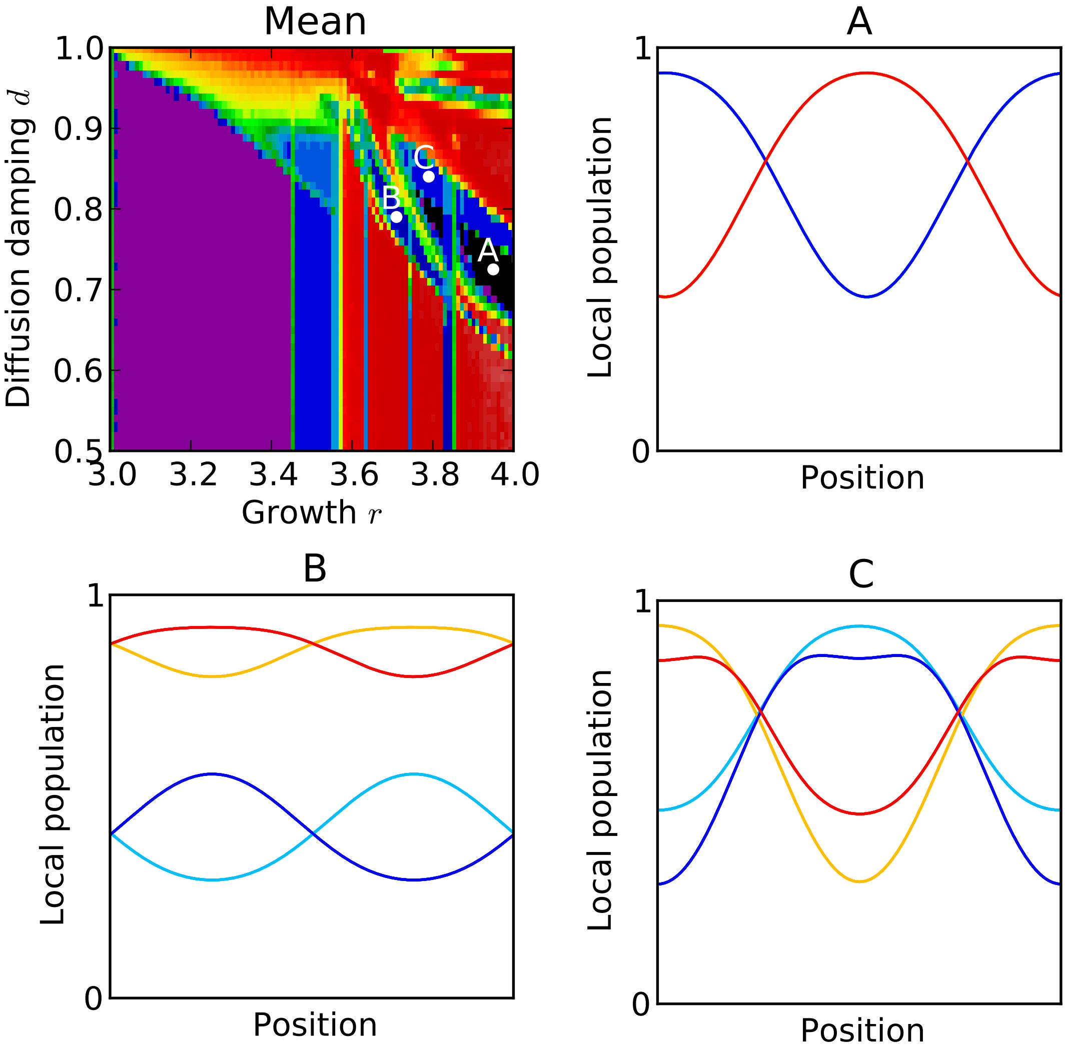


Figure 6.6: Examples of orbits observed in the full spatial model. The points in parameter space corresponding to panels A, B, and C are labeled in the top left panel, which is a reproduction of the left panel of Fig. 6.5. Orbits are plotted over the last 20 generations of a 1000 generation run, in 4 alternating colors so that the periodicity is visible.

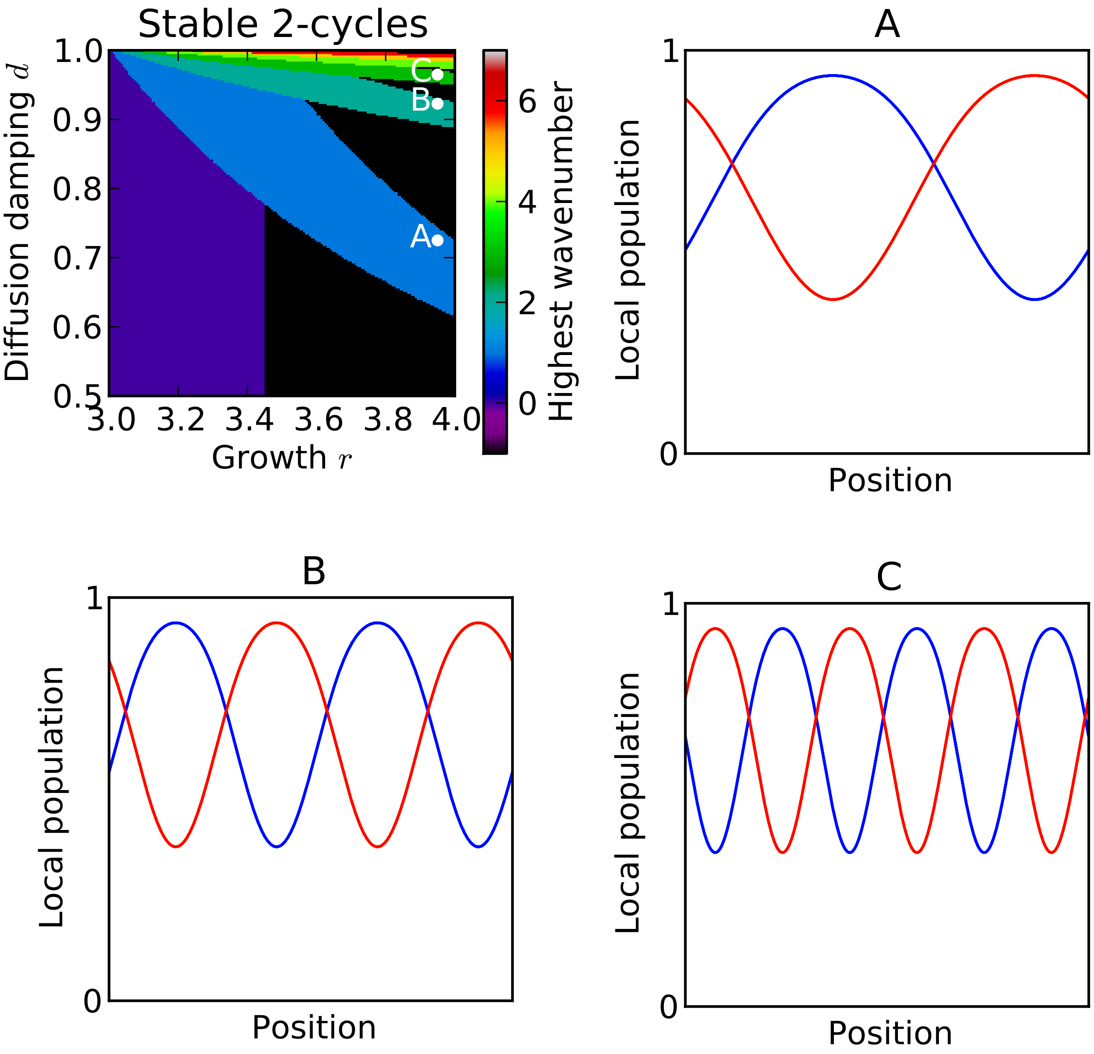
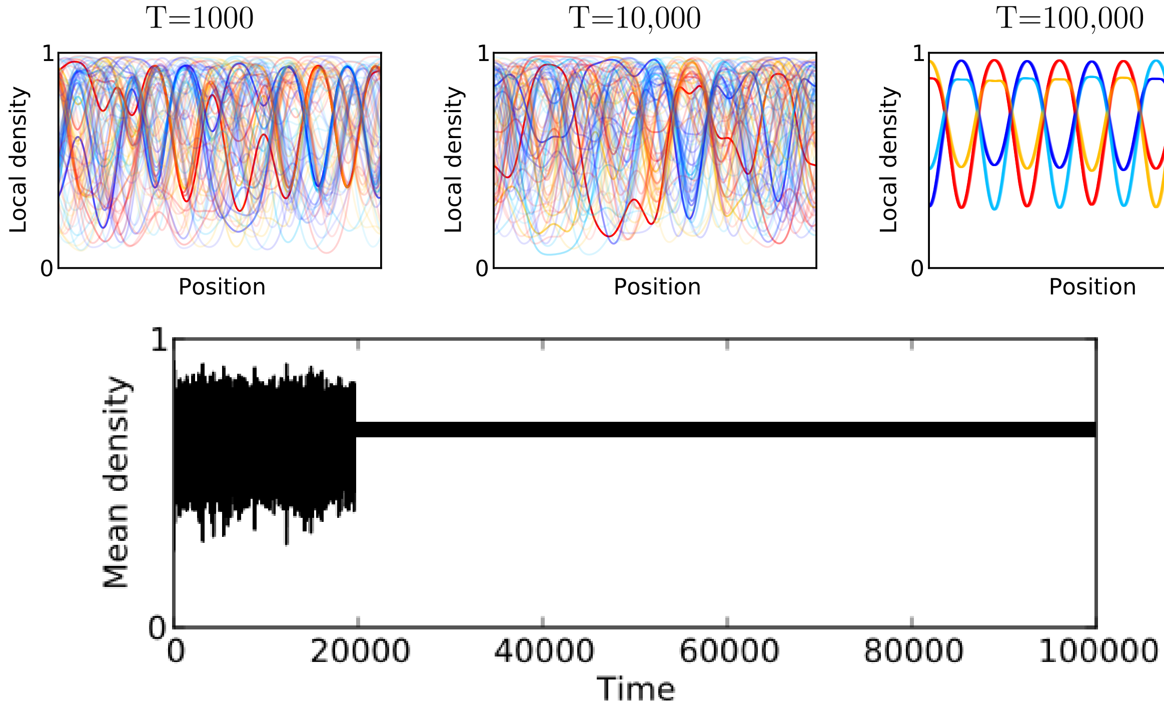


Figure 6.7: Regions with theoretically stable 2-cycles according to the single spatial frequency model, with example orbits in the full spatial model plotted in panels A, B, and C. In the top left panel, color indicates the highest wavenumber predicted to have a stable 2-cycle in the region. The damping from diffusion d is calculated according to the wavenumber 1 oscillations. The points in parameter space corresponding to panels A, B, and C are labeled in the top left panel. These points have the same growth and damping of the relevant wavenumber as point A in Figs. 6.4&6.6. The orbits in panels A, B, and C are plotted over the last 20 generations of 1000, 10000, and 100000 runs, respectively.



T=1000

T=10,000

T=100,000

Figure 6.8: Long term dynamics in at a point in parameter space where the single spatial frequency model would predict a stable spatially structured 2-cycle or wavenumber 6. The top 3 panels show the local populations for the 20 generations before generations number 1000, 10,000 and 100,000. The bottom panel shows the mean density over time. Note that, rather than the expected 2-cycle with wavenumber 6, the population eventually converges to a 4-cycle with dominant spatial wavenumber 4, although this population is not perfectly spatially periodic (compare the peaks of the dark blue line).

APPENDICES

APPENDIX A

Vaccine legacy supplementary information

A.1 Contact matrix processing and alternative contact structures

We calculated the contact rates used in our model using the reported contacts (both conversational and physical) person per day in Great Britatin in Table S8.4(a) of Mossong *et al.* (2008) and the esimated population by year of age from the United Kingdom's 2011 census data.(Mossong et al., 2008; United Kingdom. Office for National Statistics., 1988)

When using self-reported data on contact rates, it is often neccessary to correct for reciprocity in reporting in order that the rates of contacts with group i reported by group j members and vice versa predict a consistent number of contact events between the two groups.(Wallinga et al., 2006) Let c_{ij} be the actual average number of contacts per day that a member of group i has with members of group j and N_i be the number of members of group i in the POLYMOD study. Then the average number of contacts e_{ij} that occur per day between groups i and j is

$$e_{ij} = N_i c_{ij} = N_j c_{ji} = e_{ji}$$

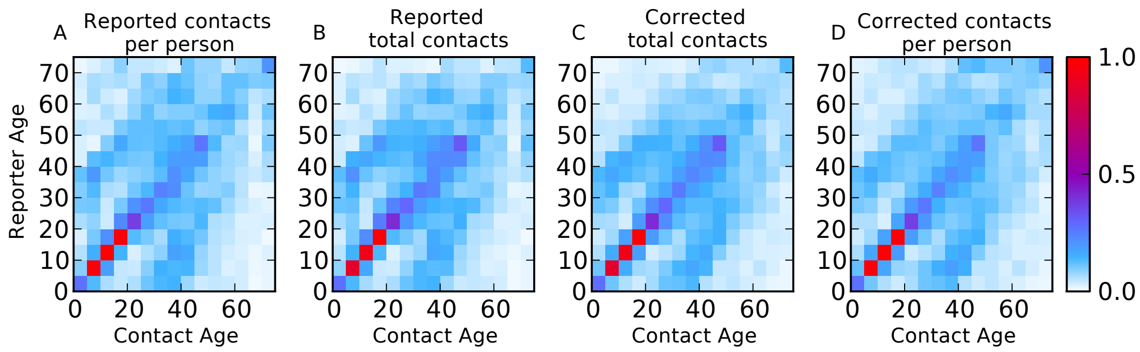


Figure A.1: Contacts between age groups. From left to right: reported per person rate of contacts, reported rate of contact events, rate of contact events after correcting for reciprocity in reporting, per person contact rates corrected for reciprocity. For this plot, each matrix is normalized to have a maximum value of one.

We estimate the number of contact events between groups i and j as

$$e_{ij} = \tfrac{1}{2}(N_i\tilde{c}_{ij} + N_j\tilde{c}_{ji})$$

where $\tilde{c}_{ij}$ is the number of contacts per day with group j reported by members of group i. This yields the symmetric matrix of contact events plotted in Figure 2.2A and Figure A.1C.

From this matrix we can recover the individual contacts per day corrected for reciprocity in reporting

$$c_{ij} = \frac{e_{ij}}{N_i} = \frac{1}{2}(\tilde{c}_{ij} + \frac{N_j}{N_i}\tilde{c}_{ji})$$

This corrected contact matrix (converted to units of contacts per year) is what we use in our simulations and is plotted in Figure A.1D.

We also carried out simulations of populations with homogeneous and proportional mixing. When mixing is homogeneous, contacts are assumed to be completely age-independent and the force of infection calculated using a single overall contact rate

$$c = \frac{\Sigma_{i,j}e_{ij}}{N}$$

In proportional mixing, an individual's total contacts per day can vary by age but who contacts who is random. So, for example, if 10-15 year olds were involved in one fifth of all contact events, then one fifth of each age group's contacts would be with 10-15 year olds.

In order to construct a contact matrix for proportional mixing, we can use our existing contact matrix. We denote the number of contact events involving age group i as

$$e_i = \Sigma_j N_i c_{ij}$$

which makes the total number of contacts

$$M = \frac{1}{2}\Sigma_j (e_j + N_j c_{jj})$$

(Since each contact involves two participants, simply adding up $\Sigma_j e_j$ double–counts contacts in which the two participants belong to different age-groups and each contribute to their own e_j.)

The chance that one's partner in a contact belongs to age group j is $\frac{e_j}{2M}$ (again, recalling that each contact involves two participants), so we can write the number of i, j contact events predicted under proportional mixing as

$$e'_{ij} = e_i \frac{e_j}{2M} = e'_{ji}$$

and the new per capita i to j contact rate is

$$c'_{ij} = \frac{e'_{ij}}{N_i}$$

The rates of contact events between age-groups with the full contact structure, with homogenous mixing, and with proportional mixing are plotted in Figure A.2.

We find a similar slow resurgence under all three mixing conditions, as seen in

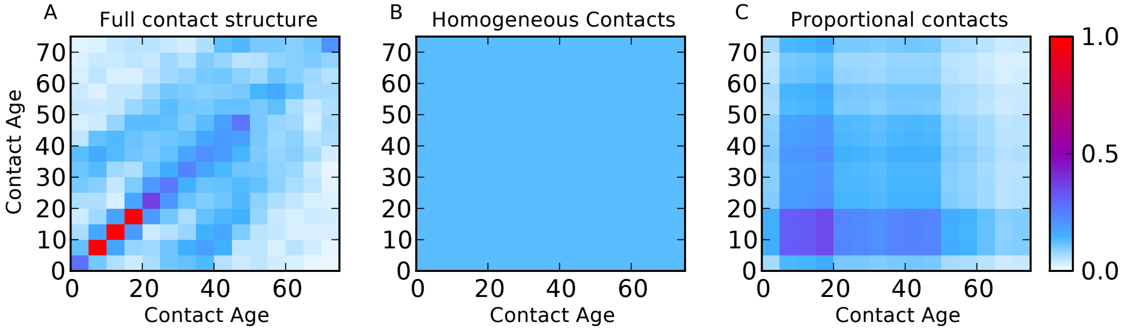


Figure A.2: Per person contact rates between age groups in three mixing models: full contact matrix (panel A), homogeneous mixing (panel B), and proportional mixing (panel C). For this plot, matrices are normalized so that the highest

Figure A.3, even though the pre-vaccine era mean age of infection is much higher without the age-assortative mixing observed in the POLYMOD. Scaling all contact rates by a constant in order to generate the historical pre-vaccine age of infection yields the same qualitative pattern (see Figure A.4) Although the age-structued incidences predicted under different contract structures vary substanitally in many respects, the gradual increase in adult cases seems to be a relatively robust feature regardless of the contact matrix used.

A.2 Historical Uptake Rates

From 1966 through 2008, we use the national vaccine coverage values reported by the Health Protection Agency (HPA) in our simulations.(United Kingdom. Health Protection Agency., 2008) However, no uptake data is available between 1957 and 1966. We examined several linear ramp ups in coverage, from coverage $\upsilon = 0.2, 0.4, 0.6$ in 1957 to the 74% coverage reported in 1966, so that coverage $u(t)$ between $t = 157$ (1957) and $t = 166$ (1966) given by

$$u(t) = \upsilon + \frac{0.74 - \upsilon}{9}$$

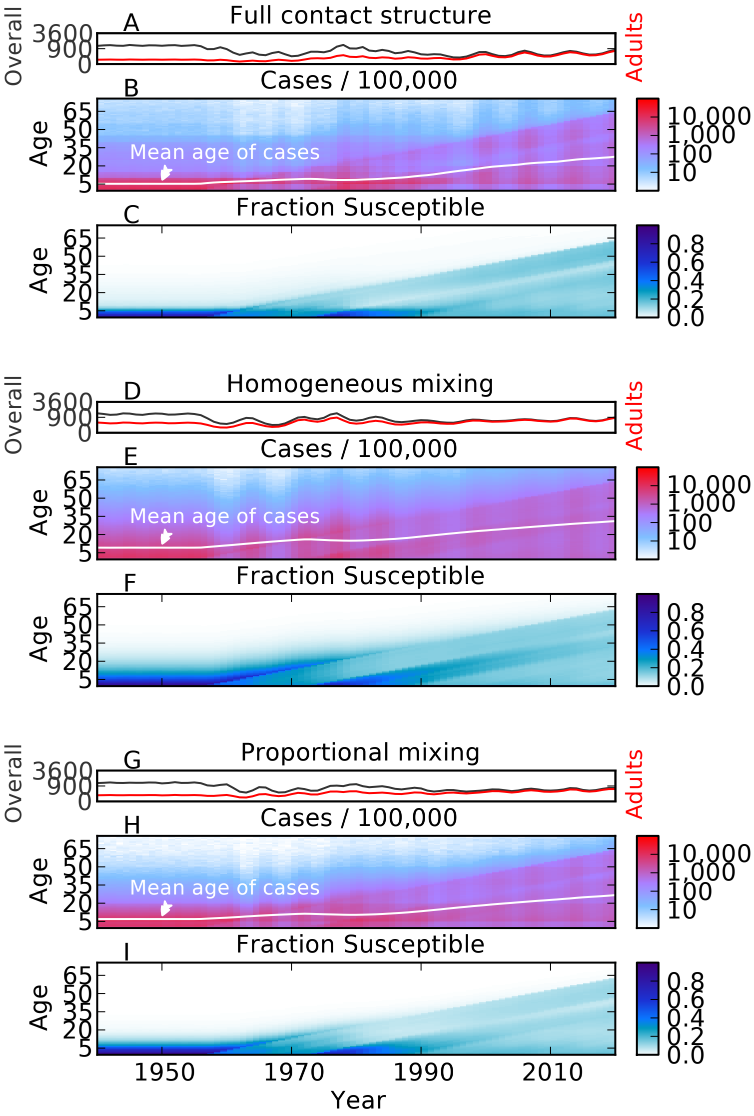


Figure A.3: Single realizations under each contact structure are plotted in panels A-C (full contact matrix), D-F (homogeneous mixing), and G-I (proportional mixing) with 60% vaccination in 1957 and 85% vaccine efficacy. For each realization overall incidence and incidence in individuals over 15 year old (panels A, D, G); the incidence by age (panels B, E, H); and fraction susceptible by age (panels C, F, I) are plotted against time.

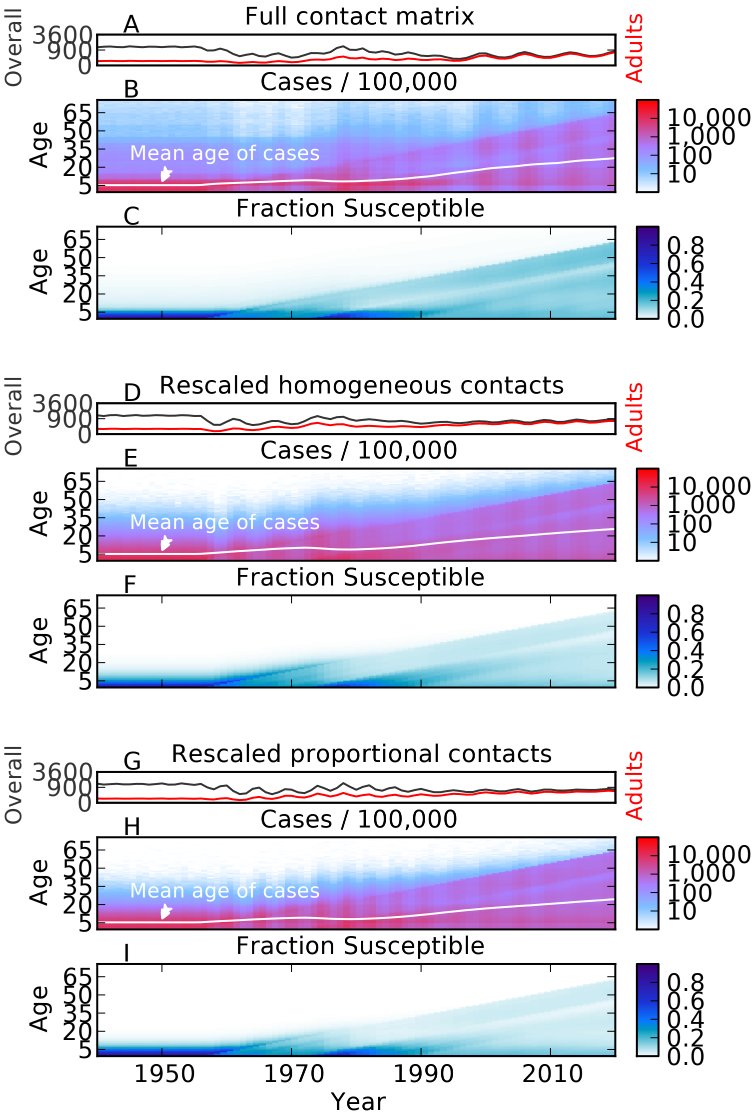


Figure A.4: Single realizations under each contact structure with homogeneous and proportional contact rates have been scaled by the ratio of the overall rate of contacts observed to the contact rate theoretically necessary to produce a mean age of infection of 5 years in a well mixed population (given our other model's 15 day mean duration of infectiousness, 75 year average lifespan, and 4% chance of transmission upon a risky contact) (Anderson and May, 1982) are plotted as in Figure A.3. All three realizations use values of 60% vaccination in 1957 and 85% vaccine efficacy.

Results with these diffrent ramp-ups are plotted in Figure A.5 As one would expect, incidences differ early in the vaccine era when incidence is declining rapidly. Higher initial uptake values lead to a faster descent with more oscillation in incidence. However, the gradual rise in adult and adolescent cases seems to be qualitatively robust to the initial uptake value.

A.3 Vaccine Efficacy

Vaccine efficacy is difficult to estimate for pertussis and may vary between vaccines.(Mills et al., 1998) We carried out simulations with efficacies of 75%, 85%, and 95%. Results with lifelong natural immunity and varying durations of vaccine derived immunity are plotted in Figure A.6. We see the gradual rise in adult cases in all cases, except for for vaccines with 95% efficacy, where lifelong immunity resulted in eradication. However, any waning of vaccine derived immunity again produced a dramatic rise in adult and adolescent cases as cohorts born in the early vaccine-era aged.

Both primary vaccine failure and waning of immunity, alone or in combination, produce the gradual increase in adult cases in our simulation results. Figure A.7 shows the age-structured incidence rates from runs with lifelong natural immunity and either primary vaccine failure (85% efficacy), vaccine waning (70 year mean duration), neither, or both. With a 95% efficacious, lifelong vaccine, eradication is acheived after about ten years of sustained, high coverage vaccination. However, in all other cases, we see the same pattern of a long honeymoon followed by a resurgence in adolscents and adults.

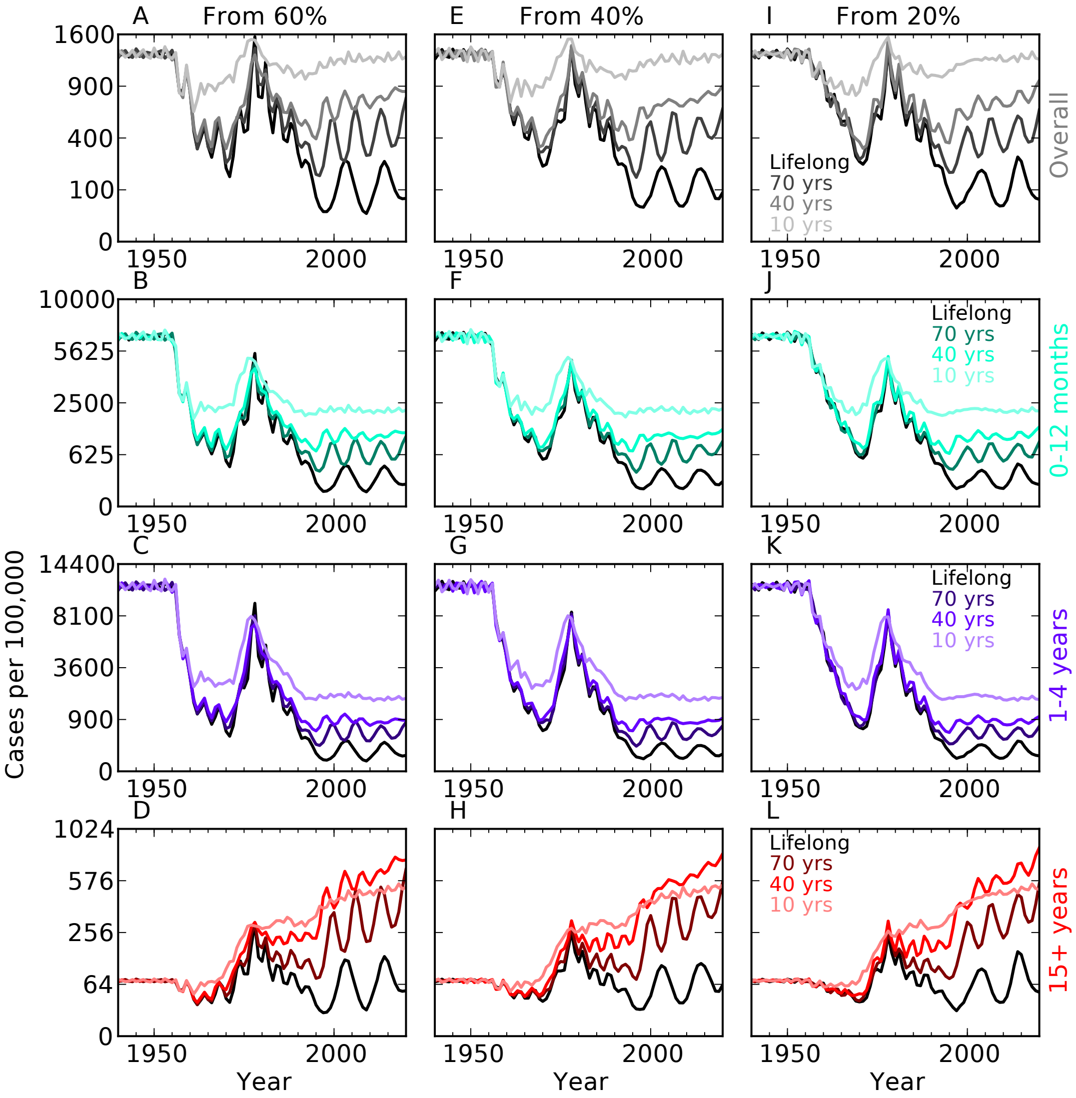


Figure A.5: Overall (top row) and adult (bottom row) annual incidences for varying vaccine the vaccine uptake in 1957 (column) and duration of vaccine derived immunity (color). In all these simulatons, vaccine efficacy was set to 85%.

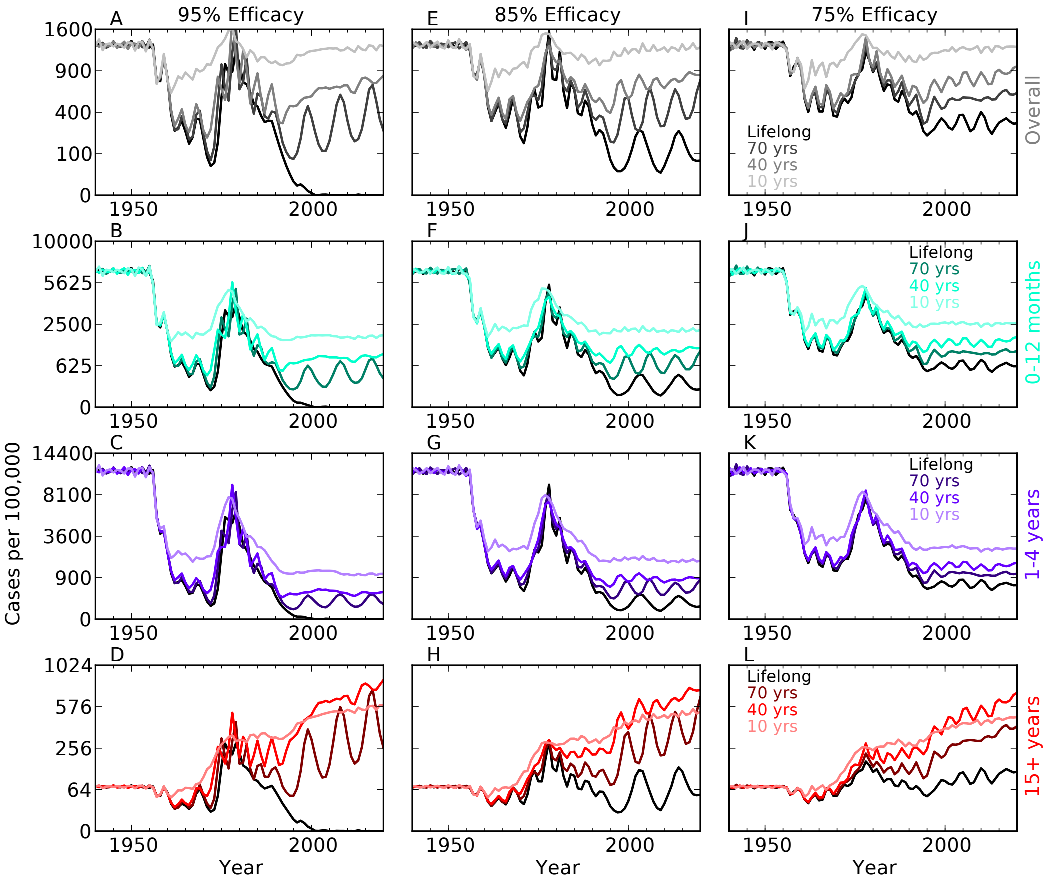


Figure A.6: Overall (top row) and adult (bottom row) annual incidences for varying vaccine efficacy (column) and duration of vaccine derived immunity (color). In these simulations, vaccine coverage is set to 60% in 1957.

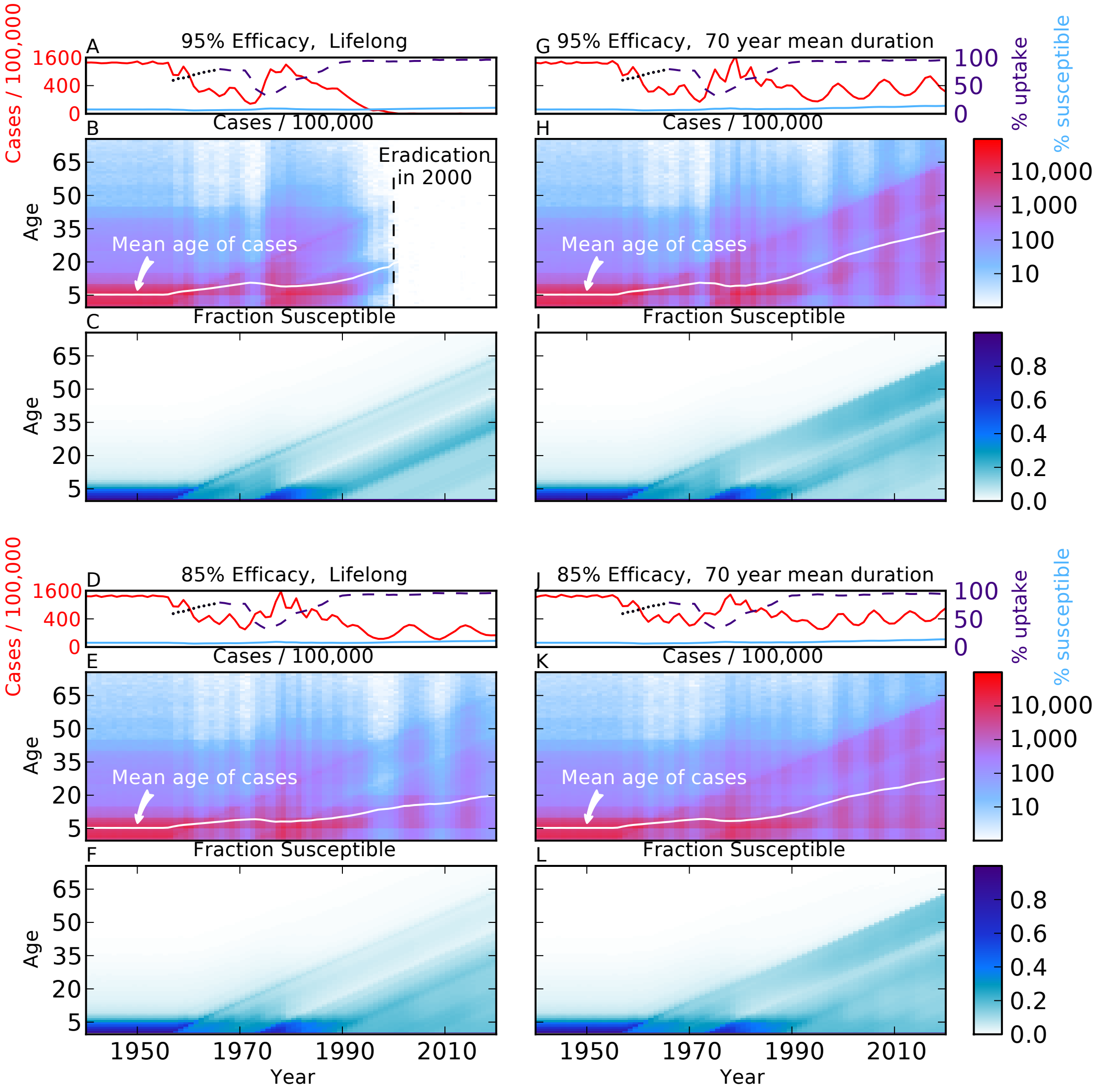


Figure A.7: Single realizations with 60% vaccination in 1957, 95% (top row) or 85% efficacy (bottom row), and lifelong (left column) or 70 year (right column) vaccine derived immunity. For each realization the incidence by age (top panel); overall incidence, vaccine uptake, and fraction susceptible (middle panel); and fraction susceptible by age (bottom panel) are plotted against time.

A.4 Additional Realizations

For clarity, Figure 2.3 was only plotted with one realization of the model. Figure A.8 shows the same figure with the omitted 19 realizations plotted as thin, gray lines. Although some variation is present, as one would expect with a stocastic model, the interepidemic period and trends in incidence seem to be reasonably consistent between runs.

A.5 Immune Boosting

There is growing interest in the possible impacts of immune boosting on pertussis transmission dynamics. (Águas, 2006; Lavine et al., 2011) We we implemented a version of immune boosting as modeled in Ref. (Lavine et al., 2011), in which some fraction of contacts between infectious and immune individuals lead to a boosting of immunity. In terms of our compartmental model, this means that individuals in the R_i' and V_i' classes move to R_i at a rate proportional to the force of infection on age group i. We simulated transmission dynamics in cases where 2%, 4%, and 80% of contacts between immune and infected individuals lead to boosting (equivalent to $\kappa = 0.5, 1, 20$ in the notation of Ref. (Lavine et al., 2011)). Three realizations are plotted in Figure A.9. When immune boosting is common enough to generate a realistically low pre-vaccine age of infection, we see that the same qualitative pattern of a gradual increase in adult cases during the vaccine era, particularly emerges.

A.6 Alternative demography and distribution of immune durations

In the classic $SEIR$ model, lifespans and immunity are both exponentially distributed, while our model has fixed lifespans and gamma distributed immune dura-

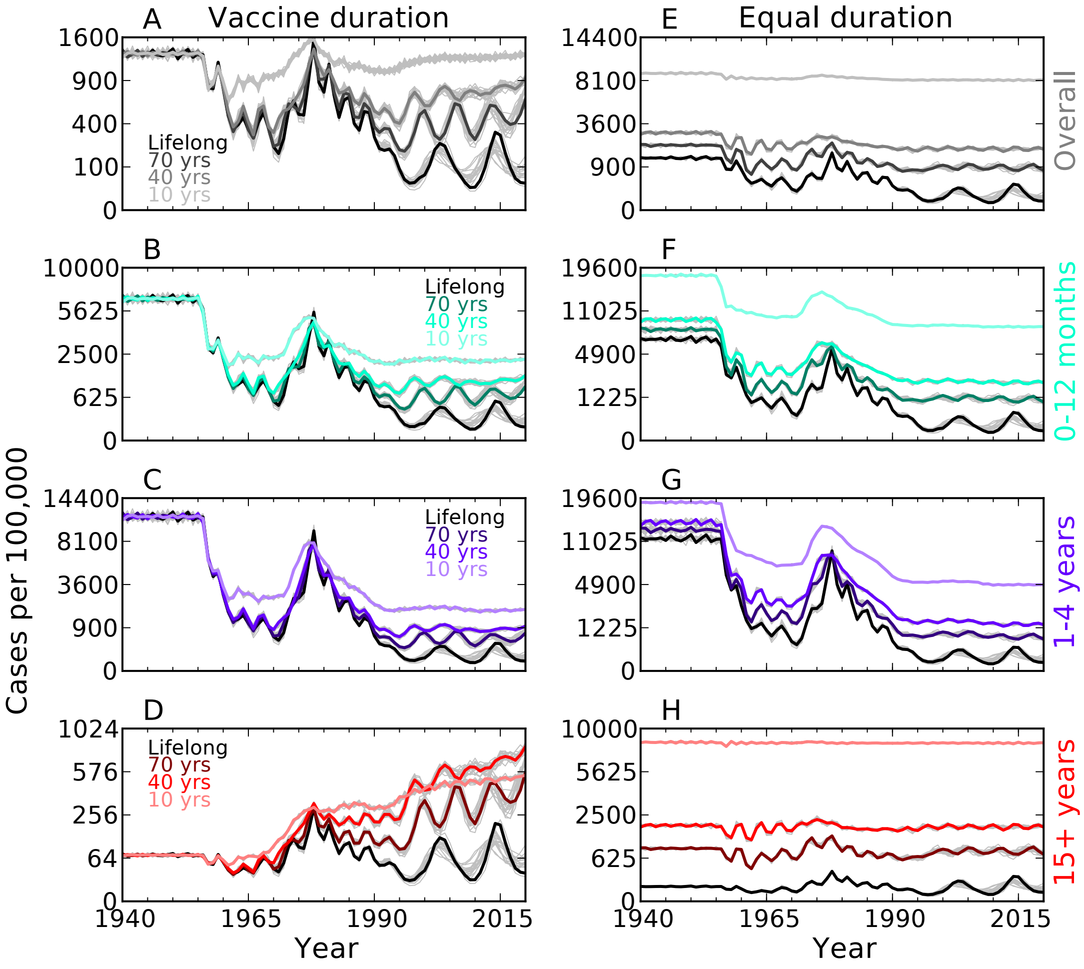


Figure A.8: Annual incidence overall, in infants, in toddlers, and in adults (top to bottom rows) for varying durations of immunity (indicated by color). Twenty realizations of each case are plotted (thin gray lines), with one run highlighted in color for clarity. In the left column, natural immunity is lifelong while in the right column natural and vaccine derived immunity have equal durations. The simulations plotted used a vaccine efficacy of 85% and an initial vaccine uptake of 60% in 1957.

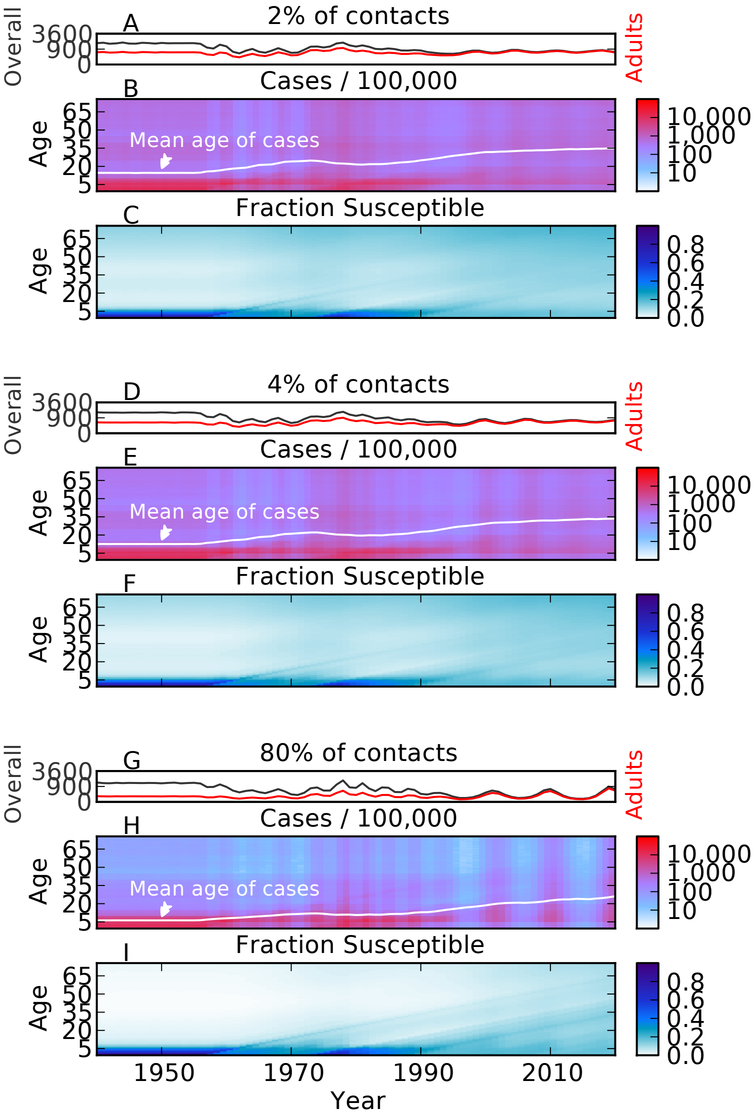


Figure A.9: Single realizations of in which 2% (panels A-C), 4% (panels D-F), and 80% (panels G-I) of contacts between immune and infectious individuals lead to boosting of immunity. The three realizations plotted use a 70 year mean duration of natural and vaccine derived immunuity, 60% vaccination in 1957 and 85% vaccine efficacy. For each realization the overall incidence and incidence in individuals over 15 year old (panels A, D, G); the incidence by age (panels B, E, H); and the fraction susceptible by age (panels C, F, I) are plotted against time.

tions. Both of these sets of assumptions are extreme simplifications of the biological reality and it is worthwhile to check that our conclusions are not sensitive to modeling assumptions about lifespan, aging, and immune waning.

We implemented an alternate model with continuous aging and death among adults and the age structure used in Ref. (Rohani et al., 2010). This alternative model uses the same age categories and annual aging for individuals under twenty years old, 0–5 months old, 6 months – 1 year old, 1–2 years old, . . . , 19–20 years old, and five year age bands with continuous aging thereafter, 20–25 years old, ... , 65–70 years old, and over 70 years old. Individuals over age twenty age at a constant rate of $\frac{1}{5}$ years^{-1} and have a constant death rate of $\frac{1}{55}$ years^{-1}, giving the same average lifespan of 75 years as in our main model.

In this alternate model, we also examined the case in which the duration of immunity is exponentially rather than gamma distributed by using one, rather than two, compartments for each immune class.

The results using 85% vaccine efficacy and 60% vaccine uptake in 1957 with the original model (A), the model with exponential lifespans and aging (B), and the model with expoenential lifespans, aging, and immune durations (C), are plotted in Figure A.10. The overall qualitative picture is very similar between the three models, even though exponential aging among adults blurs the strong cohort effects seen in the original model.

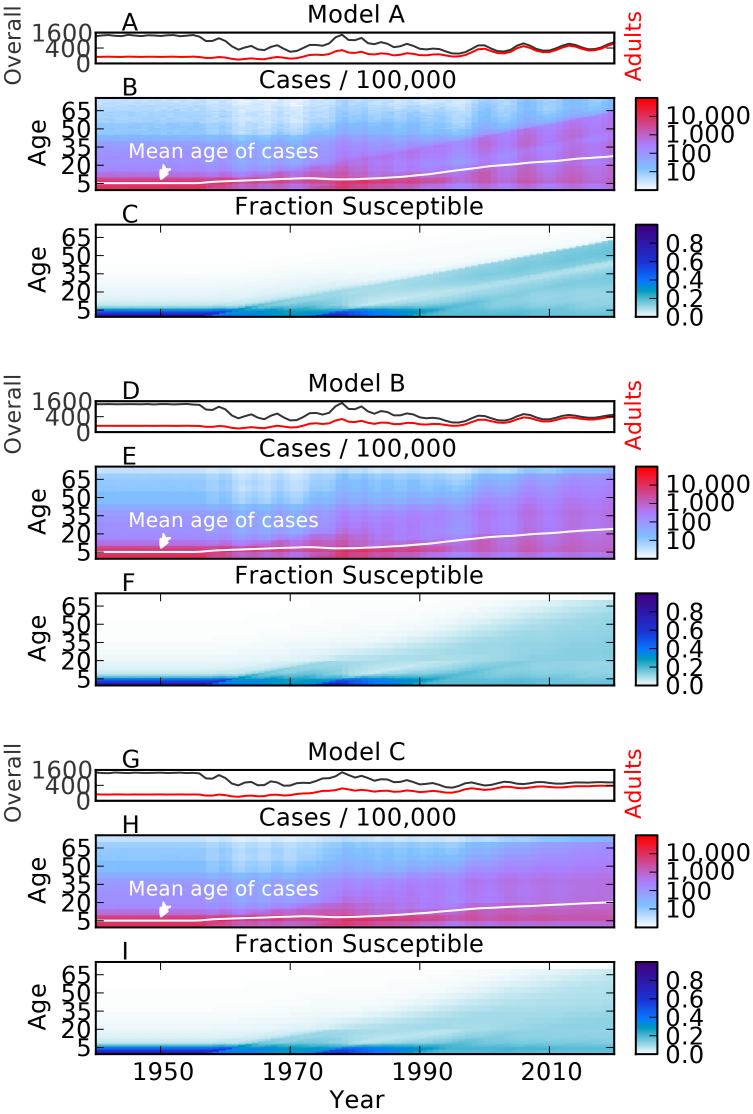


Figure A.10: Single realizations of models A (annual aging and gamma distributed immune durations), B (exponential ligespans and aging and gamma distributed immune durations), and C (exponential lifespans, aging, and immune durations) with 60% vaccination in 1957 and 85% vaccine efficacy. For each realization the overall incidence and incidence in individuals over 15 year old (panels A, D, G); the incidence by age (panels B, E, H); and the fraction susceptible by age (panels C, F, I) are plotted against time.

APPENDIX B

Evolving booster schedules supplementary information

B.1 Details of fitness calculations

For simplicity, we calculated strategy fitness as a linear combination of disease burden (measured in DALYs) and the monetary cost of vaccinations.

Although our transmission model does not account for mortality, for the purposes of assessing costs we assume a case fatality rate of 0.2% in infants one year or younger, of 0.04% in children 1-4 years of ageCrowcroft et al. (2003), and no mortality in older age groups. In all age groups, we assume that both symptoms and infectiousness last, on average, for 15 days. Thus, under our parameterization, a case in a six month old costs $\frac{15}{365} + 0.002 * 74.5 \approx 0.19$ DALYs, while a case in an adult costs only $\frac{15}{365} \approx 0.041$ DALYs.

Throughout the study, we used a cost of \$33 per dose of vaccine and \$50000 per DALY. However, it is worth noting that in our current model changing the cost per DALY is equivalent to changing the cost of vaccine (e.g. an assumption of \$66 per dose of vaccine and \$100000 per DALY would produce identical results).

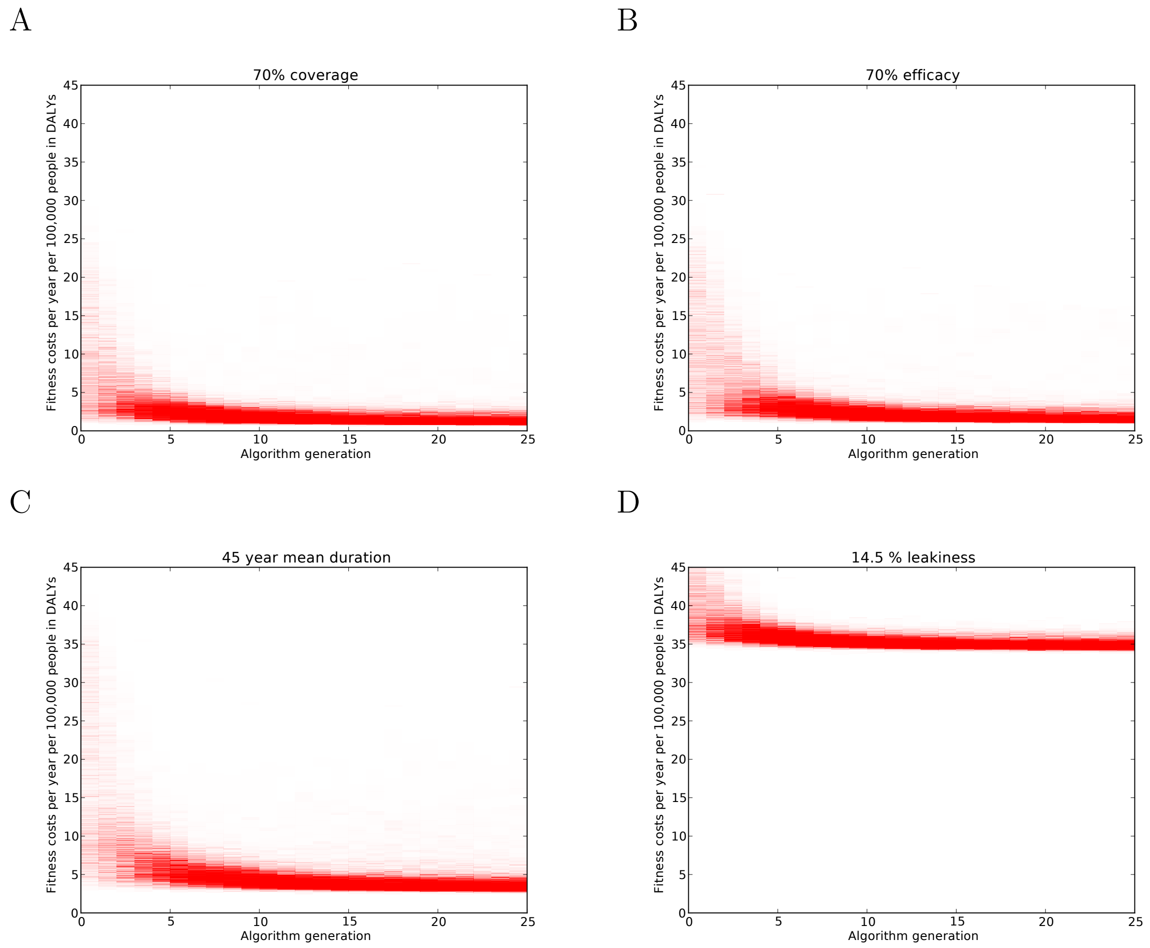


Figure B.1: Fitness costs by algorithm generation. Density maps of the fitness costs (combined vaccination effort and disease burden) of the 2000 strategies in each generation of the algorithm for (A) Scenario I: 70% infant coverage, (B) Scenario II: 70% vaccine efficacy (C) Scenario III: 45 year mean duration of vaccine derived immunity in (G,H,I), and (D) Scenario IV: 14.5% leakiness. In each case, darker color indicates a higher density of strategies.

B.2 Supplementary Genetic Algorithm results

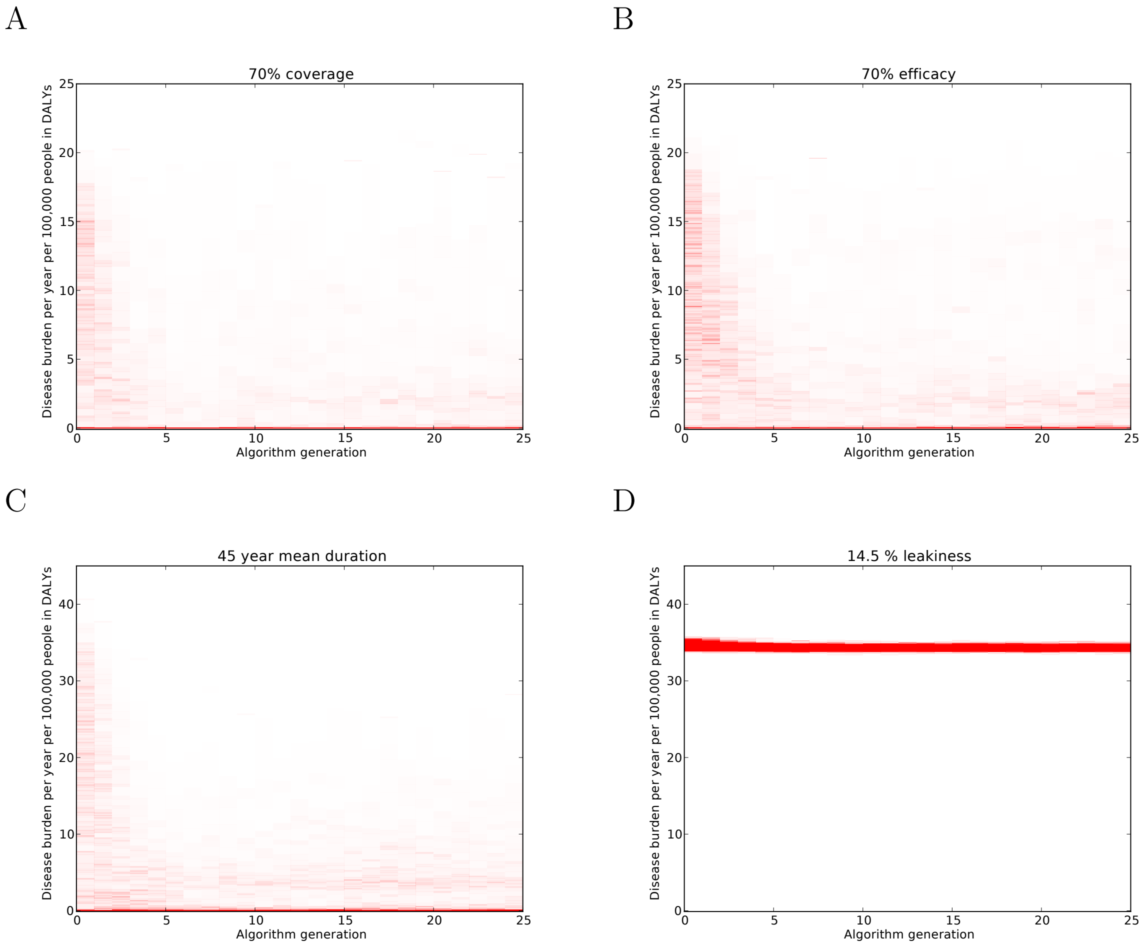


Figure B.2: Disease burden (measured in DALYs) by algorithm generation. Density maps of the fitness costs (combined vaccination effort and disease burden) of the 2000 strategies in each generation of the algorithm for (A) Scenario I: 70% infant coverage, (B) Scenario II: 70% vaccine efficacy (C) Scenario III: 45 year mean duration of vaccine derived immunity in (G,H,I), and (D) Scenario IV: 14.5% leakiness. In each case, darker color indicates a higher density of strategies.

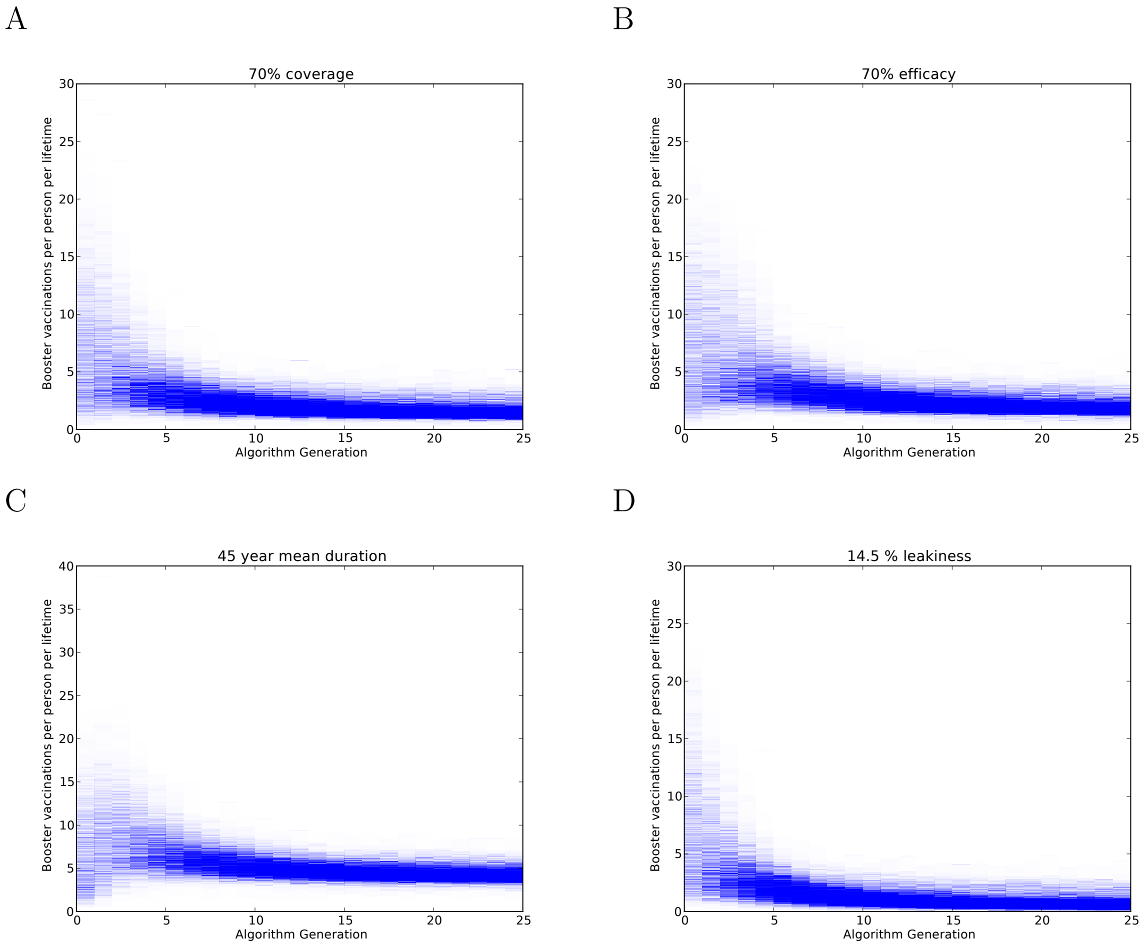


Figure B.3: Doses of vaccine given per person per lifetime by algorithm generation. Density maps of the fitness costs (combined vaccination effort and disease burden) of the 2000 strategies in each generation of the algorithm for (A) Scenario I: 70% infant coverage, (B) Scenario II: 70% vaccine efficacy (C) Scenario III: 45 year mean duration of vaccine derived immunity in (G,H,I), and (D) Scenario IV: 14.5% leakiness. In each case, darker color indicates a higher density of strategies.

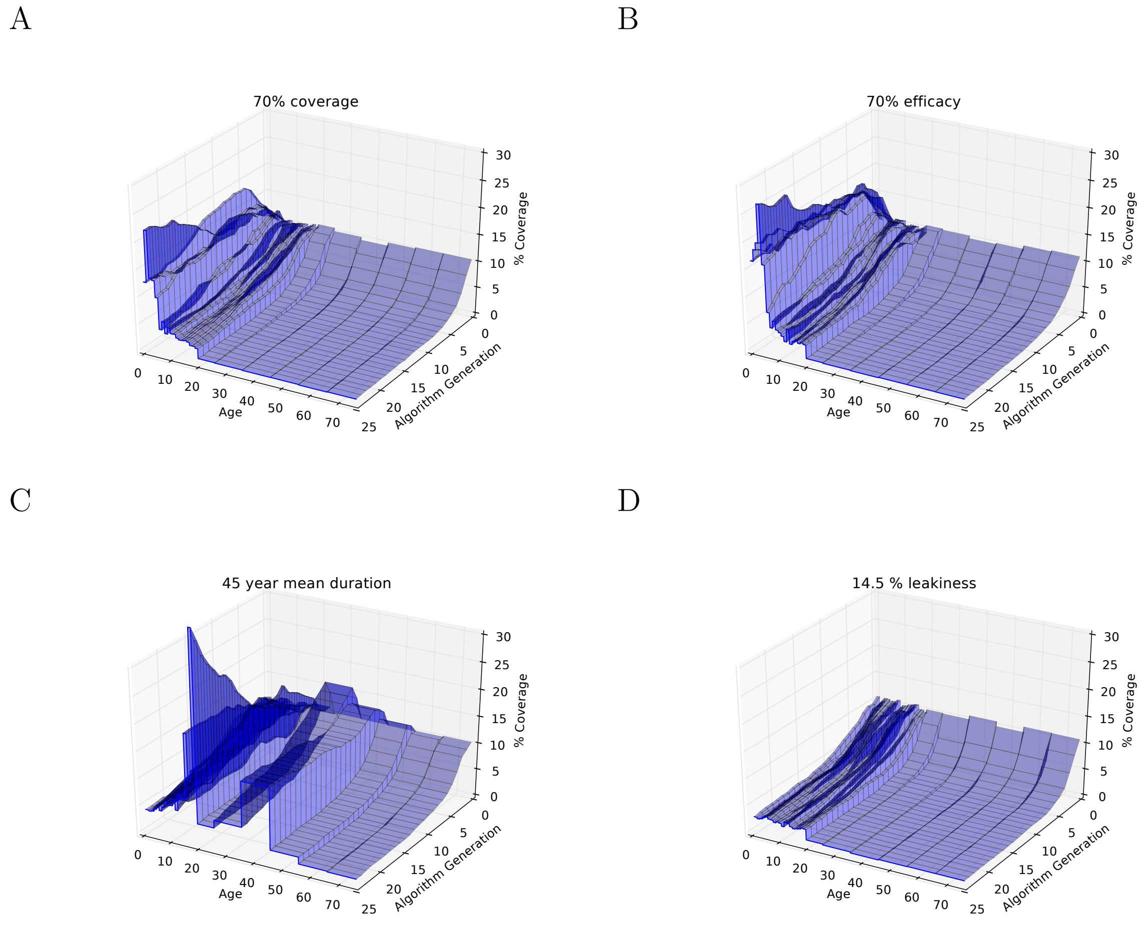


Figure B.4: Mean strategy by algorithm generation. Surfaces plot showing the average schedule during each algorithm generation (constructed by taking the mean coverage in each age cohort) of the 2000 strategies in each generation of the algorithm for (A) Scenario I: 70% infant coverage, (B) Scenario II: 70% vaccine efficacy (C) Scenario III: 45 year mean duration of vaccine derived immunity in (G,H,I), and (D) Scenario IV: 14.5% leakiness. In each case, darker color indicates a higher density of strategies.

APPENDIX C

Host-parasitoid supplementary information

Click for video of half-and-half landscape

Figure C.1: Video of a single realization of 1000 generations of host and parasitoid densities on a half-and-half landscape ($\lambda_+ = 2.8$, $\lambda_- = 1.2$). Densities on each plant are plotted in the left panel, with higher host densities indicated by more blue and higher parasitoid densities by more red (see colormap in central panel. Local populations of hosts (in blue) and parasitoids (in red) along a transect (corresponding to row 30 of the plot at the left) are plotted in the right panel.

Click for video of checkerboard landscape

Figure C.2: Video of a single realization of 1000 generations of host and parasitoid densities on a checkerboard landscape ($\lambda_+ = 2.8$, $\lambda_- = 1.2$). Densities on each plant are plotted in the left panel, with higher host densities indicated by more blue and higher parasitoid densities by more red (see colormap in central panel. Local populations of hosts (in blue) and parasitoids (in red) along a transect (corresponding to row 30 of the plot at the left) are plotted in the right panel.

BIBLIOGRAPHY

BIBLIOGRAPHY

Águas, R. 2006. Pertussis: increasing disease as a consequence of reducing transmission. Lancet Infect Dis 6:112–7.

Alpert, C. J., and S.-Z. Yao. 1995. Spectral partitioning: the more eigenvectors, the better. In Proceedings of the 32nd annual ACM/IEEE Design Automation Conference, pages 195–200. ACM.

Andersen, M. 1991. Properties of some density-dependent integrodifference equation population models. Mathematical Biosciences 104:135–157. doi: 10.1016/0025-5564(91)90034-G.

Anderson, R., and R. May. 1982. Directly transmitted infectious diseases: Control by vaccination. Science 215:1053–1060.

Andow, D. A. 1991. Vegetational diversity and arthropod population response. Annual review of entomology 36:561–586.

Bach, F. R., and M. I. Jordan. 2006. Learning spectral clustering, with application to speech separation. The Journal of Machine Learning Research 7:1963–2001.

Batch, C. E. 1984. Plant spatial pattern and herbivore population dynamics: Plant factors affecting the movement patterns of a tropical cucurbit specialist (*Acalymma Innubum*). Ecology 65:175–190.

Belyea, H. C. 1923. The control of white pine weevil (*Pissodes strobi*) by mixed planting. Journal of Forestry 21:384–390.

Bickerton, M. W. 2011. Intercropping for conservation biological control of European corn borer Ostrinia nubilalis Hübner (lepidoptera: crambidae) in bell peppers. Ph.D. thesis, Rutgers University-Graduate School-New Brunswick.

Bjørnstad, O. N., and T. F. Hansen. 1994. Individual variation and population dynamics. Oikos pages 167–171.

Bjørnstad, O. N., M. Peltonen, A. M. Liebhold, and W. Baltensweiler. 2002. Waves of larch budmoth outbreaks in the european alps. Science 298:1020–1023.

Blackwood, J. C., D. A. Cummings, H. Broutin, S. Iamsirithaworn, and P. Rohani. 2013. Deciphering the impacts of vaccination and immunity on pertussis epidemiology in thailand. Proceedings of the National Academy of Sciences 110:9595–9600.

Blennow, M., P. Olin, M. Granström, and R. Bernier. 1988. Protective efficacy of a whole cell pertussis vaccine. BMJ 296:1570–1572.

Briggs, C. J., and J. Latto. 2000. The effect of dispersal on the population dynamics of a gall-forming midge and its parasitoids. Journal of Animal Ecology 69:96–105.

Broutin, H., C. Viboud, B. T. Grenfell, M. A. Miller, and P. Rohani. 2010. Impact of vaccination and birth rate on the epidemiology of pertussis: a comparative study in 64 countries. Proceedings of the Royal Society B: Biological Sciences pages 1–7.

Cantrell, R. S., and C. Cosner. 1991. The effects of spatial heterogeneity in population dynamics. Jounal of Mathematical Biology 29:315–338.

Cappuccino, N., D. Lavertu, Y. Bergeron, and J. Régnière. 1998. Spruce budworm impact, abundance, and parasitism rate in a patchy landscape. Oecologica 114:236–242.

Celentano, L., M. Massari, D. Paramatti, S. Salamaso, and A. Tozzi. 2005. Resurgence of pertussis in europe. Pediatr Infect Dis J 24:761–756.

Chen, B., J. Wang, L. Zhang, Z. Li, and G. Xiao. 2011. Effect of intercropping pepper with sugarcane on populations of *Liriomyza huidobrensis* (diptera: Agromyzidae) and its parasitoids. Crop Protection 30:253–258.

Cherry, J. D. 2003. The science and fiction of the resurgence of pertussis. Pediatrics 112:405–406.

———. 2012. Epidemic pertussis in 2012the resurgence of a vaccine-preventable disease. New England Journal of Medicine 367:785–787.

Comins, H. N., M. P. Hassell, and R. M. May. 1992. The spatial dynamics of host–parasitoid systems. Journal of Animal Ecology 61:735–748.

Condon, A., and R. M. Karp. 2001. Algorithms for graph partitioning on the planted partition model. Random Structures and Algorithms 18:116–140.

Cronin, J. T. 2003. Patch structure, oviposition behavior, and the distribution of parasitism risk. Ecological Monographs 73:283–300.

Cronin, J. T., and J. D. Reeve. 2005. Host-parasitoid spatial ecology: a plea for landscape-level synthesis. Proceedings of the Royal Society B: Biological Sciences 272:2225–2235.

Crowcroft, N., and R. Pebody. 2006. Recent developments in pertussis. Lancet 367:1926–1936.

Crowcroft, N., C. Stein, P. Duclos, and M. Birmingham. 2003. How best to estimate the global burden of pertussis? The Lancet infectious diseases 3:413–418.

Cushing, J., B. Dennis, R. Desharnais, and R. Constantino. 1998. Moving toward an unstable equilibrium: saddle nodes in population systems. Journal of Animal Ecology 67:298–306.

Davidor, Y. 1990. Epistasis variance: Suitability of a representation to genetic algorithms. Complex Systems 4:369–383.

de Greeff, S. C., F. R. Mooi, A. Westerhof, J. Verbakel, M. F. Peeters, C. Heuvelman, D. W. Notermans, L. Elvers, J. F. Schellekens, and H. E. de Melker. 2010. Pertussis disease burden in the household: how to protect young infants. Clinical Infectious Diseases 50:1339–1345.

Denno, R. F., and M. S. McClure. 1983. Variability: a key to understanding plantherbivore interactions. In R. F. Denno, and M. S. McClure, eds., Variable plants and herbivores in natural and managed systems, pages 1–12. Academic Press, New York, New York, USA.

Eames, K., N. Tilston, E. Brooks-Pollock, and W. Edmunds. 2012. Measured dynamic social contact patterns explain the spread of h1n1v influenza. PLoS Comput Biol 8:e1002425. Doi:10.1371/journal.pcbi.1002425.

Elton, C. S. 1958. The ecology of invasions by animals and plants. London: Meuthen & Co., Ltd.

Eubank, S., H. Guclu, V. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. 2004. Modelling disease outbreaks in realistic urban social networks. Nature 429:180–184.

Evans, D. M., N. E. Turley, D. J. Levey, and J. J. Tewksbury. 2012*a*. Habitat patch shape, not corridors, determines herbivory and fruit production of an annual plant. Ecology 93:1016–1025.

Evans, L. M., J. S. Clark, A. V. Whipple, and T. G. Whitham. 2012*b*. The relative influences of host plant genotype and yearly abiotic variability in determining herbivore abundance. Oecologica 168:483–489.

Farizo, K. M., D. L. Burns, T. M. Finn, M. F. Gruber, and R. D. Pratt. 2014. Clinical evaluation of pertussis vaccines: Us food and drug administration regulatory considerations. Journal of Infectious Diseases 209:S28–S31.

Fiedler, M. 1973. Algebraic connectivity of graphs. Czechoslovak Mathematical Journal 23:298–305.

Fischbein, D., J. Bettinelli, C. Bernstein, and J. C. Corley. 2012. Patch choice from a distance and use of habitat information during foraging by the parasitoid *Ibalia leucospoides*. Ecological Entomology 37:161–168.

Foster, M. A., J. C. Schultz, and M. D. Hunter. 1992. Modelling gypsy moth–virus–leaf chemistry interactions: Implications of plant quality for pest and pathogen dynamics. Journal of Animal Ecology 61:509–520.

Garey, M. R., and D. S. Johnson. 1979. Computer and intractability.

Goldberg, D. E., and K. Deb. 1991. A comparative analysis of selection schemes used in genetic algorithms. Urbana 51:69–92.

Gower, J. C., and G. B. Dijksterhuis. 2004. Procrustes problems, volume 3. Oxford University Press Oxford.

Greco, D., S. Salmaso, P. Mastrantonio, et al. 1996. A controlled trial of two acellular vaccines and one whole-cell vaccine against pertussis. N Engl J Med 334:341–349.

Gulland, A. 2012. Teenagers and newborn babies in england and wales may get pertussis jab as cases continue to rise. BMJ 345:e5919 doi: 10.1136/bmj.e5919.

Güriş, D., P. M. Strebel, B. Bardenheier, M. Brennan, R. Tachdjian, E. Finch, M. Wharton, and J. R. Livengood. 1999. Changing epidemiology of pertussis in the united states: increasing repoted incidence among adolescents and adults, 1990-1996. Clin Infect Dis. 28:1230–1237.

Halloran, M. E., M. Haber, and I. M. Longini. 1992. Interpretation and estimation of vaccine efficacy under heterogeneity. American Journal of Epidemiology 136:328–343.

Halperin, S. A., R. Bortolussi, D. MacLean, and N. Chisholm. 1989. Persistence of pertussis in an immunized population: results of the nova scotia enhanced pertussis surveillance program. The Journal of pediatrics 115:686–693.

Hanski, I., J. Pöyry, T. Pakkala, and M. Kuussaari. 1995. Multiple equilibria in metapopulation dynamics. Nature 377:618—-621.

Hardin, D., P. Takac, and G. Webb. 1988. Asymptotic properties of a continuous-space discrete-time population model in a random environment. J Math Biol 26:361–374.

Hassell, M. P. 2000. The Spatial and Temporal Dynamics of Host–Parasitoid Interactions. Oxford University Press, Oxford.

Hassell, M. P., and R. M. May. 1973. Stability in insect host-parasite models. The Journal of Animal Ecology 43:693–726.

Hassell, M. P., H. N. Comins, and R. P. May. 1991. Spatial structure and chaos in insect population dynamics. Nature 353:255–258.

Hastings, A., and K. Higgins. 1994. Persistence of Transients in Spatially Structured Ecological Models. Science 263:1133–1136.

Helms, S. E., and M. D. Hunter. 2005. Variation in plant quality and the population dynamics of herbivores: there is nothing average about aphids. Oecologica 145:197–204.

Hill, J. K., C. D. Thomas, and O. T. Lewis. 1996. Effects of habitat patch size and isolation on dispersal by *Hesperia comma* butterflies: implications for metapopulation structure. Journal of Animal Ecology 65:725–735.

Hirzel, A. H., R. M. Nisbet, and W. W. Murdoch. 2007. Host-parasitoid spatial dynamics in heterogeneous landscapes. Oikos 116:2082–2096.

Holland, J. H. 1992. Genetic algorithms. Scientific american 267:66–72.

Holt, R. D., and M. Barfield. 2003. Impacts of temporal variation on apparent competition and coexistence in open ecosystems. Oikos 101:49–58.

Huffaker, C. B. 1958. Experimental studies on predation: Dispersion factors and predator–prey oscillations. Hilgardia 27:343–383.

Hunter, M. D. 2003. Effects of plant quality on the population ecology of parasitoids. Agricultural and Forest Entomology 5:1–8.

Hunter, M. D., and P. W. Price. 1992. Playing chutes and ladders: heterogeneity and the relative roles of bottom-up and top-down forces in natural communities. Ecology 73:724–732.

Hunter, M. D., S. B. Malcolm, and S. E. Hartley. 1996. Population–level variation in plant secondary characteristics, and the population biology of herbivores. Chemoecology 7:45–56.

Hunter, M. D., G. C. Varley, and G. R. Gradwell. 1997. Estimating the relative roles of top-down and bottom-up forces on insect herbivore populations: A classic study revisited. PNAS 94:9176–9181.

Jackson, D., and P. Rohani. 2013. Perplexities of pertussis: recent global epidemiological trends and their potential causes. Epidemiology and infection 142:1–13.

Jenkinson, D. 1988. Duration of effectiveness of pertussis vaccine: evidence from a 10 year community study. British medical journal (Clinical research ed.) 296:612–614.

Johst, K., R. Brandl, and S. Eber. 2002. Metapopulation persistence in dynamic landscapes: the role of dispersal distance. Oikos 98:263–270.

Jones, T., and S. Forrest. 1995. Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In ICGA, volume 95, pages 184–192. Citeseer.

Jury, E. I. 1964. Theory and Application of the z-Transform Method, volume 3. Wiley New York.

———. 1974. Inners and stability of dynamic systems.

Kannan, R., S. Vempala, and A. Vetta. 2004. On clusterings: Good, bad and spectral. Journal of the ACM (JACM) 51:497–515.

Kot, M. 1989. Diffusion-driven period-doubling bifurcations. Biosystems 22:279–287. doi: 10.1016/0303-2647(89)90049-X.

Kot, M., and W. M. Schaffer. 1986. Discrete-time growth-dispersal models. Mathematical Biosciences 80:109–136. doi: 10.1016/0025-5564(86)90069-6.

Lavine, J., A. King, and O. Bjørnstad. 2011. Natural immune boosting in pertussis dynamics and the potential for long-term vaccine failure. Proc Natl Acad Sci U S A 108:7259–7264. Doi: 10.1073/pnas.1014394108.

Lee, J. R., S. Oveis Gharan, and L. Trevisan. 2012. Multi-way spectral partitioning and higher-order cheeger inequalities. In Proceedings of the 44th symposium on Theory of Computing, pages 1117–1130. ACM.

Levin, S. 1976. Population dynamic-models in heterogeneous environments. Annual review of ecology and systematics 7:287–310. doi: {10.1146/annurev.es.07.110176.001443}.

———. 1992. The problem of pattern and scale in ecology. Ecology 73:1943–1967. doi: {10.2307/1941447}.

Loveless, M. D., and J. L. Hamrick. 1984. Ecological determinants of genetic structure in plant populations. Annual review of ecology and systematics 15:65–95.

Lutscher, F., and N. Van Minh. 2013. Traveling waves in discrete models of biological populations with sessile stages. Nonlinear Analysis: Real World Applications 14:495–506. doi: 10.1016/j.nonrwa.2012.07.011.

McIntyre, P. J., and T. G. Whitham. 2003. Plant genotype affects long–term herbivore population dynamics and extinction: Conservation implications. Ecology 84:311–322. doi: {10.1890/0012-9658(2003)084{[}0311:PGALTH]2.0.CO;2}.

McLean, A. R., and R. M. Anderson. 1988. Measles in developing countries. part ii. the predicted impact of mass vaccination. Epidemiology and Infection 100:419–442.

Meilă, M., and J. Shi. 1995. Learning segmentation by random walks. In T. Leen, T. Dietterich, and V. Tresp, eds., Proceedings of the 2000 Conference on Advances in Neural Information Processing Systems, pages 195–200. MIT Press, Cambridge, MA.

Meilă, M., and L. Xu. 2003. Multiway cuts and spectral clustering. Technical report, University of Washington, Department of Statistics.

Mertsola, J., O. Ruuskanen, E. Eerola, and M. Viljanen. 1983. Intrafamilial spread of pertussis. J Pediatr 103:359–63.

Meyers, L. A., B. Pourbohloul, M. E. Newman, D. M. Skowronski, and R. C. Brunham. 2005. Network theory and sars: predicting outbreak diversity. Journal of theoretical biology 232:71–81.

Mills, K. 2001. Immunity to *Bordetella pertussis*. Microbes and infection 3:655–677.

Mills, K., M. Ryan, E. Ryan, and B. Mahon. 1998. A murine model in which protection correlates with pertussis vaccine efficacy in children reveals complimentary roles for humoral and cell-mediated immunity in protection against *Bordatella pertussis*. Infect Immun 66:594–602.

Mooi, F. R., I. Van Loo, and A. J. King. 2001. Adaptation of bordetella pertussis to vaccination: a cause for its reemergence? Emerging infectious diseases 7:526–528.

Mossong, J., N. Hens, M. Jit, et al. 2008. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med 5:e74.

Murdoch, W. W., F. C. Evans, and C. H. Peterson. 1972. Diversity and pattern in plants and insects. Ecology 53:819–829. doi: {10.2307/1934297}.

Neubert, M., M. Kot, and M. Lewis. 1995. Dispersal and Pattern Formation in a Discrete-Time Predator-Prey Model. Theoretical Population Biology 48:7–43. doi: 10.1006/tpbi.1995.1020.

Newman, M. E. 2002. Spread of epidemic disease on networks. Physical review E 66:016128.

Ng, A. Y., M. I. Jordan, and Y. Weiss. 2001. On spectral clustering: Analysis and an algorithm. Proceedings of Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press 14:849–856.

Nicholson, A. J., and V. A. Bailey. 1935. The balance of animal populations.part i. Proceedings of the Zoological Society of London 105:551–598.

Nielsen, A., and S. O. Larsen. 1994. Epidemiology of pertussis in denmark: the impact of herd immunity. International journal of epidemiology 23:1300–1308.

Ntezayabo, B., G. De Serres, and B. Duval. 2003. Pertussis resurgence in canada largely caused by a cohort effect. The Pediatric infectious disease journal 22:22–27.

Oksanen, T., L. Oksanen, and M. Gyllenberg. 1992. Exploitation ecosystems in heterogeneous habitat complexes ii: impact of small–scale heterogeneity on predator–prey dynamics. Evol. Ecol. 6:383–398.

Omer, S. B., D. A. Salmon, W. A. Orenstein, M. P. deHart, and N. Halsey. 2009. Vaccine refusal, mandatory immunization, and the risks of vaccine-preventable diseases. New England Journal of Medicine 360:1981–1988.

Papadimitriou, C. H., and K. Steiglitz. 1998. Combinatorial optimization: algorithms and complexity. Courier Dover Publications.

Pennings, S. C., M. Zimmer, N. Dias, M. Sprung, N. Davé, C.-K. Ho, A. Kunza, C. McFarlin, M. Mews, A. Pfauder, and C. Salgado. 2007. Latitudinal variation in plant–herbivore interactions in european salt marshes. Oikos 116:543–549.

Pettrgrew, M., and H. Resat. 2007. Multinomial tau-leaping method for stochastic kinetic simulations. J Chem Phys 12:084101.

Pimentel, D. 1961. Species diversity and insect population outbreaks. Annals of the Entomological Society of America 54:76–86.

Pothen, A., H. D. Simon, and K.-P. Liou. 1990. Partitioning sparse matrices with eigenvectors of graphs. SIAM Journal on Matrix Analysis and Applications 11:430–452.

Pulliam, H. R. 1988. Sources, sinks, and population regulation. The American Naturalist 132:652–661.

Quinn, H., and P. McIntyre. 2007. Pertussis epidemiology in austrailia over the decade 1995-2005: trends by region and age group. Commun Dis Intell 31:205–215.

Rand, T. A., and S. M. Louda. 2006. Spillover of agriculturally subsidized predators as a potential threat to native insect herbivores in fragmented landscapes. Conservation Biology 20:1720–1729.

Reeve, J. D. 1988. Environmental variability, migration, and persistence in host-parasitoid systems. American Naturalist 132:810–836.

Riolo, M. A., and M. E. J. Newman. 2012. First-principles multiway spectral partitioning of graphs. CoRR abs/1209.5969.

Riolo, M. A., A. A. King, and P. Rohani. 2013. Can vaccine legacy explain the british pertussis resurgence? Vaccine 31:5903–5908.

Rohani, P., and O. Miramontes. 1995. Host–parasitoid metapopulations: the consequences of parasitoid aggregation on spatial dynamics and searching efficiency. Proceedings of the Royal Society of London. Series B: Biological Sciences 260:335–342.

Rohani, P., and G. D. Ruxton. 1999. Dispersal induced instabilities in. Theoretical Population Biology 55:23–36.

Rohani, P., D. Earn, and B. Grenfell. 2000. Impact of immunisation on pertussis transmission in england and wales. Lancet 355:285–286.

Rohani, P., X. Zhong, and A. A. King. 2010. Contact network structure explains the changing epidemiology of pertussis. Science 330:982–985.

Roland, J., and P. D. Taylor. 1997. Insect parasitoids respond to forest structure at different spatial scales. Nature 386:710–713.

Roos, A. D. 1991. Mobility Versus Density-Limited Predator–Prey Dynamics on Different Spatial Scales. . . . of the Royal . . . 246:117–122.

Root, R. B. . 1973. Organization of a Plant–Arthropod Association in Simple and Diverse Habitats : The Fauna of Collards (*Brassica Oleracea*). Ecological Monographs 43:95–124.

Roughgarden, J. 1974. Population Dynamics in a Spatially Varying Environment : How Population Size "Tracks" Spatial Variation in Carrying Capacity. The American Naturalist 108:649–664.

Ruxton, G. D., and P. Rohani. 1996. The Consequences of Stochasticity for Self–Organized Spatial Dynamics, Persistence and Coexistence in Spatially Extended Host–Parasitoid Communities. Proceedings of the Royal Society B Biological Sciences 263:625–631. doi: 10.1098/rspb.1996.0094.

Salathé, M., M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones. 2010. A high-resolution human contact network for infectious disease transmission. Proceedings of the National Academy of Sciences 107:22020–22025.

Salmore, A. K., and M. D. Hunter. 2001. Elevational trends in defense chemistry, vegetation, and reproduction in *Sanguinaria canadensis*. Journal of chemical ecology 27:1713–1727.

Schönemann, P. H. 1966. A generalized solution of the orthogonal procrustes problem. Psychometrika 31:1–10.

Shapiro, E. D. 2012. Acellular vaccines and resurgence of pertussis. JAMA 308:2149–2150.

Shi, J., and J. Malik. 2000. Normalized cuts and image segmentation. Pattern Analysis and Machine Intelligence, IEEE Transactions on 22:888–905.

Simon, H. D., and S.-H. Teng. 1997. How good is recursive bisection? SIAM Journal on Scientific Computing 18:1436–1445.

Skowronski, D., G. De Serres, D. MacDonald, et al. 2002. The changing age and seasonal profile of pertussis in canada. J Infect Dis 185:1448–1453.

Snijders, T. A., and K. Nowicki. 1997. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. Journal of Classification 14:75–100.

Steele, J. 1974. Spatial Heterogeneity and Population Stability. Nature 248.

Tan, T., E. Trindade, and D. Skowronski. 2005. Epidemiology of pertussis. Pediatr Infect Dis J 24:S10–S8.

Taylor, A. D. 1991. Studying metapopulation effects in predator-prey systems. Biological Journal of the Linnean Society 42:305–323.

Tenow, O., A. C. Nilssen, H. Bylund, R. Pettersson, A. Battisti, U. Bohn, F. Caroulle, C. Ciornei, G. Csóka, H. Delb, et al. 2013. Geometrid outbreak waves travel across europe. Journal of Animal Ecology 82:84–95.

Thies, C., and T. Tscharntke. 1999. Landscape structure and biological control in agroecosystems. Science 285:893–895.

Thies, C., I. Steffan-Dewenter, and T. Tscharntke. 2003. Effects of landscape context on herbivory and parasitism at different spatial scales. Oikos 101:18–25.

Tu, C.-C., C.-K. Shieh, and H. Cheng. 2000. Algorithms for graph partitioning problems by means of eigenspace relaxations. European Journal of Operational Research 123:86–104.

Turing, A. 1952. The chemical theory of morphogenesis. Phil. Trans. Roy. Soc 13.

Underwood, N. 2004. Variance and Skew of the Distribution of Plant Quality Influence Herbivore Population Dynamics. Ecology 85:686–693.

———. 2009. Effect of genetic variance in plant quality on the population dynamics of a herbivorous insect. Journal of Animal Ecology 78:839–847. doi: 10.1111/j.1365-2656.2007.0.

Underwood, N., and M. D. Rausher. 2000. The effects of host-plant genotype on herbivore population dynamics. Ecology 81:1565–1576.

United Kingdom. Health Protection Agency. 2008. Completed primary courses at two years of age: England and wales, 1966 - 1977, england only 1978 onwards. `http://www.hpa.org.uk/web/HPAweb\&HPAwebStandard/HPAweb_C/1195733819251`. (Accessed May 14, 2013).

———. 2012*a*. 2012 press releases: Whooping cough cases continue to increase. `http://www.hpa.org.uk/NewsCentre/NationalPressReleases/2012PressReleases/121130Whoopingcoughcasescontinuetoincrease/`. (Accessed May 14, 2013).

———. 2012*b*. Infection report: Confirmed pertussis in england and wales: data to end-september 2012. Health Protection Report 6.

———. 2013. Whooping cough (pertussis): Latest epidemiological data: Table 4: Pertussis table by age group, table 6: Pertussis notifications and deaths, and table 7: Pertussis notifications by age. `http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/WhoopingCough/EpidemiologicalData/`. (Accessed May 14, 2013).

United Kingdom. Office for National Statistics. 1988. Statistical bulletin: Population estimates for england and wales, mid-2002 to mid-2010 revised (national). `http://www.ons.gov.uk/ons/rel/pop-estimate/population-estimates-for-england-and-wales/`

`mid-2002-to-mid-2010-revised--national-/stb---mid-2002-to-mid-2010-revised-popu` `html`. (Accessed May 15, 2013).

Verma, D., and M. Meilă. 2003. Comparison of spectral clustering methods. Technical report, CSE-03-05-01, University of Washington.

Von Luxburg, U. 2007. A tutorial on spectral clustering. Statistics and computing 17:395–416.

Wallinga, J., P. Teunis, and M. Kretschamar. 2006. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. Am J Epidemiol 164:936–944.

Wang, M.-H., M. Kot, and M. Neubert. 2002. Integrodifference equations, Allee effects, and invasions. J Math Biol 44:150–168.

Warfel, J. M., L. I. Zimmerman, and T. J. Merkel. 2014. Acellular pertussis vaccines protect against disease but fail to prevent infection and transmission in a nonhuman primate model. Proceedings of the National Academy of Sciences 111:787–792.

Watts, D. J., and S. H. Strogatz. 1998. Collective dynamics of small-worldnetworks. nature 393:440–442.

Wearing, H. J., and P. Rohani. 2009. Estimating the duration of pertussis immunity using epidemiological signatures. PLoS pathogens 5:e1000647.

Wendelboe, A., A. Van Rie, S. Salmaso, and J. Englund. 2005. Duration of immunity against pertussis after natural infection or vaccination. Pediatr Infect Dis J 24:S58–61.

Wendelboe, A., E. Njamkepo, A. Bourillon, et al. 2007. Transmission of bordetella pertussis to young infants. Pediatr Infect Dis J 26:293–9.

Wood, N., and P. MacIntyre. 2008. Pertussis: review of epidemiology, diagnosis, management and prevention. Pediatr Respir Rev 8:291–212.

World Health Organization Department of Immunization, Vaccines And Biologicals. 2013. World health organization vaccine-preventable diseases: monitoring system 2013 global summary. `http://apps.who.int/immunization_monitoring/globalsummary/timeseries/tscoveragebcg.html`. Last accessed: May 19, 2014.

Zepp, F., U. Heininger, J. Mertsola, E. Bernatowska, N. Guiso, J. Roord, A. E. Tozzi, and P. Van Damme. 2011. Rationale for pertussis booster vaccination throughout life in europe. The Lancet infectious diseases 11:557–570.

Zhang, Z., M. I. Jordan, et al. 2008. Multiway spectral clustering: A margin-based perspective. Statistical Science 23:383–403.