Patterns of livestock-pathogen transmission and emergence

a framework for categorising transmission

Thesis submitted in accordance with the requirements of the
University of Liverpool for the degree of Doctor in Philosophy by Andy Fohrmann.

May 2011

Acknowledgements

I wish to acknowledge Professor Malcolm Bennett for offering me the opportunity to work towards a PhD in the Department of Epidemiology and Population Health, Institute of Infection and Global Health. The project was funded by the Department for Environment, Food and Rural Affairs (DEFRA) and the National Centre for Zoonosis Research (NCZR).

I am most indebted to Dr. Jonathan M. Read who acquiesced to my request to (co-)supervise the PhD project at a time the project was already in an advanced if not final stage. He took over from Dr. Helen Clough who herself replaced Dr. Sophia Latham.

I am very grateful for the data on *Salmonella* incidences provided by Susan A. Kidd of the Veterinary Laboratories Agency (VLA), Lynda Browning of the Health Protection Scotland (HPS) and Katy Hutton of the Health Protection Agency (HPA). These people were also very helpful regarding the correct interpretation of the data. Moreover, I would like to express my gratitude to Louise Taylor and Dr. Sarah Cleaveland for providing detailed information and data on their work.

Furthermore, I would like to thank all colleagues who contributed to this study, for example, Giles Constant for his contribution in data collection on the viruses, Dr. Christian Setzkorn for maintaining the database system and Professor Tom Humphrey for his advice on *Salmonella*-related issues.

There were various people who offered help and support on issues related to the thesis, though these subjects eventually were not included in this thesis. For instance, Professor Paul Goldberg and Dr. Michele Zito answered questions concerning an algorithm and Dr. Ian C. Smith and Dr. Joanne Turner supported me with the use of the Condor Computing Software. The Computing Services of the University of Liverpool also provided assistance on several occasions. Dr. John C. Davies's help to identify an issue concerning my version of Matlab, for example, was greatly appreciated.

Finally I also wish to acknowledge those people at Leahurst who made my life a little easier in various ways. These people include, amongst others, Dr. Nicola J. Willams, Dr. Trevor R. Jones, the librarian Patricia Jonker-Cholwe, Judi Young, the security men, Dr. Laura A. Hughes, Dr. Karen P. Coyne and Dr. Peter J. Cripps.

Andy Fohrmann, Neston, November 2011

Abstract

Infectious diseases continue to cause significant financial, economic and health issues. Much work has been done on a variety of pathogens and host species with a wide range of approaches and methodologies from many disciplines, yet a full understanding of pathogen emergence has not been achieved. A greater understanding of transmission would benefit the management of control strategies, help to improve decision-making at first diagnosis of a host-pathogen interaction and possibly prevent further spread of epidemics and therefore new cases of emergence. A major gap in all these efforts appears to be a comprehensive, detailed approach for categorising pathways of transmission.

Here, I develop and present a framework to categorise transmission events of pathogens. I use this framework to review the available literature on the transmission of pathogens of seven livestock species, and to derive intra- as well as inter-species transmission routes or pathways by which pathogens may emerge and infect the livestock hosts.

A general dataset of the livestock pathogens was constructed, containing information on 141 bacterial and viral pathogen species. One finds these pathogens exploit 7204 distinct transmission routes between 215 different host species. Analysis of this dataset verifies the dominance of multi-host pathogens, and suggests that non-close or indirect transmission routes tend to be associated with larger host range. The data shows the faecal-oral transmission mode is the predominant transmission pathway for livestock infection.

To attempt to link features of species-scale transmission routes with emergence rates, a second dataset was constructed. This was a more detailed description of twelve *Salmonella enterica* serovars, and holds 1716 distinct transmission routes, 110 distinct host species. One finds the connectedness of a host species within its network of transmission routes is associated to its potential for acquiring *Salmonella* infection. Analysis of both datasets highlights the substantial importance of anthropogenic mechanical vectors such as food and feedstuffs for the emergence of infectious disease in both humans and livestock animals.

The developed framework demonstrably captures important aspects of pathogen emergence, although there are limitations in the quality and availability of information required to generate the datasets. The networks of transmission routes derived using this framework show common features and structural properties which may explain emergence potential. It is hoped that, by adopting this framework, the scientific community can improve the understanding of the drivers of transmission and emergence of these important pathogens.

Contents

Preface

"Bring in your idols to tell us
what is going to happen.
Tell us what the former things were,
so that we may consider them
and know their final outcome.
Or declare to us the things to come,
tell us what the future holds,
so we may know that you are gods..."

The Bible, New international version,
Isaiah, 41.22-23

The author would like to stress that this bible quotation is not meant to reflect any religious, personal, scientific etc. attitudes whatsoever of the people partaking in the project this work arose from. The author alone chose this quotation merely to highlight mankind's ancient desire to obtain useful information and knowledge from past data as well as the concurrent awareness of the intricate challenges when tackling this task. However, the author strongly hopes that at least one structuring force does exist and leaves the choice of terms and concepts in order to refer to such forces to the reader. A complete lack of structuring forces appears to make the universe a place of pure chance and arbitrariness, thereby excluding the possibility of causal chains. This in turn would seem to render scientific research and the expectation to discover any governing laws and rules pointless.

It should also be mentioned that most, if not all, relevant terms in the field of epidemiology and infection biology are highly ambiguous and are often used quite loosely. This holds true even for fundamental terms like, for example, "concept" [68] [109], "model" [93] [97], "ontology" [3] [7] [13] [82] [174] etc. The reader is clearly encouraged to make up his own mind about the applied terminology and to disagree with the author. Moreover, terms might have been used before the attempt of a definition has been made. The reader should therefore consider to use the list of contents, section 2.1.1 and the index interactively with the rest of this document.

Chapter 1

Introduction

"It is time to close the book on infectious diseases, declare the war against pestilence won, and shift national resources to such chronic problems as cancer and heart disease ..."

U.S. Surgeon General W.H. Stewart, 1967 [108]

"To write about infectious disease is almost to write of something that has passed into history [...] the most likely forecast about the future of infectious disease is that it will be very dull ..."

Sir Frank McFarlane Burnett and David O. White, 1962 [72]

In August 2007, outbreaks of foot-and-mouth disease had been confirmed on two cattle farms in South-East England [47] [173]. Soon afterwards additional cattle-keeping premises as well as sheep were affected with the foot-and-mouth disease virus [46]. These outbreaks caused severe economic losses to local farmers and affiliated enterprises [29] [71] but in the end also affected the tax-payer. The foot-and-mouth outbreaks were preceded by an outbreak of the avian influenza virus earlier that year [49] and were immediately followed by outbreaks of the bluetongue virus [48] [190]. These examples have been chosen mainly because they coincided both with the starting year and geographic area which this PhD thesis has been conducted in. The list of infectious disease outbreaks and the host species affected is very large and clearly contradicts the quotations above.

1.1 General motivation

> *"The expectation that infectious disease was in decline and would continue [...] to decline toward complete elimination, held on tenaciously even in the face of well known and dramatic exceptions. [...] The accumulation of "exceptions" [...] forced a new awareness that diseases rise and fall, evolve and spread and retreat and spread again, and that we have to prepare for a more complex tomorrow ..."*
>
> Richard Levins, 1994, [104]

Financial and economic issues as major motivational drivers for conducting research in the field of infectious diseases of livestock animals have already been mentioned. In fact, few infectious diseases affect exclusively any single one of the stated groups of host animals but rather exist within a so-called *host-parasite (ecological) continuum* [36] - or alternatively called *multihost-pathogen community* [60] - that include many host groups [36]. Hence the emergence and reemergence of infectious diseases in any non-human host species may pose a fundamental threat to society by means of zoonotic transmissions. Such an event could potentially trigger epidemics and pandemics in humans [95] [102]. Infectious diseases still remain the major cause of human death on a global scale and their incidence seems to increase [102] [150]. Over approximately the past two decades, ca. 75% of all emerging human diseases originated from pathogens that initially have been assigned to non-human animals [17] [115].

Another motivation for reasearch on infectious diseases is conservation biology. Human migration and the expansion of grassland could bring livestock and humans in close contact with wildlife populations [1] [94], thereby increasing the chance of transmissions between these host categories. This might put precious and often endangered wildlife species at risk of becoming extinct [36] [145]. It should be noted that even parasites and/or pathogens constitute a biodiversity in their own right. Declines and extinction of wildlife species may in turn have cascading negative effects on the biodiversity of affiliated parasites and/or pathogens [2]. This may sound as a wellcome side effect at first but we have to keep in mind that we are dealing with a highly nested, complex, non-linear system in which all parts are interrelated. Changing one part of the system could lead to drastic changes in other parts too.

A further motivation is the domestication[1] of pathogens by exploiting the latest available techniques, for example through the development of live attenuated vaccines, plant vector systems like *Agrobacterium tumefaciens* or the control of pests using pathogens like *Bacillus thurigiensis*. Strategies of bringing in a disease and/or a genetically modified pathogen in order to control a pest can be highly risky and might pose a threat to other host species too [140]. It is certainly within the realms of possibility that newly developed pathogens, even if designed by state-of-the-art techniques, could unexpectedly broaden their host range [80] [184]. On top of that, cultivation and/or manipulation of microorganisms might also be used in the course of an act of biocrime or bioter-

[1] That is the selection of seemingly useful traits which are potentially concurrent with maladaptive traits [177].

rorism [17] [77] [55]. Such scenarios are even more worrying when considering that some authors still consider research on the spread and transmission of infectious diseases in its infancy [123].

The potential of infectious diseases to lead to catastrophic outbreaks has already been metaphorically described as the microbial equivalent of a *perfect storm* [17]. Interestingly, the potential relationship between both multifactorial events, i.e. geophysical disasters on the one hand[2] and disease outbreaks in their aftermath on the other hand, still seems to be a matter of debate [61] [195].

The previous paragraphs highlight the need for a better understanding of the ecology, evolution and social effects of infectious diseases which are important from a self-preservation but also financial and economic points of view. In order to tackle such issues, governmental and non-governmental organisations as well as private companies have become increasingly interested in collecting and analysing potentially relevant information. The resulting patterns need to be evaluated for their usefulness and/or meaning. Such efforts might confirm existing hypothesis or give rise to new ones which ideally will contribute to our understanding of the nature of the underlying processes. Organisations that conduct or are involved in such data collections comprise, for example, the VLA (Veterinary Laboratories Agency), the HPA (Health Protection Agency), the HPS (Health Protection Scotland), the WHO (World Health Organization), the OIÉ (Office International des Épizooties), the HHS (Department of Health and Human Services), the USDA (United States Department of Agriculture), the CDC (Centers for Disease Control and Prevention) and the NIAID (National Institute of Allergy and Infectious Disease), just to name a few. Even organisations such as the FBI (Federal Bureau of Investigation) [77] [55], Google [70] and the Bill & Melinda Gates foundation [12] are contributing to this field.

1.2 Concepts, terms and definitions

> *"The search for the* mot juste *is not a pedantic fad but a vital necessity. Words are our precision tools. Imprecision engenders ambiguity and hours are wasted in removing verbal misunderstandings before the argument of substance can begin ..."*
>
> *"Words are meant to convey thought; if you take trouble in the use of words you are bound to clarify the thought which you wish to convey ..."*
>
> *"What appears to be a sloppy or meaningless use of words may well be a completely correct use of words to express sloppy or meaningless ideas ..."*
>
> Anonymous [206]

Shakespeare [164] highlighted the importance of understanding the terms and expressions used in the study of any scientific subject from the outset. This is to comprehend the underlying idea[3] to

[2] For example earthquakes, volcanic eruptions, tsunamis, cyclones etc.

[3] The terms idea, notion and concept are often used synonymously.

which a term is referring. It seems obvious to start with some of the terms that have already been used so far.

1.2.1 Infection and infectious disease

The first cluster of terms that we are going to investigate comprises the terms *infection* and *infectious disease*. Infection implies that a pathogen has taken up residence in, or even on [66], a host and is multiplying and/or developing. This is synonymous with the term *colonisation* [101] [102] [66]. However, this does not necessarily lead to obvious disease and clinical signs. Infection is therefore not the same as infectious disease [101] [102]. In fact, colonization "... is a term that describes growth of a microorganism in the body without damage to tissues" [66] which is why the "... normal flora may be described as colonization" [66]. Infection is therefore the required but not the sufficient condition for infectious disease in this context.

Key to infection and infectious disease is the transmission of a pathogen to a susceptible host. Therefore it seems appropriate to challenge the suitability of these expressions. Alternatively one could use the expression *host-pathogen interaction*. This term would on the one hand refer to an infectious interaction between a host and a pathogen. On the other hand it also accounts for the fact that a host can be infected without showing any disease or clinical signs. This makes even more sense given that today's screening techniques render it possible to discover a pathogen before its associated disease symptomes [102]. In other words, the term draws attention to the pathogen and its ability to interact with a host rather than its actual impact on the host.

For the sake of completeness it should be mentioned that some authors not only refer to an infectious disease as an illness caused by the initial entry and multiplication of a pathogen in a host. They also include illnesses caused by the mere uptake of toxic products originating from a microorganism [101]. This point of view is clearly at odds with some of the statements made earlier in this section. Infection would not then be a required condition for an infectious disease. Transmission of a pathogen to a susceptible host would also no longer be key to infection and infectious diseases.

The idea put forward in the foregoing paragraph requires to reconsider the the term pathogen. Another term that can be encountered regularly in the literature is *microbial threat*. This term was introduced by the Institute of Medicine in order to include "... both the agent and the disease" [102]. The term agent is crucial in this context. It was meant to comprise "... *all* classes of infectious agents" [102]. In fact, the Institute of Medicine also took toxicological expertise into account. An agent of disease, according to Last [101], is any factor which is essential for disease occurrence. This would leave it open if the actual threat is posed by a microorganism or any of its toxic products. A microbial threat is therefore any agent of direct microbial origin which poses a threat to hosts in terms of infection and/or disease. Transmission of a "microbial threat" is therefore still key to an infectious disease with respect to a definition including toxic products of microorganisms.

The literature uses a remarkable number of terms in order to refer to agents of infection and/or disease. This multitude of terms is full of synonyms, homonyms, contradictions and specifications. Some of these terms have been used earlier already. It was not our intention to discuss them in detail but rather to list a few:

pathogen, parasite, microparasite, macroparasite, pathogenic microbe, agent, infectious agent, transmissible agent, communicable agent, pathogenic agent, disease agent, etiologic agent, microbial agent, parasitic agent, microbial threat etc.

I shall henceforth favour the use of the terms microbial threat and pathogen which will be used interchangeably, though the first term is broader in its meaning. A microbial threat may be a viable microorganism itself, any other form of microbial existence (e.g. a virus), a biochemical product of a microbial threat (e.g. a toxin) or a microbial survival stage (e.g. a spore). A microbial toxin on its own would certainly not be regarded as a pathogen but this should not be of great importance with regard to this thesis.

1.2.2 Transmission

Transmission is viewed as "...the driving force in the dynamics of any infectious disease" [9] and its suppression is deemed key to the control of the spread of infectious diseases [102]. For example, vaccination of the host species whose presence dominates the transmission of a microbial threat may contribute to its eradication [50].

Transmission has been defined by Last [101] as any direct or indirect mechanism by which a microbial threat is spread from a source or reservoir to a susceptible host. Woolhouse *et al.* suggested the term transmission event[4] to refer to a process by which a pathogen exits a sender host, is then transmitted between sender and receiver host and finally enters a receiver host [207].

Microbial threats can be categorised by their ways or strategies of transmission [105] [143]. Previous efforts distinguished a few broad and, according to most references, mutually non-exclusive transmission categories. These categorisations, though very similar, differ slightly depending on the actual reference. The literature mainly mentions two categories:

1. Transmission via close/direct contact [2] [143] [145] [182];
2. Transmission via non-close/indirect contact [143] [145] [182].

Transmission via close/direct contact involves physical contact which requires close proximity of the partaking hosts. It can include such events as, for example, touching, biting, scratching, transmission via wounds, sexual contact, vertical transmission [2] [143] [145] and also inhalation [182]. The term "inhalation" and associated terms and issues need a separate investigation which can be found in the subsection 2.2.3.1. For now let it suffice to assume that inhalation in this context requires close proximity between hosts. The second class is known as non-close/indirect contact

[4]The authors actually made use of the expression "route of transmission" [207]. This expression was already assigned to a slightly different idea in the context of this thesis (see section ??). I therefore replaced 'route of transmission' with 'transmission event'.

transmission and can involve contact with contaminated environments such as soil or water, or with contaminated food and/or faeces [143] [145] [182].

Additional categories that can be found in the literature are either explicitly part of the non-close/indirect category [2] or have been singled out as separate categories [143] [145] [182]. The latter ones include the categories 'transmission via arthropod vectors' [143] [145] [182] and 'transmission via intermediate hosts' [143] [145]. The description and/or use of these two categories can be quite ambiguous and contradictory. For example:

- At least arthropods and helminths [146] - in fact even bacteria - can serve as hosts for other parasites or pathogens too. However, it can be acted on the assumption that the authors of [143], [145] and [182] did not view any of these parasites as hosts, even though certain statements might cast some doubt on that. This point of view originally derived from the medical field which focus is on vertebrate hosts in the first place and therefore treats hosts other than vertebrates differently [11].
- For the category 'transmission via arthropod vectors' the arthropods are not even viewed as parasites anymore but rather provide an exclusively mechanical link which facilitates pathogen transmission between the hosts in focus. The definition of Woolhouse *et al.* includes "... biting or mechanical transfer by arthropods" [182]. This wording might suggest that an arthropod bite is fundamentally different from a mechanical transfer. It seems likely that these authors meant to distinguish between a more active and a more passive role of the arthropod in the transmission. In the active scenario the pathogen is simply attached to the mouthparts of the arthropod and enters the receiver host via an "active" bite through its skin.
- The difficulty with the category 'transmission via intermediate hosts' consists of identifying the parasites and pathogens it applies to. An intermediate or secondary host apparently harbours a sexually immature "parasite", usually until a developmental stage of the parasite's life-cycle is completed. Such a definition would obviously exclude all parasites/pathogens which do not reproduce via a sexual process (e.g. viruses and bacteria). Reference [143] includes those parasites that exhibit a complex life-cycle and/or are transmitted via the food chain (i.e. trophic transmission, see also [99]). Therefore, arthropods with complex life-cycles such as ticks as well as helminths should belong to this category. However, according to [143] is trophic transmission a sufficient characteristic of this category. This would actually include viruses and bacteria which would result in a conflict since they were already excluded based on the requirement for sexual reproduction.

Another expression that is to be encountered regularly in the context of pathogen transmission is *microbial traffic*. Microbial traffic has been defined in slightly differing ways. However, it is mainly known as the processes by which existing microbial threats are disseminated and transferred into (new) host populations (regardless of countries or international boundaries) [95] [102] [124] [125]. This definition, by the way, appears to be very close to Last's definition of the term transmission [101] as stated at the beginning of this subsection. The main difference between these viewpoints appears to be that the term microbial traffic explicitly emphasises the fact that transmission can happen over large geographical distances. As a result long-distance transmission would probably be linked to an increased risk of causing infection in hitherto isolated host populations.

With respect to the foregoing paragraph, it is now obvious to assume that transmission and in particular microbial traffic could be major drivers of so-called *Transboundary Animal Diseases*

(TAD)[5]. Transboundary Animal Diseases have been described as diseases "...which are highly contagious or transmissible and have the potential for very rapid spread, irrespective of national borders, causing serious socio-economic and possibly public health consequences" [63]. The last part of this description suggests that transboundary animal diseases should not be considered as an issue affecting only non-human animals [91]. It seems obvious to assume that TAD's to be the prime suspects for initiating what has been described as the microbial equivalent of a perfect storm [17].

The spread of infection and its underlying mechanisms are still considered elusive in many respects [123] [177]. It is also debatable if we can ever gain a sufficiently profound understanding of these processes which would enable us to keep microbial threats at bay for good [95]. Therefore it seems more likely that humankind has to continue and refine both research and surveillance permanently [95].

1.2.2.1 Basic reproductive ratio R_0

An important concept in the study of the spread of infections is the *basic reproductive ratio* R_0. It "...often serves as a threshold parameter that predicts whether an infection will spread" [79]. In this context[6] R_0 is usually defined as the expected (i.e. average) number of secondary (i.e. new) cases of an infection caused via the introduction of a typically infected individual (i.e. the primary case) into a sufficiently large population consisting of susceptible (i.e. non-infected, unexposed) host individuals only [78] [204]. The precise definition might slightly differ depending on the actual reference. Even the term 'basic reproductive ratio' itself has not been standardised so far. An often used and accepted alternative is the term *basic reproductive number*[7]. Some authors advise against the usage of yet another term, namely *basic reproductive rate*, since R_0 is meant to be a dimensionless quantity [78]. The definition of R_0 already indicated that it is supposed to characterise the interaction of a particular pathogen with a particular host population. This is reflected in figure 1.1 by the R_0's depicted in the bottom left corner of each population box. In fact, it has already been pointed out that, even when focusing on one particular pathogen, R_0 can still vary considerably between different populations of the same host species [204] as indicated, for example, by the turkey populations T1 and T2 in figure 1.1. This might be explained by various factors such as different farming practices, differences in the age and/or sex structure of populations, differences in the spatial structure of farms or the number and/or density of animals present etc. Such significant variations of R_0 can also be seen between distinct strains of the same pathogen species when focusing on one particular host population (compare swine populations P1 and P2 in figure 1.1). In other words, certain strains might be more adapted to a certain host species and/or host population. If a single primary case - on average - gives rise to more than one secondary case,

[5] Transboundary Animal Diseases were originally identified on the OIÉ List A [91]. According to [27] they were identified on the OIÉ Lists A and B. However, both of these lists have already been discarded by the OIÉ as no longer helpful for classifying animal diseases [91] [209].

[6] R_0 is not only used in epidemiological modelling with respect to the spread of infections but also in demography and ecology with an analogous interpretation.

[7] Another term that is sometimes used in order to refer to R_0 is *transmission potential* [207].

then $R_0 > 1$. Such a scenario bears the potential of an exponential increase in the number of cases which in turn could result in larger outbreak sizes. In contrast, a single primary case should - on average - result in self-limiting sequences of transmission events if $R_0 < 1$ [204]. $R_0 = 1$ could be considered a borderline case. The expected number of secondary cases would not be based on exponential growth. The infection would, at least theoretically, continue to spread and remain present in the population. However, the infection might also cease to exist within the population after a prolonged sequence of transmission events as compared to $R_0 < 1$. The outbreak size itself is supposed to be nonlinearily related to both R_0 and the number of primary cases introduced into the host population [207] [154][8]. Nonlinearity implies that "...small variations in either can lead to large variations in outbreak size" [207]. It therefore seems to be considerably difficult to predict the scale of outbreaks within host populations based on R_0 and the number of primary cases.

The arrows in figure 1.1 symbolise transmission events from a donor host individual to a receiver host individual. They imply two kinds of uncertainty. The first is the mode (i.e. route) of transmission by which a transmission event takes place. Usually it is only possible to make more or less well-informed assumptions about how exactly a pathogen is or was transmitted from the donor to the receiver in any particular case. Perhaps there is more than one mode of transmission possible for a particular transmission event. The second uncertainty addresses the likelihood or potential of a particular transmission event to take place which might also differ between distinct modes of transmission. Figure 1.1 distinguishes three kinds of transmission events with respect to:

- Intra-population transmission;
- Inter-population-intra-species transmission;
- Inter-population-inter-species transmission.

In the first case, R_0 would be defined in the already stated fashion, i.e. the individual unit in focus is the individual itself in order to model within-population spread. The general concept of R_0 might also be applied to different other levels, e.g. the level of farms or populations in order to model the spread between farms and populations or even the level of cells within a particular host individual in order to model in-host dynamics [79]. It should also be noted that there can be considerable and important differences between individuals of the same species and populations of the same species etc. The *next generation* approach to model R_0 allows for multiple, discrete and disjoint classes of host individuals within a population rather than simplifiying matters by only distinguishing populations merely by their location and infection status as a whole [79]. Figure 1.1 accommodates such R_0-derivatives referring to the second and third bullet points via the R_{02}'s in the bottom left corners of the dashed host species boxes and the R_{03}'s in the pathogen boxes.

[8]The authors of [207] and [154] were actually referring to 'host species' but given the statements on R_0 so far, 'host population' appeared to be more appropriate.

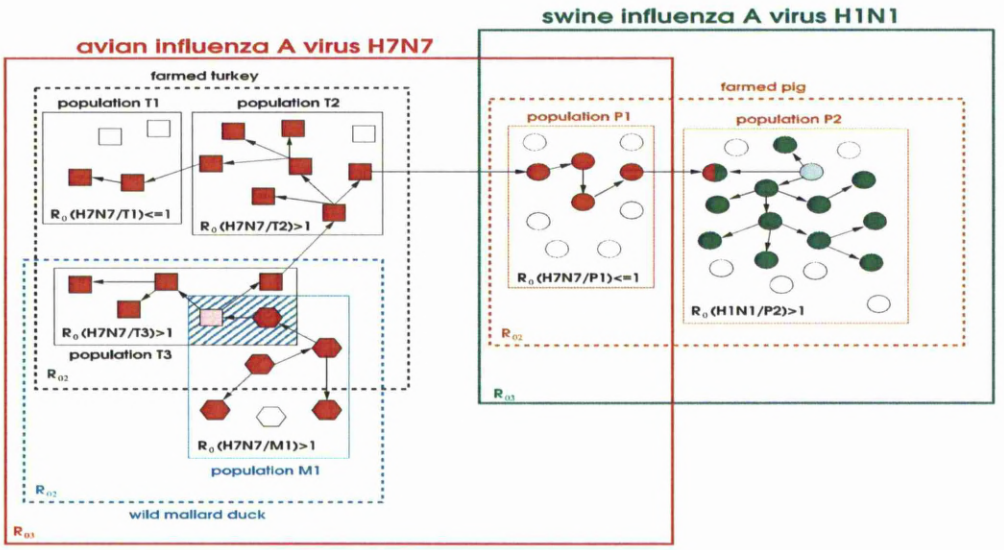

Figure 1.1: Transmission of pathogens and spread of infection: This figure uses the example of influenza virus to summarise some aspects of the complex nature underlying the transmission and spread of pathogens and the resulting emergence of host-pathogen interactions. Information from references [18] and [41], on two different strains of the (-)ssRNA influenza A virus, has been used to create a roughly realistic scenario. The figure will be referred to and further explained at different passages in the thesis. Parameters like geographical areas, population sizes, associated R_0 values etc. are likely to change over time. The figure is therefore meant to be a snap-shot, covering a relatively narrow time frame.

Circles (pig), ***squares*** (turkey) and ***hexagons*** (mallard duck) refer to host individuals whereas ***arrows*** refer to transmission events. Empty or white hosts refer to unexposed individuals and otherwise coloured hosts to infected individuals. The lighter red and lighter green hosts are meant to be the first identified cases of infection. ***Solid pathogen boxes***: The area that is considered to be inhabited or contaminated by an influenza A strain either inside or outside of any host i.e. red=H7N7 and green=H1N1. ***Dashed host species boxes***: The area in which populations and/or individuals of the corresponding host species have been reported to occur, i.e. black=turkey, brown=pig and cyan=mallard. ***Solid host population boxes***: The area which a population is inhabiting during the time of the snap-shot, colour code see host species boxes.

The H7N7 strain is introduced into the farmed turkey populations by an individual belonging to the wild mallard population M1 [41] [18] which shows a spatial overlap with the turkey population T3 (see cyan-striped area). Infected mallards might have accidentally occupied a farmed area of free-range turkeys leading to temporary environmental contamination of this area. Once introduced into a turkey population the infection spreads to other farmed populations via any form of indirect mechanical transfer. The situation in the turkey population T2 promotes a rapid and exponential spread within the population via a combination of direct and indirect transmission routes, but moreover causes the spread to other populations including the pig population P1 via an indirect route of transmission. P1 and T2 might, for example, belong to the same mixed farming sytem rearing turkeys and pigs [41]. Though both host species might be kept in different spatial locations, farm workers and/or farming equipment could have been involved in an indirect form of transmission. The H7N7 strain is not able to cause epidemic spread in pig populations [41]. Nevertheless it is capable of prolonged pig-to-pig transmission [18] and it eventually emerges in a hitherto H7N7-free area, namely farmed pig population P2. The half-red, half-green circle refers to a coinfected pig which might give rise to a new emerging variant of influenza A [127] [138].

1.2.3 Emergence

"*Each event is quite unique.*
Nothing ever happens twice.
What occurs will not recur.
There can be no second time ..."

Flux by Theodore Melnechuk [120]

The preceding subsections highlighted the importance of transmission to infections and infectious diseases. Two closely associated terms are *emergence*[9] and *reemergence* (see also figure 1.2). The 'Institute of Medicine' [102] defined emerging and reemerging infectious diseases as those whose incidence[10] has increased within the past two decades or that threatens to increase in the near future [178]. The difference between emerging and reemerging is that the second term refers to an infectious disease or infection which had been 'previously recognized', whereas the first term refers to a 'newly defined' infectious disease or infection [178]. 'Previously recognized' is likely to address an increase in incidence of a particular host-pathogen interaction that had been previously recorded within a given geographical area. The pathogen therefore resurges within a particular host species and in a particular geographic area following a preceding decline in its incidence with respect to this very host as well as this very area [17] [150]. The decline of incidence is sometimes referred to as *demergence* and might in fact be of equivalent interest for the study of the dynamics of infectious diseases [168]. 'Newly defined'[11] on the other hand describes a situation in which either a known or hitherto unknown pathogen infects any given host species for the first time. Alternatively it can refer to a situation in which a familiar pathogen infects a susceptible and already established host species in a hitherto unaffected geographic area [17]. In other words, a newly defined pathogen implies both a newly defined host species as well as a newly defined geographic area of incidence. A newly defined host species, however, neither necessarily points to a newly defined geographic area nor to a newly defined pathogen. Consequently does a newly defined "geographic area" neither hold any information on the 'newly defined' status of the host species nor on that of the pathogen.

[9] Reference [117] provides both a short historical background on the term 'emergence' as well as corresponding ideas from a systems theoretical point of view. These ideas have neither been derived from, nor are they aiming at, the context of host-pathogen interactions. They are also presented in a slightly confusing manner. Nevertheless, the concepts put forward in this book suggest that emerging phenomena might be understood as the result of the interaction between highly nested and hierarchical systems.

[10] Referring to the human population, i.e. species, as a whole.

[11] Probably synonymous with 'newly identified'.

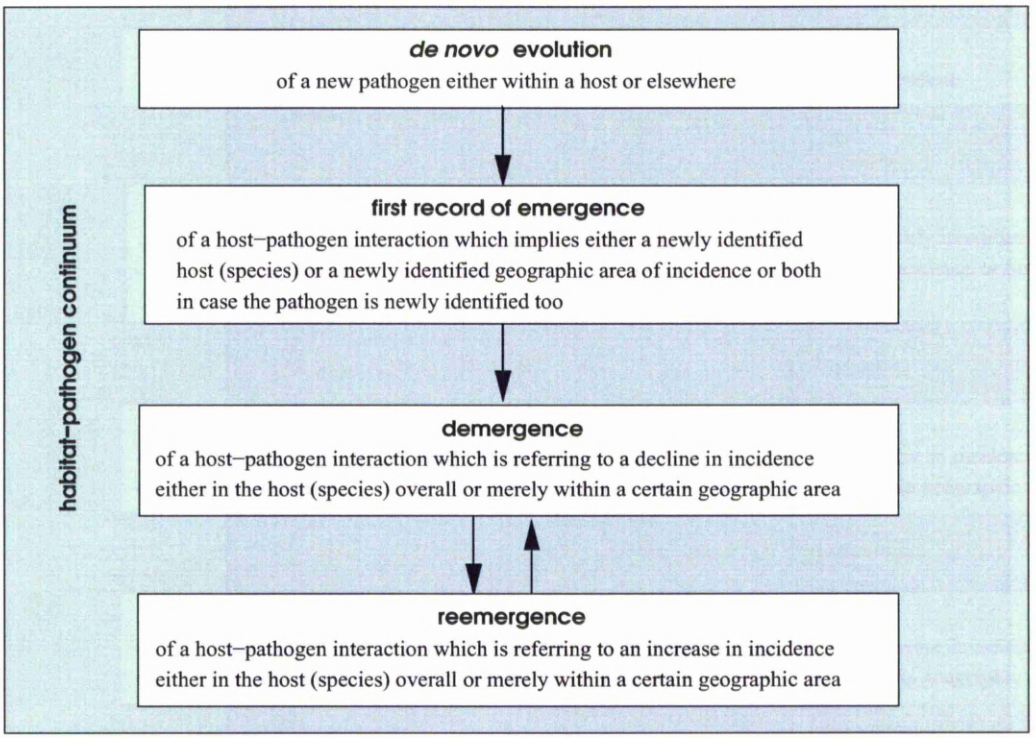

Figure 1.2: The concept of emerging host-pathogen interactions: The flow chart summarises some major terms and events which are associated with the concept of emerging host-pathogen interactions. In order to simplify matters all events are meant to address one pathogen and one host (species) only. Note that a host might not necessarily be viewed on a species level but could also be looked at on a population, genus etc. level, which is why species appears in parentheses. All events take place against the background of an omnipresent and ubiquitous real world truth which will be referred to as the habitat-pathogen continuum (here indicated in green colour). The *de novo* evolution (or alternatively *de novo* synthesis) of a pathogen could take place either within a host (see intra-host speciation in subsection 1.2.5) or elsewhere. In the former case would the *de novo* evolution coincide with the "genuine emergence" of a new host-pathogen interaction. In the latter case, though, would the "genuine emergence" of the new host-pathogen interaction take place at a later point in time. Though perhaps unlikely, it might potentially coincide with the 'first record of emergence'. Note that demergence might also refer to the complete extinction or disappearance of a pathogen in which case any subsequent reemergence should be rather unlikely. Many host-pathogen interactions will finally oscillate between the stages of demergence and reemergence.

The *de novo* evolution of a new pathogen and/or its "genuine emergence" might in fact remain unnoticed in most if not all cases. Hence, what is commonly referred to as "emergence" might

actually only be the "first record of emergence". It is therefore debatable if the expression 'newly originated', in order to refer to so-called new pathogens [165], is appropriate. Perhaps one should omit this expression in favour of either 'newly defined' or 'newly identified'. Emergence might actually just be due to newly available screening techniques, enhanced research efforts and/or broadened or improved surveillance strategies [168]. In fact, most emerging pathogens (or host-pathogen interactions) are believed to have "...existed previously in nature and in most cases have not changed their genetic structure but rather gained access to new host populations or immunocompromised hosts" [150] / [102] [125]. Thus, new ways of transmission [147] rather than new pathogens (or host-pathogen interactions) would be the cause of the majority of (re)emerging host-pathogen interactions. Looking at figure 1.2 it could be proposed that one part of the overall habitat-pathogen continuum is actually formed by the dynamic processes of emerging, demerging and reemerging transmission pathways. This part does not result in a continuous, omnipresent and ubiquitous network. Transmission pathways rather come into existence, disappear and reappear again in different places and contexts[12].

1.2.4 Host range of a pathogen

The *host range* of a "pathogen" depends on the taxonomic level of the pathogen. Henceforth it will be referred to the species level unless stated otherwise. The host range of a pathogen can then be given, for example, as the number of host species which are susceptible to (i.e. can be infected by) the pathogen in focus. One might also prefer to view the host on any other taxonomic level. A major shortcoming in terms of host ranges is that the literature rarely takes such quantitative approaches. Instead, most information on host ranges is of qualitative or semi-quantitative nature. For example, pathogens are classified as singlehost/specialist (infecting one host species only), multihost/generalist (infecting $>$ 1 host species), zoonotic (equivalent to multihost but must include the human host) etc. Pathogens assigned to any of the multihost classes could in fact vary considerably in their actual host ranges. For example, one multihost pathogen might be able to cause infection in two host species whereas the host range of another multihost pathogen could include hundreds of host species. Even if both multihost pathogens are able to infect the same number of host species could their actual host range still be of very different nature. That is if one of them is specialised to infect only members of a few taxonomic categories and the other is able infect hosts species belonging to various distinct families, orders etc. Nevertheless, attempts to provide a more detailed view on the multihost issue have been put forward already. For example, the authors of reference [145] suggested five distinct categories in terms of *host specificity*[13]:

1. **Species-specific:** infecting members of a single host species only;
2. **Genus-specific:** infecting members of a single host genus only but of $>$1 host species;
3. **Family-specific:** infecting members of a single host family only but of $>$1 host genus;

[12]This part of the habitat-pathogen continuum might be called a *transmitome* with reference to ideas derived from the field of neuroscience [44].

[13]That is the level of specialism or generalism of a pathogen.

4. **Order-specific:** infecting members of a single host order only but of >1 host family;

5. **Multi-order:** infecting members of >1 host order.

A major question associated with the host range of pathogens is if evolution would actually be in favour of either specialism or generalism. Some authors argue that evolution should generally favour specialism [207]. In fact, the authors of [207] suggest that specialism should be favoured particularly with respect to pathogens. According to their point of view pathogens are considered to be under selection pressure to coevolve with their hosts. Multihost pathogens would therefore face functional trade-offs which would make it difficult for them to reach a level of fitness equivalent to that of specialist pathogens. Consequently the evolution of a pathogen would proceed faster within narrower host ranges.

Interestingly even closely related pathogens can show very different host ranges. Woolhouse *et al.* [207] argue that pathogens may rapidly adapt to generalism or specialism throughout their evolutionary history as a result of "...powerful selective pressures in both directions" [207]. These selective pressures would act upon transmission-related features of the pathogen which influence both the opportunities for generalism and the costs of specialism. The pathogen might, for example, face reproductive costs in case there are no susceptible host species available. In order to keep such costs at a minimum, the pathogen would probably need to evolve towards increasing the likelihood of encountering and/or entering hosts. This might involve prolonged survival times outside a host and/or an enhanced ability to adapt to new host species.

Transmission-related features of the pathogen act upon the opportunities for generalism [207]. The ability to contact multiple hosts are mainly associated with events that help bridging the spatial distances between the pathogen's current whereabouts and potential receiver host individuals, populations and/or species. Close geographical proximities appear advantageous for obvious reasons. However, the pathogen might also develop the ability to travel increasing distances by exploiting environmental factors such as airflows or waterways. It could also be mechanically attached to birds, insects or any element involved in human traffic. Another option is that yet undiscovered host species participate in establishing contact.

So far the discussion has assumed that all susceptible host species contribute equally to the host range of a pathogen. However, the literature sometimes distinguishes so-called *maintenance* or *reservoir host* species from *accidental* or *dead-end host* species. The former category addresses host species (or rather populations) that have been established for any particular pathogen, i.e. the pathogen can persit in them independently [207][14]. In other words, transmission from sources other than the maintenance host species is not required. Host species belonging to the latter category are considered transient phenomena, i.e. the pathogen will sooner or later disappear from the host species or population if transmission from other sources is coming to a halt [207]. It now seems obvious to assume that a dead-end host "population" is likely to be characterised by

[14] "For many macroparasites the picture is more confusing as they often have complex life cycles, with different stages in different host species. The community of hosts therefore acts as the 'reservoir', although if the term 'reservoir' is used in the sense of 'source' of infection to a target host, e.g. humans, then only part of that community is often classed as the reservoir..." [11].

an $R_0 \leq 1$ which would not allow for extensive chains of transmission events thereby leading to the disappearance of the pathogen from this host population. Consequently a maintenance host "population" would be characterised by an $R_0 > 1$.

If the differences between host species indeed turn out to be of relevance, then we would need to reconsider the multihost label as we know it. In fact, the authors of [60] argue that so-called *spillover pathogens* and *apparent multihost pathogens* should be distinguished from *true multihost pathogens*. Spillover pathogens are allegedly those that occur only occasionally in any given host species. The *transmission rate*[15] within these very host species [16] as well as the transmission rate between any of these host species and the corresponding source(s) of the pathogen are too low to account for the pathogen's persistence in this very host species. Apparent multihost pathogens are sustained in a given host species but only because the transmission rate between this host species and the pathogen's source(s) is sufficiently high. The transmission rate within the given host species does not contribute to the pathogen's persistence. The idea goes that only the true multihost pathogens can be sustained in their corresponding host species without constant reintroduction from an outside source.

1.2.4.1 Host shift

The dynamic nature of hosts and pathogens and their interactions can lead to *host shifts* [60] [144]. In order for a pathogen to undergo a host shift, i.e. to emerge within a hitherto unaffected host species, it is reliant on inter-species transmission events in the first place. In the course of a host shift the pathogen might either broaden its host range it might turn it into a specialist pathogen for the new host species, henceforth being unable to cause infection in the initial donor species. In fact, it is also possible that it causes the extinction of newly affected host species[17]. A conceptualisation of such a conversion from a specialist pathogen of one host species into a pathogen that subsequently specialises on a different host species has been put forward in [203]. This conceptualisation was later adapted to the emergence of new pathogens in the human host in general [204]. Figure 1.3 tries to reconcile, complete and adjust the ideas of these two papers. I will start the discussion by assuming an initial source I of any particular pathogen and a given new (i.e. hitherto unaffected) host species N in which this pathogen - or one of its derivatives as we will see - is about to emerge. Let us assume that the source I is, for example, another host species. Figure 1.3 distinguishes five stages involved in a host shift:

[15] The transmission rate is proportional to the contact rate between susceptible individuals and the sources of infection (including so-called "direct" and "indirect" sources) and the probability that the contacts will lead to infection which depends on the infectivity of the pathogen and susceptibility of the host.

[16] It is questionable whether the authors of [60] did in fact mean to refer to host species or were rather referring to host populations.

[17] "Extinction by infection ..." [36] has not only "...been implicated in the local extinction of a number of species and the global (species) extinction" [36] but has also been "...definitely proven" [36] in at least one case. However, "pathogens" "...rarely have been implicated as a direct cause of host extinction risk relative to other factors such as habitat loss, hunting, and invasive species [...] Only 3.7% of 833 known species extinctions have been attributed in part or directly to infectious disease" [145].

① : The pathogen is neither present in N nor has N ever been exposed to it under natural conditions. However, it is present in I.

② : The pathogen has been exposed to at least one population of N under natural conditions. The exposure might cause minor, local infections in the individuals of N but these would neither lead to disease - or any noteworthy reduction of the host's fitness - nor do they enable the pathogen to transmit between individuals of N.

③ : The pathogen can cause severe infection in the exposed individuals in at least one of N's populations. However, it still lacks the ability to transmit between individuals of N, i.e. no secondary infection.

④ : The pathogen is now able to transmit between the individuals of at least one of N's populations. However, the extent of the transmission is restricted, i.e. there will be no exponential spread. It might still lead to shorter or longer sequences of secondary transmission[18]. Outbreaks will therefore either soon die out or might be sustained for a while. Reference [203] subdivided this stage with respect to limited and prolonged outbreaks.

⑤ : The pathogen is able to cause outbreaks that can spread exponentially in at least one of N's populations.

Transitions between these five stages are symbolised in figure 1.3 by arrows. Greek letters have been placed next to the arrowheads and refer to absolute, though perhaps not necessarily sufficient, conditions that need to be met in order for the corresponding transition to take place[19]. I distinguish four absolute conditions:

α : Transmission between the pathogen's source I and at least one of N's populations is required. In other words, contact of N with the pathogen needs to be established. This contact might be entirely of mechanical nature at first if no initial entry portal is available.

β : The pathogen requires (on average) a sufficiently high infectivity[20] and/or effective portal of entry with respect to individuals of N. In other words, the individuals of N require (on average) a sufficiently high or effective susceptibility with respect to the pathogen.

[18] Also called "stuttering" chains of transmission [11].

[19] Such absolute conditions have also been referred to as barriers [204] [203] [60] or thresholds [60], though perhaps with different implications.

[20] Ability of a pathogen to establish infection, i.e. to enter, survive and multiply within a host.

γ : This condition might involve either any one or both of the following two aspects:

1. The individuals of N require (on average) a sufficiently high or effective infectiousness[21]. This might include sufficient excretion of the pathogen by gaining access to certain body parts like the alimentary, respiratory or urogenital tract or the blood system etc. It might also include the time period within which individuals of N remain infectious, i.e. continue to excrete the pathogen. On top of that, it could involve pathogen features that enable the pathogen to stay infective outside the host thereby increasing the chances of non-close/indirect contact transmission.
2. The second aspect is a sufficient contact structure within the given population of N. It could be determined, for example, via behavioural patterns and/or density of individuals within the population.

δ : Sufficiently *enhanced* contact structure within the population(s) of N and/or sufficiently *enhanced* average infectiousness of the individuals of N.

As can be seen are the arrows in figure 1.3 both colour- and line-style-coded. The *colour red* refers to transitions which introduce the pathogen from the current source into a new host species. The *colour green* highlights those transitions that either take place within one or that might involve two populations of the new host species N. In the latter case the pathogen could be introduced into a hitherto unaffected population of N. It is important to note that this kind of introduction would not stem from an "outside" source like I anymore, but from within N. The *solid line-style* refers to transitions that involve the current, i.e. unchanged, pathogen. A pathogen might face significant alterations during a host shift which could be regarded as a *de novo* evolution, i.e. genuine emergence of a new pathogen. The host population carrying this novel pathogen could then constitute a new source, i.e. I_2, potentially initiating another host shift. The *dashed line-style* consequently refers to transitions leading to these kind of "ramifications".

Solid-red arrows address the introduction of a pathogen into the current new host species N via the current source I. Any required changes with respect to the pathogen's molecular assembly in order to meet any of the absolute conditions would need to take place within I and not within N.

[21] Ability of a pathogen to be transmitted to other hosts. In fact, this depends on the contact structure, too.

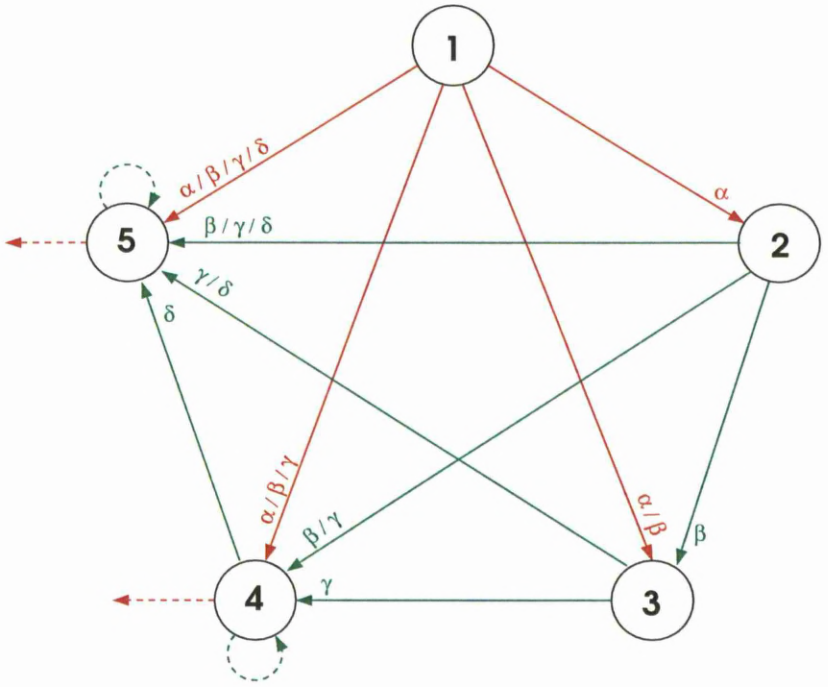

Figure 1.3: Suggested stages and conditions during host shifts: This framework distinguishes five transition stages that might be encountered during the processes involved in host shifts. The Greek letters symbolise absolute conditions to be met in order to allow for a transition to take place. The arrows stand for transitions. Detailed information on the colour and line-style code of the arrows, the absolute conditions and further information can be found in the text.

The set of absolute conditions that can be satisfied will determine the stage(s) that could potentially be entered. An introduction of the pathogen into N via I might therefore result directly in any of the stages (2), (3), (4) or (5). Once one of these stages has been reached, the opportunity for proceeding to any of the following stages might arise subsequently within N. This is symbolised by the *solid-green arrows*. The pathogen is henceforth marooned within N - initially perhaps in a single population - and might need to adapt to N relatively quickly in order to maintain its newly obtained position. If the absolute conditions in order to proceed from stage (2) or (3) to any higher stage cannot be met in due time, then the pathogen will sooner or later face extinction from N. The same might apply to stages (4) and (5) but the ability to transmit within N at these stages could enable the pathogen to finally prevail. Infections that lead to exceptionally lethal outcomes might form a different case.

Many host species are likely to be composed of a set of distinct populations. These populations could potentially differ considerably in relevant parameters such as behavioural patterns, geographic range, population size, density of host individuals etc. However, it does not seem likely that all of

N's populations will simultaneously undergo the same transitions at the same time. The pathogen is therefore supposed to thrive and progress within some of the exposed populations whereas in others it will remain in any of the stages or disappear. As a consequence, it could be suggested that each stage - except stage (1) - can exhibit a heterogeneous pattern of N's populations. In other words, any particular population of N does not necessarily need to be in the stage that has been assigned to the species of N. This calls for some clarifying remarks on the host shift stages:

(1) : All of N's populations belong to stage (1).

(2) : At least one population of N has entered stage (2). The rest of the populations remain unexposed.

(3) : At least one population of N has entered stage (3). The rest of the populations remain either unexposed or in stage (2).

(4) : At least one population of N has entered stage (4). The rest of the populations remain unexposed, in stage (2) or (3).

(5) : At least one population of N has entered stage (5). The rest of the populations remain unexposed, in stage (2), (3) or (4).

It has been already pointed out that the dashed line-style accounts for ramifications with respect to the genuine emergence of a new pathogen in the course of a host shift. Dashed lines have only been attached to stages (4) and (5). This was deemed reasonable since both stages are considered to have developed sufficiently effective ways of transmission, at least on the intra-population level of at least one population. The *dashed-red arrows* then refer to a situation in which one of N's populations is now viewed as a new source I_2. As a source, I_2 would be considered to be in stage (1) and potentially moving on to one of the next stages if a new receiver host species N_2 is available.

The *dashed-green loop-arrows*, on the other hand, indicate potential transitions involving only populations of N itself. So, rather than introducing the pathogen of I_2 to a new host species N_2, I_2 would introduce this new pathogen to a distinct population of its own species, i.e. N. The new pathogen has already undergone all the adaptations required for stages (4) or (5). Since the pathogen is not about to conquer a totally new host species, one could expect that any subsequent transition is mainly dependent on characteristics of the newly affected population of N. It could therefore be assumed that the so far unaffected populations of N should convert more easily to one of the higher stages, provided of course, that they get in contact and that their population-specific parameters are not particularly disadvantageous. Keep in mind that the source population I_2 remains either in stage (4) or (5). Therefore p.d. is the entire host species N considered to be either in stage (4) or (5).

A *de novo* evolution in the course of a host shift could therefore form the genuine emergence of a new specialist pathogen. In order to persist in any population of the so far exclusive host species, the pathogen is allegedly reliant on an $R_0 > 1$ [207]. Woolhouse *et al.* [204] suggested to express and quantify the distinction between certain stages during a host shift in terms of R_0. These stages would correspond most likely to stages (4) and (5) of our conceptualisation as presented in figure 1.3. Although R_0 is often used in the literature to refer to the transmission potential of a pathogen with respect to a whole host "species", I already emphsised that it is probably more accurate when referring to single populations only. In order to express the distinction between stages (4) and (5) with respect to R_0, one would need to take N's potentially heterogeneous pattern of populations

into account. If N is therefore assigned to stage (4), all of its populations are thought to exhibit a basic reproductive ratio of $R_0 \leq 1$. If N, however, is assigned to stage (5), then there has to be at least one population with a basic reproductive ratio of $R_0 > 1$. From this follows that in fact only stage (5) would have the ability to sustain the status of a specialist pathogen. Stage (5) and to a lesser degree stage (4) are on the other hand perhaps the most likely candidates for providing a source I (note: not I_2!) for another host species N (note: not N_2!) themselves, thereby initiating another host shift which bears the potential of "broadening" the host range. This consequently raises the question of how a pathogen can keep its specialist status under these circumstances. The minimum requirement appears to be that the pathogen should not conform to the absolute condition α, i.e. it needs to avoid the establishment of contact structures with additional host species. For example, the pathogen might inhabit a rather isolated host population which prevents both close/direct as well as non-close/indirect contact with most potentially new host species. The pathogen might also exhibit very low survival rates outside its host, lowering the chances of non-close/indirect transmission events even further. The chances of close/direct transmission could be lowered significantly if the pathogen would adapt itself entirely to transmission via sexual contact. On top of that, the chances of getting into close/direct contact with new host species could be expected to be even lower if the specialist pathogen inhabits a host species that is considered to be at the top end of the food chain with respect to natural wildlife situations. Such a host species is simply less likely to be consumed by other predatory species.

1.2.4.2 Zoonosis

According to references [45], [94], [142] and [164] the 'World Health Organization' (WHO) defined the term *zoonosis*[22] initially as a disease or infection which is naturally transmitted between vertebrate animals and humans [45] [94] [142] [164]. This definition suggests that the disease or infection can be transmitted from vertebrate animals (or more precisely non-human vertebrates) to humans (or more accurately human vertebrates) and vice versa [32]. However, a look at the WHO's current presentation on the world wide web reveals a clear restriction "...from vertebrate animals to humans" [208] only. This restricted view on the term zoonosis is also supported by Last [101] and the 'Institute of Medicine' [38]. An earlier definition by the 'Institute of Medicine' [102], although appearing to be very similar, makes instead no restrictions in terms of the underlying way of transmission (e.g. naturally transmitted) and also none with respect to the involved groups of non-human animals. Krauss *et al.* [94] are in line with these non-restricted views but rather prefer the idea that zoonotic transmissions can occur in both directions. Considering the direction of zoonotic transmissions, the terms *zooanthroponosis*, i.e. from animals to humans, and *anthropozoonosis*, i.e. from humans to animals, have been suggested [94]. Nelson [132] formerly used these very terms but with reference to the exact opposite directions. According to him does the term zooanthroponosis refer to those infections which are naturally but incidentally transmitted from the human maintenance host to other vertebrate animals [132]. Therefore does the term anthropozoonosis describe the reverse direction, i.e. from a non-human vertebrate being the maintenance host to the incidental human host [132]. The situation where both host groups, i.e. human and non-human animals, are considered as potential maintenance hosts in a dynamic exchange of infection between both of them has been termed *amphixenosis* [132]. Nelson's [132] view of anthropozoonosis is therefore interchangeable with the notion of the term zoonosis as provided by the 'Pan American

[22] *pl.*: zoonoses, *adj.*: zoonotic

Health Organization' (PAHO). The PAHO regards zoonosis as an infectious disease transmitted from non-human vertebrates to humans in which the non-human part plays "... an essential role in maintaining the infection in nature, and man is only an accidental host" [1]. Madigan *et al.* agree on the facts that a zoonosis is primarily occurring in non-human animals - though not limiting the term animal to vertebrates - and that it is only "occasionally transmitted to humans..." [105]. For this reason are Madigan *et al.* [105] and the PAHO closer to the "proposed" original meaning of zoonoses as being primarily animal diseases (*zoon*: Greek for 'animal') [94], which - according to Rudolf Virchow, who apparently coined the term in 1855 [17] - also *must* infect man [132][23].

In order to increase the level of confusion even further it should be mentioned as a closing remark, that it still seems a matter of debate if intoxications caused by, for example, snake and spider venoms or the botulinum toxin of *Clostridium botulinum* should be regarded as a zoonosis [142].

It is questionable whether the terms discussed in this subsection add anything conceptually useful. The limited definition of non-human hosts as 'vertebrates' is ecologically not useful but perhaps makes thinking about zoonoses easier for medics [11]. In this thesis the term zoonosis is simply referring to an infection that affects humans and at least one additional non-human host species.

1.2.5 Pathogen richness of a host

It can be generally assumed that hosts should have an interest in warding pathogens off or at least to hold them at bay since they still remain a potential cause of disease, reduction of host fitness and death. The authors of [151] state three distinct ways for a host species to acquire a pathogen species:

1. **Inheritance:** The pathogen species is apparently "... passed down"[24] [151] to a daughter host species from an ancestral host species during the evolutionary speciation of this ancestor. Inherited pathogen species will therefore be shared by closely related host species.
2. **Host shift**[25]**:** See subsubsection 1.2.4.1.
3. **Intra-host speciation:** This would constitute a *de novo* evolution or genuine emergence within a host, i.e. a pathogen species changes within a host, giving rise to one or more distinct daughter species or strains. Compare also 1.2.4.1.

Pathogen species, on the other hand, can also get lost from a host species. For example, the pathogen species might be unable to adapt to evolutionary changes of the host species [151]. The likelihood that a host species loses or acquires pathogen species has been assumed to be related

[23] Note that this point of view could be wrong since apparently did Virchow mean "... any animal transmissible disease" [11].

[24] The authors did not explain the actual process of passing down a pathogen species. It might involve some form of transmission, e.g. vertical transmission from the mother to the offspring (i.e. *in utero*/congenital).

[25] These authors prefer the term 'host switching' instead [151].

to ecological characteristics of the host species [151]. It is debatable whether this also applies to inherited pathogens since they would not originate from ecological or evolutionary processes [151].

Pathogen species richness does not necessarily need to be viewed on the host species level. In fact, there seems to be a multitude of possible organisational levels with respect to the host. Depending on the context or research interest one might, for example, focus on[26]:

- Any single host individual as a whole [151] [98];
- Any single host population as a whole [151] [98];
- Any single host species as a whole [151] [98];
- Any defined set of host individuals as a whole;
- Any defined set of host populations as a whole;
- Any defined set of host species as a whole [98];
- Any defined anatomical structure (e.g. the gastrointestinal tract etc. [151]):
  - Within any single host individual;
  - Within any single host population;
  - Within any single host species;
  - Within any defined set of host individuals;
  - etc.
- Any defined host community or ecosystem;
- etc.

In order to estimate the pathogen (species) richness of hosts, the literature suggests various frameworks. Such frameworks were, for example, derived from:

[26] It should be noted that this list is based on suggestions put forward in [151] and [98]. The actual terminology, however, appears to be slightly ambiguous and has therefore been largly omitted here. Take, for example, the term *parasite fauna*. Poulin defined it as "...the sum of the parasite species found in the various populations [...] that make up the host species" [151]. It is debatable if the author is in fact referring to an actual "sum" in this context. Alternatively he could be referring to the set-theoretical "union", i.e. the number of "distinct" parasite species found in the entirety of populations with respect to one host species. On the other hand, the same reference also states that parasite fauna is a "...level of organization of parasite assemblages" [151]. These assemblages actually "...refer to all parasite species found within one microhabitat in or on the host (e.g. the gastrointestinal tract [...])" [151]. Therefore, it seems possible that the author is referring only to particular anatomical structures of a host rather than a host as a whole. Note also that the term parasite fauna has been used differently elsewhere [98]. Here it can either refer to a single host species or to multiple interacting host species.

1. The theory of island biogeography;
2. Epidemiological R_0 models;
3. Biogeographical patterns.

The first framework is, as the name suggests, based on biogeography which is the study of the distribution of biodiversity [154]. Biodiversity can be expressed, for example, by the number of species over space and time [154]. Hosts[27] can be regarded as island in terms of isolated habitats [98]. Isolation would then require transmission in order to introduce pathogens to the host-islands [154]. Infection of a host with a pathogen would correspond to a successful immigration of a new species whereas the loss of a pathogen would correspond to the extinction of a species. Extinction rates (number of species that go extinct/unit time) rise with the number of species present on the island and fall with the area of the island. Immigration rates (number of immigrated species/unit time), on the other hand, fall both with the number of species present on the island and with the island's distance from the potential source(s) of new (pathogen) species [154]. It has been proposed that the number of species on a given island should reach an equilibrium when these two rates equal each other [154]. The rates themselves are allegedly influenced by island (i.e. host) characteristics such as body-size, host complexity, age and/or lifespan, population density and/or size, geographical ranges, variety of diet, behavioural patterns etc. [151] (see also Kuris *et al.* [98]). In fact, the framework qualitatively predicts that host species average body size, average lifespan, average population density, its geographical range and its variety of diet [151]. Note that such characteristics are usually highly correlated[28] which poses a considerable challenge in terms of identifying chains of cause and effect between these features [151]. It is questionable if such predictions can be expanded to microparasites such as viruses and bacteria.

The application of the island metaphor to host-pathogen interactions might neglect specific aspects of the host biology which distinguish hosts from geological islands. In order to allow for predictions on pathogen richness based on the theory of island biogeography, the theory would need modifications which accommodate aspects of the host biology. The actual modifications might also depend on the considered organisational level of the host [98], i.e. islands might refer to host individuals or host populations etc. To simplify matters, let us assume that an island corresponds to a particular host individual. What aspects of the host biology would not be covered by the island metaphor? For instance:

- The inter-island distance (i.e. distance between the host individual in focus and other host individuals) might be correlated to temporal fluctuations of the host density [98];
- The inter-island distance could be correlated to seasonal differences in the behaviour of the host individual and also other individuals;
- Growth and aging processes of the host individual [98];

[27] In fact, individual hosts, host populations or an entire host species etc.

[28] For example, host species with larger average body sizes tend to have longer lifespans and tend to occur in lower population densities [151].

- The pathogen species richness might never reach the suggested equilibrium between immigration and extinction rates due to short lifespans of host individuals [98];
- Changing characteristics of host individuals due to the presence of pathogen species [98]. Such characteristics might even involve behavioural patterns influencing the inter-island distance or the distance between an island and any other infectious source. Manipulation of host behaviour via "parasites" is a well known fact [23] but perhaps more commonly recognised in the context of helminths so far [99]. Nonetheless, "microparasites" such as viruses and bacteria are not only able to induce morphological and/or physiological changes in hosts but might also bear the potential to have an effect on behavioural patterns[29] [19] [197].
- Immune response, i.e. the host individual will actively pursue the prevention of pathogen immigration [151];
- Vertical transmission [11].

It has been pointed out that the identification of suitable modifications would require sufficient data on pathogen species richness in the first place, preferably on all interacting organisational host levels [98].

The second framework is based on epidemiological R_0 models. It basically suggests that populations with characteristics favouring high R_0 values ($R_0 > 1$) would be easier to colonise by new pathogens. As a result are such populations expected to exhibit an increased pathogen species richness [151][30].

The third framework has been derived again from biogeography. Organismal diversity is known to follow a latitudinal gradient with higher diversity in the tropical regions [151]. One might consequently expect a higher pathogen diversity towards tropical regions. However, since there is a general increase of organismal diversity towards the tropical regions, the actual ratio of pathogen species per host species might not change at all [151].

I already highlighted the issue of accidental *vs.* maintenance hosts with respect to the host range of a pathogen. Similarly, should an accidental pathogen species contribute equally to the pathogen species richness of a host as an already established pathogen species? Moreover, one might argue to state the number of pathogen species that have ever been reported to cause infection in a particular host species. This does not seem to account for the dynamics of immigration and extinction over time. Last but not least, one might also need to account for differences between individual hosts, between host populations etc.

As a closing remark it should be mentioned that the pathogen species richness of hosts might also be influenced by direct or indirect in-host interactions between distinct pathogens. This characteristic also seems to apply to geological islands, i.e. animal and/or plant species will interact, compete, feed on each other or set up mutualistic relationships (e.g. symbiosis) etc. In fact, coinfections of hosts

[29] Rabies, for example, makes foxes travel further distances and bite other hosts they would not normally meet [11]. Diarrhoea, coughing, sneezing etc. can also be viewed as behavioural changes [11].

[30] R_0 was actually used in [151] with reference to host species rather than populations.

have been viewed as common in natural systems [144]. It has been argued that they are abundant and relevant *in vivo* for many viral systems [138]. Schnitzler *et al.*, however, view the "... double infection with two different strains of influenza virus in a single host" [163] as a rare event. Then again, certain hosts or rather host populations might be more prone to coincidentally acquire infections from multiple pathogens. "Following interspecies transmission [...] an influenza virus may undergo many pig-to-pig transmissions because of the continual availability of susceptible pigs ..." [18]. The pig "... is the only domesticated mammalian species which is reared in abundance and is susceptible to, and allows productive replication of [...] influenza viruses from mammalian and avian hosts" [18]. In other words, the pig appears to have a broader richness regarding influenza viruses. This is highlighted in figure 1.1 by the fact that pigs can be infected by both the avian H7N7 and the swine H1N1 influenza A virus, though with different abilities to spread among pigs. In the "... event of a double infection with two different strains of influenza virus in a single host, reassortment of the genome sequence might occur, producing a series of completely novel combinations of genome segments in the progeny viruses. Such reassorted strains maybe the source of new pandemic influenza variants" [163]. Pigs are therefore being viewed as "... the leading contender for the role of [...] a mixing vessel for reassortment between influenza viruses from mammalian and avian hosts with unknown implications" [18]. The half-red, half-green circle in population P2 of figure 1.1 is meant to address such in-host pathogen interactions. Furthermore, pathogens might also interact in other ways within the host. Coinfection with distinct strains of foot-and-mouth disease virus, for example, that differ in their virulence (i.e. cell-killing ability and/or ability to invade and colonise host tissues) can result in an overall decreased virulence due to processes of competition and interference between the distinct viral strains [138].

1.3 Related work: Facts and figures

"Nobody can check everything, we're all interdependent for information ... "

Ben Goldacre [69]

As I am about to explain in the subsequent chapter is the data of this thesis mainly concerned with the transmission of infectious viruses and bacteria towards certain livestock animals. Nevertheless, an extension to other pathogens as well as host species is generally possible and of course desirable. Consequently, although the literature provides additional information on pathogens such as protozoa, helminths, fungi etc., this information has been largely omitted. However, I did compare certain data on these pathogens with data on viruses and bacteria in order to provide a justification for focusing on the latter in the first place.

1.3.1 Transmission

1.3.1.1 Close/direct contact

Let us begin the review on transmission-related facts and figures by looking at 'close/direct contact' transmission. A study on wild primate hosts including "pathogens" such as viruses, bacteria,

protozoa, helminths and arthropods indicated 'close/ direct contact' transmission for 42% of all recorded "pathogen" species [2]. A similar study, which additionally included fungi, subdivided the 'close/direct contact' category into three subcategories [143]:

- **Close non-sexual contact:** accounted for 185, i.e. 44.8%, of all "pathogen" species;
- **Sexual contact:** accounted for 20, i.e. 4.8%, of all "pathogen" species;
- **Vertical transmission:** accounted for 20, i.e. 4.8%, of all "pathogen" species.

These figures add up to 54.2% for the overall 'close/direct contact' category. However, these subcategories were not supposed to be mutually exclusive, i.e. pathogens could have been assigned to more than one subcategory including the 'non-close/indirect contact' category. The true percentage could therefore be expected to be somewhat smaller and consequently closer to the previously stated value of 42%. In fact, looking at the figures of [143] in detail the true percentage for the 'close/direct contact' category can be assumed to be in the range of [44.6, 54.2][31]. Thus, studies [2] and [143] are roughly matching up but they both appear to be at odds with the finding of yet another study. A significantly higher proportion of "pathogen" species, in fact 75%, had been assigned to the 'close/direct contact' category in study [145]. The included range of pathogen types in [145] is in accordance with [143]. However, [145] merely aimed at "pathogens" of endangered mammal species that were also documented either to be a cause of population declines or that were known causes of considerable host fitness reduction. Provided that all three works are representative and of sufficient quality, what could have caused this difference? In fact did study [145] take a very different approach. Three parameters could have caused the discrepancy:

- The included variety of hosts. Whereas papers [2] and [143] are looking at primate hosts only, paper [145] broadened the focus to mammalian hosts in general;
- A label required to be assigned to the hosts, i.e. a mammalian host was required to be officially registered as an endangered species;
- A label required to be assigned to the pathogen, i.e. a parasite needed to be a known cause of host population decline and/or of host fitness reduction.

[31] 34% of all pathogens, i.e. 141 pathogens out of 415 in total, are assigned to more than one category [143]. If none of them are to be found in the 'close/direct' subcategories, which is mathematically possible, then the upper boundary would be 54.2%. On the other hand, the maximum number of pathogens to be assigned to more than one category within the 'close/direct' categories cannot exceed 40, i.e. 20 pathogens are assigned to both 'close non-sexual' and 'sexual' and another 20 are assigned to 'close non-sexual' and 'vertical'. From this follows 225-40=185 and (185*100)/415=44.6.

1.3.1.2 Non-close/indirect contact

Altizer *et al.* [2] state that 70% of all pathogens fall into the 'non-close/indirect contact' category but this includes the categories 'transmission via arthropod vectors' and 'parasites with intermediate hosts'. Pedersen *et al.* [143], however, made a distinction between these three categories. They assigned 175 distinct "pathogens" (42.2%) to the 'non-close/indirect' category, 132 (31.8%) to 'transmission via arthropod vectors' and 59 (14.2%) to 'transmission via intermediate hosts'. The resulting sum is 366 (88.2%). Yet, the actual proportion of the 'non-close/indirect contact' category can be expected to be in the range of [42.2, 63.9][32].

With respect to pathogens that pose a risk in terms of population declines in wild mammals, 48% were assigned to the three 'non-close/indirect contact' categories. 45% were actually able to exploit both 'close/direct contact' as well as 'non-close/indirect contact' [145]. This also means that only a "...relatively few threatening parasites" [145], i.e. 3%, relied entirely on 'non-close/indirect contact' transmission. This could lead to the hypothesis that "pathogens" require ways of 'close/direct contact' transmission and perhaps also close contact structures in order to have a severe impact in terms of host population declines. However, keep in mind that the host species of this study were, at the same time, restricted exclusively to endangered mammal species in the first place.

1.3.2 Emergence

Table 1.1 compares the findings of four papers on emerging human pathogens. More precisely, the papers focus on emerging human-'pathogen species' interactions. Overall, viruses and bacteria were the most likely taxonomic groups associated with emerging infectious diseases (EIDs).

[32] 34% of all pathogens, i.e. 141 pathogens out of 415 in total, are assigned to more than one category [143]. At least 101 of these multi-category pathogens would be found in the 'non-close/indirect' category, since 40 can be maximally assigned to the 'close/direct' one. These 101 multi-category pathogens need to be assigned to at least any two of the 'non-close/indirect' categories, which leaves 164 of the 336 assignments to pathogens assigned to only one of these categories. From this follows (101+164)*100/415=63.9, i.e. the upper boundary. The lower boundary follows from assuming all 141 multi-category pathogens to fall into the 'non-close/indirect' category and also looking for the maximal number of pathogens to be assigned to either all three and then two of the subcategories. 59 could be assigned to all three, 132-59=73 to two and 43 to only one subcategory, i.e. (43+73+59)*100/415=42.2.

reference	year of publication	# total pathogen spp.	# emerging pathogen spp.	% virus spp.	% bacterial spp.	% fungal spp.	% protozoan spp.	% helminth spp.
[182][33]	2001	1415	175	44	30	9	11	6
[206]	2005	1407	177	44	31	12	8	6
[204]	2007	1399	87	68[34]	13	15	3	1
[85][35]	2008	335	335	25	54[36]	6	11	3
range				[25,68]	[13,54]	[6,15]	[3,11]	[1,6]
average				47	34	11	7	4

Table 1.1: Emerging human pathogen spp.: Note that references [182], [206] and [204] assigned also prions to the group of viruses (the proportion of prions, though, is negligible [28]). **Column one:** References as listed in the bibliography. **Column two:** Year of first publication. **Column three:** Total number of included pathogen species. **Column four:** Total number of emerging pathogen species. Note the darker shade of grey. **Columns five to nine:** The proportions of emerging pathogens by taxonomic group. **The range:** Smallest and largest value encountered with respect to the corresponding column above. **The average:** Rounded arithmetic mean of the two range values only. For further details see text and also compare table S2 in the supplementary information of reference [85].

On closer inspection of table 1.1 one can spot substantial differences in particular with respect to the proportion of emerging viruses and bacteria. The results on viruses and bacteria as presented in [182] [206] [204] are in agreement with an often quoted parameter that has been sus-

[33] These results have been published also in [27].

[34] The authors in fact state 75%, which seems incorrect given the data listed in the reference's table 1.

[35] In this work the authors also state additional figures that I view as incorrect and therefore chose to omit. Of all emerging pathogens 37-44% account for viruses whereas 10-30% account for bacteria. These ranges are apparently based on references [27] [182] [206] which are likely to be the result of a misinterpretation of reference [206]. In there, 37% of all viruses and 10% of all bacteria (not of all emerging pathogens!) have been reported as emerging. Be aware that a similar misinterpretation of [206] also led to wrong figures in table S2 of the supplementary information of [85]. The correct figures can be found in row two of table 1.1 referring to reference [206].

[36] The authors state 54% in the text of this publication and are referring to table S2 of the supplementary information in which they in fact state 49.3%.

pected to be related to patterns of emerging host-pathogen interactions. This parameter is the spontaneous rate of mutation which is assumed to be positively correlated with the likelihood of emergence [27] [143] [204] [206] [207] and also with the genetic variability and adaptability of a pathogen. From this point of view should viruses outperform bacteria [182] [206] [204]. Estimates of spontaneous mutation rates for viruses ranged roughly between [0.1,1] per genome per replication. Although virus genomes consist of either DNA or RNA, the genome of bacteria contains strictly DNA. Then, microorganisms with a DNA genome were estimated to have a considerably lower rate of roughly 0.0033 per genome per replication [52] [53] [54]. However, note that the estimation of spontaneous mutation rates remains a very active field of research and present estimates might change as the field is to progress.

The previous findings contrast considerably with the results of study [85]. In this study bacteria provide by far the most of the species that were considered as emerging. Genetic variability and adaptability of the pathogen caused by higher rates of spontaneous mutation does therefore not account for this observed pattern of emerging host-pathogen interactions. However, Lederberg *at al.* suggested that "...new variants of nonviral pathogens, such as bacteria, would be more common than new forms of viral pathogens since nonviral organisms are less constrained by host requirements" [102]. Moreover, bacteria are generally considered to be more sustainable outside the host which could provide a means of contact establishment that many viruses are less likely to take advantage of.

The observed discrepancies might be due to some form of bias. To start with, the time period when the studies have been conducted is a potential source of bias. Both quantity and quality of available data might change as research and evolution advances. The year of first publication has been consulted as an approximation of this time period. All four studies fall roughly into the same frame of time, i.e. 2001 to 2008. Therefore it can be assumed that the authors should have had access to more or less equivalent data.

In order to gain an insight on the comprehensiveness and balance of the studies, the total number of included pathogen species as well as the proportions for each pathogen type were considered. References [182] [204] [206], however, were in good agreement regarding both of these issues[37]. They incorporated either 14 or 15% viruses, either 38 or 39% bacteria, either 22 or 23% fungi, either 4 or 5% protozoa and all three included 20% helminths. The study of Jones *et al.* [85], however, differed considerably with regards to both the total number of incorporated pathogen species and the proportion of pathogen types. The corresponding figures can be taken directly from table 1.1.

So far this does not explain the increased proportion of emerging viruses and the decreased proportion of emerging bacteria observed in [204]. Therefore one needs to examine the applied concepts of the term emergence next. The definition of emergence of studies [182], [206] and [204] included a time-related cut-off regarding the first record of human-pathogen interactions. These three references roughly agree on this point in time. A human-pathogen interaction was viewed as emerging if it was first reported from 1980 [204], 1981 [182] or 1985 [206] onwards [38]. They also included

[37] All three studies might in fact be based on the same initial dataset. The dataset used in study [206] is definitely an updated version of the one created for [182].

[38] [182] and [206] state the last 20 years as the corresponding time frame which results in 1981 and 1985 given the year of publication.

such human-pathogen interactions for which the corresponding disease had actually been reported before the established cut-off time if the causative pathogen was firstly discovered after the cut-off time. However, [182] and [206] only included such human-pathogen interactions if the discovery of the causative pathogen was accompanied by an increase in incidence after the cut-off time. It is debatable if this accounts for the corresponding discrepancies. In fact, it seems more likely that the following aspect holds the key to explain the observed differences. Studies [182] and [206] considered an increase in incidence after the cut-off time as sufficient in order to label a host-pathogen interaction as emerging. In other words, these studies p.d. allowed for "reemerging" human-pathogen interactions. Nevertheless, it is likely that even study [204] included a certain amount of reemerging host-pathogen interactions. Woolhouse *et al.* [204] already pointed out that newly emerging human-pathogens interactions could have been around for an indefinite time, oscillating between demergence and reemergence. Perhaps they have been associated only lately with a known disease or they have been reported only recently for the first time[39].

The approach of reference [85] is very different in certain respects and deserves a separate inspection. Note that some of the following statements have been drawn from the reference's supplementary information. The most obvious difference is that this approach focused entirely on emerging "human-pathogen interactions" from the outset. In fact, the underlying database is said to contain so-called *EID events*. Therefore, key to understanding the differing results of this paper could be the concept of an EID event itself. The definition of an EID event apparently broadly followed the definition of "emerging pathogens" as presented in [182] and [206][40] Anyway, an EID event has been defined as "...the first temporal origination of an EID (that is, the original case or cluster of cases representing an infectious disease emerging in human populations for the first time [...])" [85]. Given this definition and the fact that the database is also said to consist of EID's, it seems that an EID event is meant to be a specific kind of an EID. From that point of view it can be assumed that the database actually consists of EID's in the first place. The authors state that such an EID can be a diseases that:

- "...has recently increased in incidence, impact or geographic range" [85];
- "...is caused by a pathogen that has recently evolved or entered the human population for the first time" [85][41];
- "...has occurred previously, but is increasing in incidence" [85];
- "...has occurred previously, but is [...] expanding into an area in which it has not previously been reported" [85];

[39] The term 'emerging' is best replaced by 'first record of emergence'. Consequently should the term 'reemerging' be replaced by 'rediscovered' or 'reportedly reemerging'.

[40] It is also mentioned that criteria of a few other sources have been applied too [?].

[41] Jones *et al.* [85] included newly evolved drug-resistant strains. It is questionable if this could have caused the discrepancy since Woolhouse *et al.*, for example, also included "...any recognized variant" [206] of a pathogen species which refers to drug-resistence too [205]. This can also be taken from reference six of [206].

- "...has occurred previously, but [...] which has significantly changed its pathological or clinical presentation" [85];

It should be noted that we do not know how Jones *et al.* defined the term 'recently' in this context. Perhaps they were referring to the last twenty years with respect to the start of the project [182] [206]. However, EID's were not completely restricted to the last twenty years since the database is also said to contain data from 1940 to 2004. Details put aside, this database was then probably analysed in order to identify 'EID events', i.e. the "...origins of EIDs" [85]. In other words, the EID events constitute an actual result of this work and are therefore a subset of the EID's that were collected for the database.

Keep in mind that the preceding explanations are based on my interpretation of the paper's content. In fact, the paper also holds evidence for different interpretations. The major point is that I deem the suggested concept of an "EID event" - as presented in [85] - as largely impenetrable.

It is sometimes difficult to figure out if a reference is actually referring to emerging infectious diseases, emerging pathogens, emerging host-pathogen interactions (i.e. emerging infections) or, indeed, what is meant by the terms "host" and "pathogen". These concepts are not mutually exclusive but can make a huge difference when it comes to the interpretation of results. Nevertheless, it has been suggested that viruses are a clear risk for disease emergence in humans and domestic mammals [27]. Moreover, among livestock pathogens, among OIÉ-listed pathogens as well as among human pathogens, RNA viruses were more likely to emerge than DNA viruses [27]. In fact, the ssRNA viruses might be of particular importance. For example, they made up the largest subset of new species to infect humans [204] and 70% of all the virus species that are causing declines in endangered mammals [145]. However, the latter result did not differ significantly from the percentage of ssRNA viruses in a comprehensive (i.e. not restricted to threatening parasites) database of mammalian "parasites"[42] [145]. Furthermore, multihost and multi-order pathogens were overall more likely to emerge than singlehost and single-order pathogens [27], a trend that was also discovered in [206]. This seems to be in disagreement with some of the figures of tables 1.2 and 1.1 as we are about to see in the following subsection. Here, helminths and protozoa were the pathogens most likely to be labelled as multihost but were the least likely to be labelled as emerging. This discrepancy might be due to an insufficiency of the multihost label. It also could have been caused by focusing on different host groups. For example, the results of table 1.1 were derived from studies focusing on the human host only. Table 1.2 is in contrast based on results derived from humans, domestic non-human hosts as well as wildlife hosts. As a final remark it has been shown that those pathogens of humans and domestic mammals which can also affect wildlife hosts are more likely to emerge [27].

1.3.3 Host range of a pathogen

In a study on pathogens (including viruses[43] , bacteria, protozoa, helminths and fungi), which focused on host species of three general host categories (i.e. humans, domestic livestock and domestic

[42] Global Mammal Parasite Database, www.mammalparasites.org; [136]

[43] Viruses included prions but the amount of prions was negligible small [28].

carnivores), the multihost category accounted for 63% of all pathogen species. Moreover, 58% of all these pathogen species affected members of at least two of the three listed host groups [27]. A more detailed perspective on the multihost issue based on an updated version of this reference's dataset [28] is given in table 1.2. Here, with 88.6% helminths were the group of pathogens with the highest proportion of multihost species. These proportions might change when looking closer at the variety of affected hosts. The table therefore distinguishes host species of four general categories, namely humans (H), domestic species (D), domestic ungulate species (U) and wildlife species (W). Interestingly, none of the highest scoring proportions is still presented by helminths. In fact, bacteria scored highest in the HDW category (i.e. infecting hosts from these three categories), viruses both in the HW and the HU category, fungi in the HD category and protozoa in both the DW and the D category. Note also that the vast majority of pathogen species of all pathogen types infects hosts of at least two host categories. A significant proportion is actually able to infect hosts of three categories (HDW).

Another study concentrated on wild primate species [143]. The authors report helminths to be the most host specific pathogens with only 52% (85 out of 163) multihost species. Moreover, both bacteria (29 out of 32) and fungi (9 out of 10) contained 90% multihost species, viruses 87% (71 out of 82) and protozoa 72% (59 out of 82). The updated dataset as provided by [28] also contained data on primates. All viruses, bacteria, helminths, protozoa and fungi infecting these primate hosts exhibited a multihost proportion of 100%.

	helminths	protozoa	viruses	bacteria	fungi
# total spp.	499	145	312	633	329
% multihost spp.	88.6	71.0	67.3	50.4	39.5
# multihost spp.	442	103	210	319	130
% HDW	33.7	21.4	25.2	37.9	35.4
% HW	23.8	16.5	30.0	14.4	12.3
% DW	27.8	36.9	17.1	3.4	0.0
% HD	6.6	7.8	6.2	27.0	44.6
% HU	2.3	1.0	16.7	8.2	0.0
% D	5.7	16.5	4.8	9.1	7.7

Table 1.2: Multihost pathogens: The figures of this table were derived from a recent version of the dataset that has been created for study [27]. This updated dataset was kindly provided by Cleaveland and Taylor [28]. The viruses included prions but the amount of prions was negligibly small. The percentages shown in the second row refer to corresponding figures in the yellow cells above. For example, 88.6% of all helminth species turned out to be multihost pathogens. All remaining percentages refer to the figures of the second yellow row. For example, 33.7% of all multihost helminth species were able to infect hosts from three host categories, i.e. H, D and W (HDW). The **host categories** were: **H**=human species, **D**=domestic non-human species (cattle, sheep, goats, pigs, horses, dogs, cats), **U**=domestic non-human ungulate species (cattle, sheep, goats, pigs, horses) and **W**=non-human wildlife species.

Study [143] also holds evidence that ca. 68% [44] (250 out of 369) of all primate pathogens (including species of viruses, bacteria, protozoa, helminths and fungi) are multihost pathogens, many of which were also not restricted to primate hosts. In fact, most bacterial species that were able to infect multiple primate species could also cause infection in non-primate hosts from multiple orders. Viruses were even dominated by multi-order pathogens, i.e. infected hosts from multiple orders. Strikingly, the majority of the multi-order and also multi-family viruses turned out to be RNA viruses, whereas DNA viruses were more evenly distributed with respect to these host specificity categories (i.e. species-specific, genus-specific, family-specific etc.). Broader host ranges in RNA viruses than in DNA viruses were also acknowledged in [204]. The data provided by [28] revealed that only 39% of all DNA virus species but 77% of all RNA virus species are multihost pathogens. Trends alike had been proposed already and were based on the idea that a higher mutation rate - in RNA viruses ca. 300 times higher than in DNA viruses [52] [54] - should be more likely to result in a broad host range because it would facilitate the adaptation to new host species [207] [85] [204].

A study on both endangered mammal species and threatening[45] "pathogens"[46] assigned the multihost label to all included "pathogen" species [145]. Furthermore, even the most host-specific "pathogen" species[47] that the authors came across were still family-specific, i.e. infected hosts from > 1 genus. 66% of all "pathogen" species were actually able to infect species from multiple host orders which in some cases included non-mammalian hosts [145].

Moreover, the authors of [145] discovered that viruses and bacteria with broad host ranges, which in most cases included domesticated animals, posed the main threat to the extinction of wild mammals. It is also interesting to note that pathogens which are relying entirely on 'close/direct contact' transmission, more precisely sexual or vertical transmission, tended to be highly host specific [143]. Such pathogens "... may have limited exposure to multiple species and thus transmission modes that decouple host-to-host contact (i.e. waterborne or soilborne transmission) will increase the opportunity for between-species transmission" [60].

An often used specification of the multihost label is the zoonotic label. Two closely related literature surveys on human pathogen species[48] estimated 58% [206] or 61% [182], respectively, to be zoonotic. Reference [182] further assigns the zoonotic-label to 31% of all bacterial species and to 19% of all viral species (including prions). A study taking humans, livestock and wildlife pathogens[48] into account found 42% of all recorded pathogen species to be zoonotic [27]. Furthermore, the authors of this study assigned the zoonotic label to 66% of all multihost pathogens, to 70% of all pathogen species affecting domestic mammalian carnivores (dogs and cats) and to 39% of all pathogen species affecting domestic mammalian livestock species (i.e. ungulates: cattle, sheep, goats, pigs and horses). 44% of the zoonotic pathogen species were able to affect domestic mammalian carnivores and wildlife hosts whereas 28% affected domestic mammalian livestock and

[44] This proportion is not to be confused with the value stated in the reference itself, i.e. 68% (282 out of 415), which includes ectoparasites such as arthropods [143].

[45] That is, documented threats to population sizes and/or host fitness.

[46] Including viruses, bacteria, protozoa, helminths, fungi and arthropods.

[47] This group made up 9% of all included "pathogen" species.

[48] Including viruses, bacteria, protozoa, helminths and fungi.

wildlife hosts [27]. Within the group of pathogen species that had been characterised as emerging 64% were zoonotic [27].

Rather than continuing to extend the catalogue of figures on the host range of pathogens, it is more important to highlight the key point. That is, the majority or at least a significant amount of pathogen species do not seem to be restricted to a single host species. In fact, many pathogens even seem to be able to utilise hosts of distant taxonomic groups. This clearly contradicts the previously mentioned idea that the evolution of pathogens should favour specialism.

1.3.4 Pathogen richness of a host

The pathogen richness of hosts is not well covered in the literature, particularly in terms of quantitative data. A possible reason for this could be the hitherto prevailing one-host-one-pathogen framework applied in many studies. The need to extend this framework and to look at multihost-multipathogen communities has been addressed already [144]. Nevertheless, it has been demonstrated for wild primate populations that pathogen species richness increased with the population's density and with its geographic range [143]. Such trends were in fact predicted earlier in a work by Poulin [151] who derived his ideas from the theory of island biography. As already mentioned did this author as well as Kuris *et al.* [98] further predict that the host's body size should covary with its "pathogen" (at least referring to helminths) species richness. He finally concludes that, although such a "...relationship indeed exists" [151], it is relatively weak. More precisely, larger animal species[49] harbour, on average, more pathogen species than smaller animal species but this trend is mainly explained "...by inheritance of ancestral parasites and not by more recent processes linked with their body size per se" [151]. However, a study on wild primate populations including 117 distinct primate species has shown that both "...population density and body mass [...] correlate significantly with parasite[50]richness" [2]. Moreover, numerous studies supposedly indicated that age - which ironically is often correlated with body size - was positively correlated with pathogen species richness and pathogen density [98].

1.4 Aims of the thesis

Transmission events is the recurring issue throughout the introduction. This thesis aims at zooming in on the transmission issue. More precisely, rather than using the previously outlined transmission categories, I intended to increase the level of granularity by extracting basic information from the corresponding literature. The idea is to create so-called transmission routes of pathogens. These are literature-based representations of (assumed) real world transmission pathways.

Section 1.3 highlighted the special status of viruses and bacteria in the context of infectious diseases. Therefore, it was decided to concentrate on these types of pathogens.

[49]Here, mammal species.

[50]The study included viruses, bacteria, protozoa, helminths and arthropods.

The aims of the work described in this thesis were:

- To develop a framework that enables clear categorization of real world or rather literature reported transmission pathways;
- To create two datasets of transmission routes based on the framework that could be interrogated in order to ...
- ... investigate patterns in pathogen transmission that might provide insights into the emergence or reemergence of host-pathogen interactions.

Chapter 2

Materials and methods

2.1 Data representation: The transmission framework

> *"Our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature* [...] *But* [...] *is not something we are normally aware of* [...] *The essence of metaphor is understanding and experiencing one kind of thing in terms of another* [...] *The concept is metaphorically structured,* [...] *and, consequently, the language is metaphorically structured ..."*
>
> Lakoff and Johnson [100]

Currently the available data on transmission events is mainly of qualitative nature. The so far suggested and applied conceptualisations, as summarised in subsection 1.2.2, are not of sufficient granularity to provide detailed information about the elements potentially involved in transmission events. To recap, a transmission event is the process by which a pathogen exits a sender host, is then transmitted between sender and receiver host and finally enters a receiver host. As was pointed out in figure 1.1 these events can involve, for example, environmental factors, people working with animals or the equipment used in their rearing and management. Therefore, in order to address questions concerning the exact course of events involved in establishing contact between a pathogen leaving a donor host and a yet unexposed host, it is necessary to apply a more refined representation of transmission events. Deciding on such a representation or conceptualisation constitutes the first step towards a structured literature review. This conceptualisation will be referred to as the transmission framework. Ideally it will capture and structure the basic concepts and terms encountered in the corresponding literature in a meaningful way. In fact, the literature information on transmission events can be expected to be highly conceptualised in the first place. Certain concepts might therefore be inappropriate, redundant or even in conflict with each other. The framework thereby implicitly conceptualises other concepts and could be viewed as a concept of concepts. It will eventually consist of a finite set of terms which denote concepts that are perceived

as being relevant for the transmission of pathogens[1]. Henceforth these basic concepts will be referred to as items of the transmission framework. These items are described both by means of a definition (or rather description) and via a set of structuring rules which form links and relations between the items. In this sense does the transmission framework resemble an *ontology* specified for the field of pathogen transmission [7] [13]. The term ontology itself has historically various meanings [13] [112]. Nevertheless, it is nowadays often used as "...a formal way of representing knowledge in which concepts are described both by their meaning and their relationship to each other" [7] or in terms of "...an explicit and formal specification of a conceptualization" [3]. An ontology thereby assists in representing a formalised current state or understanding of a field or a knowledge domain [7] [13]. In other words, an ontology is a knowledge representing language with a well-defined semantics that renders it possible not just to express facts about a knowledge domain but to make the domain and its' current facts accessible to computerised large scale analysis [13].

The transmission framework, as presented here, was not completed at the beginning of the project. Starting with an initial framework, which consisted of expandable hierarchies or lists regarding the first four items in the following enumeration, the framework has been continuously extended and adjusted motivated by the information encountered during the literature search. Hosts and viruses that were not present in the initially provided taxonomic hierarchy have been added mainly according to information derived from the NCBI Taxonomy Browser [131]. For bacteria we relied on the regularly updated 'List of Procaryotic names with Standing in Nomenclature (LPSN)' [59] but also took the NCBI Taxonomy Browser into account. In order to minimise and control for redundancy, the hosts and pathogens along with the taxonomic hierarchies were kept on a collectively accessible, relational database system (see [129]). However, the transmission framework was finally composed of the following entities:

1. A cladistic taxonomy for the pathogens;
2. A taxonomy for the habitat items which contains a cladistic taxonomy for the hosts[1] (see figure 2.1);
3. A finite set of entry portals for the hosts;
4. A finite set of exit materials for the hosts;
5. The structuring rule '*is instance of*' which structures both pathogen and habitat taxonomies in a way that assigns every item to exactly one item higher up in the hierarchy;
6. The structuring rule '*transmits pathogen X to* (where applicable: *via entry portal and/or exit material*)' or shortened '*transmits to*' which links any two habitat items to form a transmission route (see red arrows in figure 2.1.

[1] A concept in this context is an idea or a notion of a real world object or occurrence.

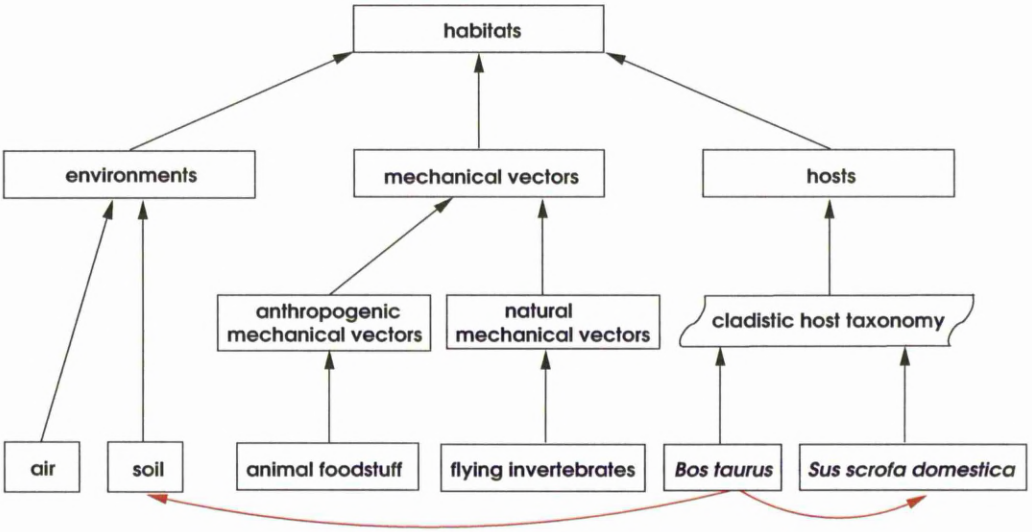

Figure 2.1: The habitat hierarchy. The items of the habitat hierarchy are depicted as rectangles whereas the **black arrows** denote the structuring rule 'is instance of' in a bottom-up manner, i.e. an item the black arrow is adjacent from 'is instance of' the one the arrow is adjacent to. Note that the host items are part of a much more complex hierarchical structure which is indicated by the field 'cladistic host taxonomy'. Also note that the term 'habitat hierarchy' has been used elsewhere in the literature already [10]. The red arrows exemplify the structuring rule 'transmits to'. For example, *Bos taurus* 'transmits pathogen X to' soil and also to *Sus scrofa domestica*.

In this context the term *habitat* refers to the potential location or whereabouts of pathogens. As shown in figure 2.1 the current framework distinguishes between three general habitats, namely environments, mechanical vectors and hosts. These three habitats might be of very different nature. For example, a host habitat by definition implies a replication and/or multiplication event, whereas mechanical vectors by definition exclude any form of replication and multiplication event. Environments on the other hand might involve some degree of replication and/or multiplication. The habitat items can then be subdivided further which is indicated in figure 2.1 by the black arrows representing the structuring rule 'is instance of'. Consult subsection 2.1.1.1 for a complete list of habitat items and definitions.

The structuring rule 'transmits to' was extended by adding an exit material and entry portal where appropriate. A pathogen is released from an infected host via an exit material and gets access to a susceptible host by means of an entry portal. Subsection 2.1.1.3 contains a complete list of exit materials and entry portals. This rule requires any two habitat items to be linked which consequently yields nine general configurations. These will be referred to as *transmission configurations*:

1. host-to-host
2. host-to-environment

3. environment-to-host
4. host-to-'mechanical vector'
5. 'mechanical vector'-to-host
6. environment-to-'mechanical vector'
7. 'mechanical vector'-to-environment
8. environment-to-environment
9. 'mechanical vector'-to-'mechanical vector'

Each transmission configuration can be assigned to one of four *transmission types* based on the information required on the exit materials and entry portals (see figure2.2):

1. **Transmission type A**: Host item X 'transmits to' host item Y via an exit item and entry item, also called an *entry-exit pair*. In other words, pathogen Z is released by the donor host X through the exit material i and is then relatively immediately taken up by the receiver host Y via an entry portal j which leads to infection of host Y. This includes the transmission configuration host-to-host.
2. **Transmission type B**: Host item X 'transmits to' environment item *or* mechanical vector item Y via an exit item. In other words, pathogen Z is released by the donor host X into or onto the environment (mechanical vector) Y by means of the exit material i. This leads to contamination of the environment (mechanical vector) Y. Type B includes the transmission configurations host-to-environment and host-to-'mechanical vector'.
3. **Transmission type C**: Environment or mechanical vector item Y 'transmits to' host item X via an entry item. In other words, pathogen Z is taken up "from" the donor environment (mechanical vector) Y "by" the receiver host X via the entry portal j. This leads to infection of host X. Type C includes the transmission configurations environment-to-host and 'mechanical vector'-to-host.
4. **Transmission type D**: Environment or mechanical vector item X 'transmits to' environment or mechanical vector item Y. In other words, pathogen Z is transferred "from" the donor environment (mechanical vector) X "to" the receiver environment (mechanical vector) Y. This leads to contamination of environment (mechanical vector) Y. Type D includes the transmission configurations environment-to-'mechanical vector', 'mechanical vector'-to-environment, environment-to-environment and 'mechanical vector'-to-'mechanical vector'.

It is important to note that a particular transmission event, i.e. transmission from one particular donor to one particular receiver host, can consist of or rather can be represented by one or more transmission routes.

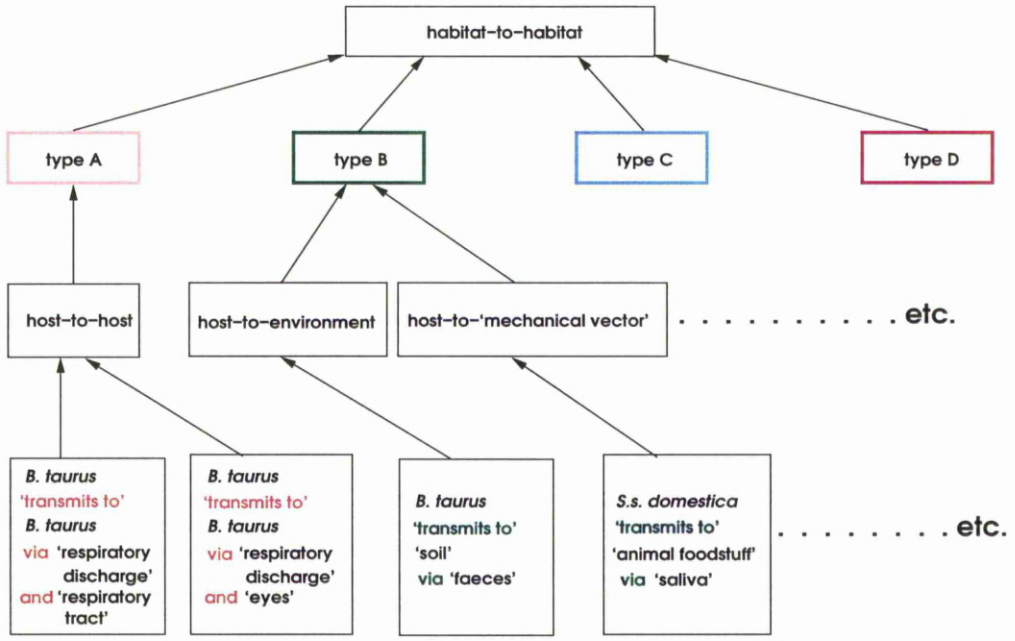

Figure 2.2: Transmission routes, configurations and types. The bottom level depicts four examples of transmission routes. Actual pathogens were omitted in this figure. Each transmission route 'is instance of' exactly one transmission configuration as indicated by the arrows. Each transmission configuration is then 'instance of' exactly one transmission type. All routes, configurations can be viewed as 'habitat-to-habitat' transmissions.

2.1.1 Items of the transmission framework

> *"Language is a process of free creation; its laws and principles are fixed, but the manner in which the principles of generation are used is free and indefinitely varied. Even the interpretation and use of words involves a process of free creation ..."*
>
> Noam Chomsky

> *"A word can only serve to indicate that someone else may have a valuable idea – that is, some useful structure to be built inside the mind. Each new word only plants a seed: to make it grow, a listener's mind must find a way to build inside itself some structure that appears to work like the one in the mind from which it was "learned" ..."*
>
> Marvin Minsky [120]

In addition to the main items of the transmission framework, two additional items were incorporated to account for either gaps in the transmission framework or missing information in the literature:

1. **not in list**: This substitutional item can be used in case the literature does supply information on 'environments', 'mechanical vectors', 'hosts', 'exit materials' or 'entry portals' for a particular transmission route but our list of items does not seem to provide an appropriate match to assign that information to.
2. **unknown**: This substitutional item can be used in case the literature does not supply the information on any particular item which is required to set up a particular transmission route.

2.1.1.1 Habitats: Environments, mechanical vectors and hosts

A 'habitat' item has been defined as any potential location or whereabouts of a pathogen which can be involved in a transmission event. Note that even pathogens can form habitats for other pathogens. For example, certain viruses such as bacteriophages can inhabit bacteria. However, these scenarios have not been taken into account.

Environments:

In this context environments could probably be most suitably described as habitats which form the surroundings of hosts as well as mechanical vectors. Certain environments might also provide a growing medium for particular pathogens. The corresponding environment items are listed subsequently in alphabetical order:

1. **air**: Involved in "airborne"transmission over larger distances (see also subsection 2.2.3.1).
2. **bedding**: Materials such as straw, hay, matting etc. which are used as a ground layer in animal stables.
3. **lakes and ponds**: Freshwater habitats with a comparatively low water flow rate, i.e. standing freshwater habitats.
4. **man made**: Anthropogenic constructions like buildings (not the floor area), canals, aqueducts, drinking water systems, distribution mains, stables, fencing posts etc. which cannot be assigned to any one of the other items. It also comprises sewers, dung piles, slurry reservoirs etc., i.e. such environments which might contain artificially high numbers of pathogens caused by human intervention.
5. **natural**: Any additional objects of the natural environment that do not seem to fit in one of the other environment items, e.g. trees, bushes, twigs, hedgerows etc.
6. **oceans**: Marine salt water habitats.
7. **rivers and streams**: Freshwater habitats with a comparatively high water flow rate, i.e. flowing freshwater habitats.
8. **soil**: The ground floor including pastures, meadows and indoor floors.

Mechanical vectors:

A 'mechanical vector' was defined as any living or inanimate structure involved in the dissemination and spread of a pathogen from one habitat to another. The important point is that the pathogen is not meant to cause infection of the 'mechanical vector', i.e. no appreciable replication or multiplication events on or in the 'mechanical vector' are involved. Mechanical vectors have been further subdivided into 'anthropogenic' and 'natural' mechanical vectors.

1. **animal equipment** (anthropogenic): Any device, tool, gadget or contraption that disqualifies as a vehicle or as a medical device but is involved in the husbandry of non-human animals. This includes, for example, harnesses, brushes, ringing, hoof trimmers/cutters, milking machineries etc.
2. **animal foodstuff** (anthropogenic): Food or foodstuff used for consumption of non-human animals. This includes, for example, silage, hay, feed meals etc.
3. **animal products** (anthropogenic): Anything that is produced by or produced from non-human animals. For example, provisions like eggs, milk and meat but also products for the clothing industry such as leather would fall into this group.
4. **birds** (natural): All members of the taxonomic class *Aves*. The main reason to incorporate this item was the ability of most birds to fly and potentially to cross large distances relatively quickly.
5. **flying invertebrates** (natural)
6. **large terrestrial vertebrates** (natural): Comprises non-human animals such as, for example, dogs, cats, cattle, sheep or pigs.
7. **local invertebrates** (natural)
8. **medical equipment** (anthropogenic): For example syringes, catheters, clinical examination gloves and scalpels.
9. **people** (anthropogenic): The human being as a mechanical vector and not as a host. This could involve the clothing, footwear, hands, hair etc.
10. **small terrestrial vertebrates** (natural): Comprises non-human animals such as, for example, mice and rats.
11. **vehicles** (anthropogenic): This includes the exterior and/or interior of, for example, cars, lorries, planes, helicopters, ships and boats.

Hosts:

A host according to our definition is *any* living animal which can be infected by a pathogen. Infection in this context involves a replication and/or multiplication event of the pathogen inside or on the exterior of a host [66]. A developmental stage of a pathogen taking place inside or on the exterior of a host can also be viewed as process of infection. However, we did not view the mere uptake of a microbial toxin as an infection as has been suggested elsewhere in the literature [101].

The term 'colonisation' is often used synonymously with 'infection' [101] [102] [66] and applies even to the normal (micro-)flora of hosts, i.e. to commensal microorganisms. It was not our intention - and it would have been clearly far beyond the scope of this work - to account for all microorganisms that might belong to the normal (micro-)flora of the main seven livestock host species.

2.1.1.2 Microbial threats (pathogens)

This includes all microorganisms such as bacteria, archaea, viruses, helminths, protozoa, fungi etc. as well as their corresponding toxins, developmental or survival stages if they are able to cause an infection of a host. Toxins, developmental or survival stages were viewed as integral parts of the respective microorganism and do not constitute a separate threat. Note also that the uptake of a microbial toxin was not viewed as a process of infection as opposed to elsewhere in the literature [101].

Commensal microorganisms were included only if they are also known to be opportunistic pathogens.

2.1.1.3 'Transmission type'-specific items: Exit materials and entry portals

Depending on the transmission type a corresponding transmission route might involve an exit material and/or an entry portal.

Exit materials:

Exit materials are organic discharges by which a pathogen is released from and/or passed on by a donor (sender) host to another habitat.

1. **blood**: This includes dried blood but not scabs. Scabs have been assigned to the item 'shed skin'.
2. **eggs**: Includes the interior as well as the exterior of eggs and can be involved in vertical but also horizontal transmissions. Roundly or ovally shaped bodies laid by a female, egg-laying animal. An egg consists of an ovum, membrane layers and an outer shell.
3. **faeces**: Note that birds (class *Aves*) discharge urine and faeces through the cloaca which makes cross-contamination from bird urine to bird faeces or vice versa very likely. Therefore, it has been decided to assign both bird urine and bird faeces to the item faeces.
4. **flesh**: This item includes all of the remains (carcass, carrion, cadaver, corpse, body) of a dead animal (including humans). Bones are part of this item. It may also refer to living creatures. For example, wound picking birds might ingest bits of flesh when picking on wounds.
5. **milk**
6. **ocular secretion**

7. **placenta/aborted material**: Can be involved in vertical and horizontal transmissions.
8. **pus**
9. **respiratory discharge**: All excretions and secretions of the upper (oronasal) respiratory tract as well as the lower (pulmonary, bronchial, alveolar) respiratory tract that can be exhaled, i.e. breathed out, sneezed out, coughed out etc.
10. **saliva**: Note the overlap with 'respiratory discharge'.
11. **semen**
12. **shed skin**: This item includes dandruff, scabs, external mucosas (e.g. penile mucosa) etc.
13. **urine**: Note that the urine of birds (class *Aves*) has been assigned to the item 'faeces'.
14. **vaginal secretion**

Entry portals:

Entry portals are the ways by which a pathogen initially enters a susceptible host. They do not necessarily relate to the final location(s) or tissue(s) where actual infection takes place.

1. **alimentary tract**: This item has been called initially 'ingestion'. It covers the whole upper parts of the digestive tract but not the rectum and anus. It should be used for what is usually referred to as ingestion. Note that the oral cavity is part of both the upper alimentary as well as upper respiratory tract.
2. **blood stream**: Direct injection or entry into the vascular system, for example, via a mechanical vector such as 'medical equipment' (syringes etc.), via a host such as biting arthropods etc. This item also includes the haemolymph of arthropods.
3. **damaged skin**: Abrasions and deeper wounds of the skin.
4. **eyes**
5. ***in utero*/congenital**: This item covers vertical transmission events from the mother to the unborn offspring exclusively. The pathogen can be acquired at any point after the fertilisation of the ovum has taken place up until the offspring is finally released from the mother's body. It includes the process of parturition, e.g. ingestion of the pathogen by the offspring during parturition.
6. **intact skin**: Only to be used if the pathogen itself exhibits an active strategy to cross the barrier formed by intact skin.
7. **mammary tract**: The teat opening or teat canal.
8. **rectum**: Entry of a pathogen via the anus and rectum of the lower alimentary tract (e.g. during homosexual behaviour of rams).

9. **respiratory tract**: This item has been termed initially 'inhalation'. It has been retermed because of the ambiguities laid out in subsection 2.2.3.1. The respiratory tract involves the upper (oronasal) respiratory tract as well as the lower (pulmonary, bronchial, alveolar) respiratory tract. The new term is also more consistent with the other entry portals. It refers to a location rather than an act of entry. Inhalation of smaller or larger particles towards the upper or lower respiratory tract will be involved in the infective process. Note that the oral cavity is part of both the upper alimentary as well as upper respiratory tract.

10. **urogenital tract**:

2.2 Literature search

2.2.1 Search strategy and search procedure

The literature on the transmission of pathogens can be expected to be extensive. Therefore, the basic search strategy was guided by recent textbooks on infectious diseases which served as a kind of master candidate list to identify important pathogens, relevant host species and fundamental data and information on transmission-related characteristics. It seems justified to assume that the majority of pathogens and additional information which are covered by these books, both to be of significant importance to the field and to be studied in greater detail than those which have not been mentioned in major textbooks so far. These books included references [152] [30] [88] [159] [179] but also [181] [86] [110]. An ancestor search of the books' reference lists was applied as a further search procedure in cases where the information stated in the books was too general, incomplete, ambiguous, insufficient or even contradictory. If necessary further data and information was sought either via additional, more specific textbooks or via electronic sources such as two bibliographic databases, i.e. NCBI's PubMed and ISI's Web of Science, the internet using the search tool Google and the electronic library of the University of Liverpool. The online searches included - apart from the name of the pathogen, host or associated disease - any combination of the key words "transmission", "infection", "epidemiology", "entry portal" (or entry route), "exit material" (or excretion, secretion). In a few cases non-peer-reviewed sources were included and/or expert opinions were consulted. The basic search strategy was not restricted to any particular type of publication, journal, study design, date of publication or language to maximise available data. However, up-to-date information was generally favoured if available. References were added to a collectively accessible online reference management system, namely RefWorks, and later uploaded to the database system [129].

Moreover, the search was carried out by three reviewers who independently used the transmission framework to review the literature. Two reviewers focused on viral and one on bacterial pathogens. Issues arising during the search process were sought to be resolved in meetings between the reviewers. In order to facilitate the creation of transmission routes, the information was initially collected in Excel spreadsheet tables (see table 2.1). In total, data and information from at least 17 additional textbooks, 188 peer-reviewed articles on bacterial pathogens and 37 peer-reviewed articles on viral pathogens has been collected for the general dataset.

		host species 1	host species 2	host species 3	etc.
information on ...	...entry portal	ingestion (ref.1); etc.			
	...exit material				
	...environments				
	...mechanical vectors				
	...survival outside a host				
	...miscellaneous aspects				

Table 2.1: Spreadsheet used to collect data and information: Relevant data and information encountered during the search was stored along with a unique identifier for its reference in the appropriate cell within a pathogen-specific spreadsheet, i.e. one spreadsheet table for each pathogen.For example, host 1 gets infected by the pathogen in focus via ingestion, i.e. via item alimentary tract.

The information in the tables was subsequently consulted to decide on the creation of transmission routes. Each transmission route was temporarily stored in additional spreadsheet tables before being uploaded to the collective database system. Every transmission route is finally described by:

- Unique ID for the transmission route;
- Transmission category;
- Pathogen name;
- Taxonomic path of pathogen;
- Sender host species name were applicable;
- Taxonomic path of sender host were applicable;
- Exit material were applicable;
- Receiver host species name were applicable;
- Taxonomic path of receiver host were applicable;
- Entry portal were applicable;
- Mechanical vector item were applicable;
- Environment item were applicable;
- Supporting data and information along with reference ID's.

2.2.2 Search constraints

The number of potential pathogens and associated host species is too large to be reviewed by this approach comprehensively. The search for information was therefore restricted exclusively to those pathogens which infect at least one of seven predefined livestock host species:

1. *Bos taurus*, i.e. domestic cattle, cows;
2. *Sus scrofa domestica*, i.e. domestic pigs, swine;
3. *Ovis aries*, i.e. domestic sheep;
4. *Gallus gallus domesticus*, i.e. domestic chicken;
5. *Meleagris gallopavo*, i.e. domestic turkeys;
6. *Anser anser*, i.e. domestic geese;
7. *Anas domesticus*, i.e. domestic ducks.

Such a restriction also seems reasonable from an economic as well as public health point of view. The second and final restriction addressed the transmission routes. Figure 2.3 gives an overview of the included transmission routes and those ignored for this study. Certain transmission routes were omitted owing to a lack of information (in particular with respect to wildlife hosts [143] [115] [175]) and a lack of time.

It should be noted that the search for information was not necessarily carried out on the species level of pathogens. That is, where the literature justified the inclusion of a subspecies or strain level, the corresponding variants have been treated separately. This decision is supported by Woolhouse *et al.* who pointed out that "...the species [...] unit [...] ignores a wealth of important and interesting variation within species in traits such as [...] host specificity" [204].

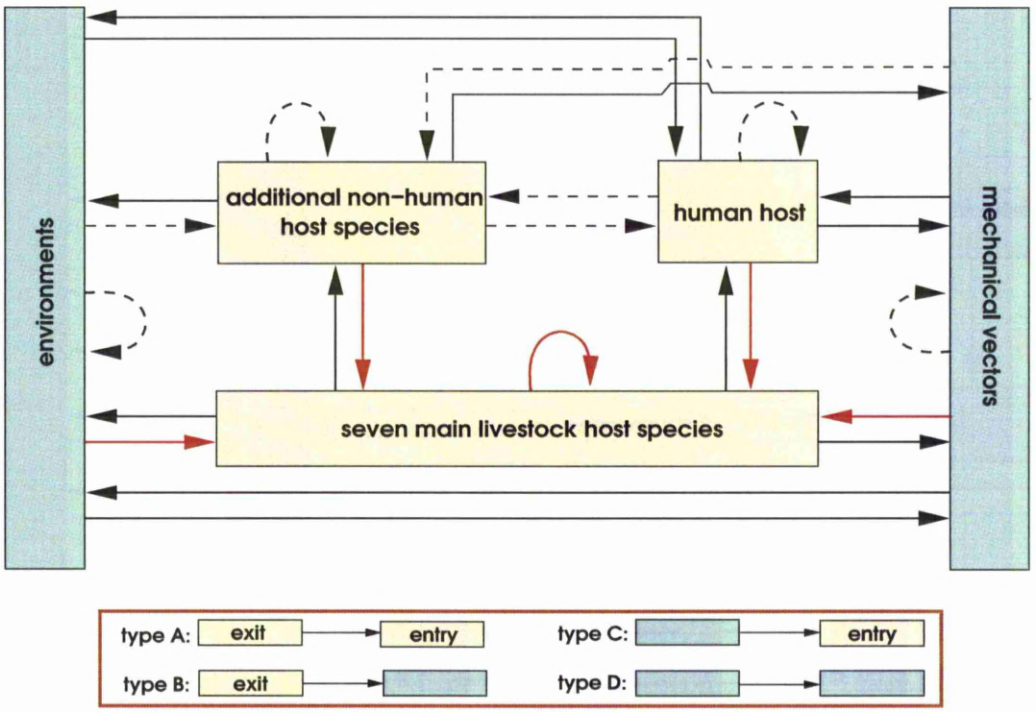

Figure 2.3: Overview of the included transmission routes. The figure gives an overview of the transmission routes of interest. In fact, in order to simplify matters it rather depicts the corresponding transmission configurations. Note that the host habitats have been split into three groups (see yellow boxes) in order to highlight the different treatment of these host groups with respect to the inclusion and exclusion of transmission routes. The host habitats are flanked by the habitats 'environments' and 'mechanical vectors' which are coloured green. Arrows denote the structuring rule 'transmits to'. **Solid arrows** symbolise included transmission routes whereas those shown as **dashed arrows** were omitted. Red arrows point to an immediate threat to the main livestock host species. The brown-framed box at the bottom of the figure highlights the additional information needed on exit materials and entry portals.

Quality assessment and exclusion criteria:

Transmission routes that were regarded as unclear or uncertain were marked and later discussed in group meetings and then either rejected or included. The final database consisted of 7808 distinct bacterial and viral transmission routes. Transmission routes were excluded from analysis if they did not show the complete set of data required for a specific transmission route. For example, the use of the items 'unknown' or 'not in list' as well as an incomplete taxonomic host or pathogen path led to exclusion of the corresponding transmission route. However, representatives of higher taxonomic host groups were allowed in case the actual host species could not be identified. More

precisely, if the literature stated the involvement of rats without specifying the actual rat species, this was accounted for by incorporating a substitutional label such as "Rattus spp.". After exclusion the final dataset consisted of 7204 distinct transmission routes, 215 distinct host species and 202 distinct pathogens.

2.2.3 The transmission framework in use

> *"... most concepts are partially understood in terms of other concepts [...] Are there any concepts at all that are understood directly, without metaphor? If not, how can we understand anything at all?"*
>
> *"In allowing us to focus on one aspect of a concept [...], a metaphorical concept can keep us from focusing on other aspects of the concept that are inconsistent with that metaphor..."*
>
> Lakoff and Johnson [100]

This section considers two concrete examples in order to investigate nature and degree of difficulties that can be expected when dealing with literature information. Such information might be contradictory and/or it might be presented by means of terms and concepts that cannot be applied directly to the transmission framework. A thorough understanding of newly encountered terms and concepts is therefore a basic requirement. Only then can we make educated decisions on how we are to reconcile their informational content with the transmission framework. The potentially problematic nature of epidemiological data has already been addressed elsewhere in the literature. For example, it has been stressed that the "... free use of concepts and terms in a rather loose fashion contribute to inaccuracy of meaning and confusion of issues" [4]. The same author also pointed out that "... fundamental contradictions, not only" [4] exist between "... the opinions of different authors, but also within a single article" [4]. Moreover, the author views the interpretation of existing data as one of the main sources of discrepancies [4]. Nevertheless, interpretation might be a necessary evil since "nothing is intrinsically meaningful we must..." [42] build on interpretations in order to create meaning [42]. But "what people call "meanings" do not usually correspond to particular and definite structures, [...] meanings are rarely sharp, and we cannot always expect to be able to "define" them in terms of compact sequences of words. Verbal explanations serve only as partial hints ..." [120].

2.2.3.1 Example 1: Airborne transmission - a "simple" concept ?

Tuberculosis is a disease associated with the bacterium *Mycobacterium bovis* well-known airborne pathogen. In the following paragraphs I will run through the intuitively simple [134] concept of airborne transmission. Here are a few facts taken from the literature:

1. The primary host species is cattle, i.e. *Bos taurus* [67]. Cattle is said to be the main source of infection for other cattle [152], i.e. cattle-to-cattle transmission has been recognised as an important factor [67].

2. The bacterium is concentrated in the respiratory system [67] and is excreted in exhaled air [152].

3. Exhaled air, i.e. respiratory exhalations, are produced by means of sneezing [123] [167] [176] [212], coughing [123] [167] [176] [212], breathing [176] [212], talking [31] [123] [134] [167] or singing [31]. These processes create so-called *droplets* [167] [212] "and" *droplet nuclei* [31] [212].

4. Inhalation is considered as the principal mode of transmission in this context [152] and challenge of the respiratory tract with a very low dose of this bacterium initiated infection [67].

5. The disease caused by this bacterium can be airborne and US control strategies focus on *airborne transmission* [81]. The "...hypothesis that airborne transmission is the principal route of infection" [67] for cattle with this pathogen has been supported by current findings [67].

6. "Airborne particles containing micro-organisms can either originate from liquids as *droplets or* from dry matter. Droplets present a large surface to the air and evaporate quickly. They are thus reduced in size and weight and can *remain airborne* over long time periods. The residues of these evaporated droplets are called droplet nuclei..." [176] which "...are small enough to be inhaled" [176]. In other words are droplets airborne and can 'remain airborne', although when reduced in size are called droplet nuclei. Droplet nuclei are "...infectious particles of respiratory secretions that are aerosolized by coughing, sneezing" [31] etc.

Item two of the preceding list justifies 'respiratory discharge' as an exit material whereas item four provides sufficient evidence for the entry portal 'respiratory tract'. Given the information so far, it intuitively suggests the creation of a direct host-to-host transmission route:

- *Bos taurus* 'transmits to' *Bos taurus* via exit material 'respiratory discharge' and entry portal 'respiratory tract'.

This appears to be a sound decision at first glance. But is it really that simple? Did we look closely enough at the provided information? Can we think of any alternative scenario?

In fact, the list of information does also suggest the involvement of air as an intermediate, though perhaps rather transient, part of the transmission event. Hence, should it rather be entered as two separate transmission routes, namely 'cattle-to-air' followed by 'air-to-cattle'? This view is supported by the notion that airborne transmission is an *indirect* way of spreading a microbial threat from a source or reservoir to an animal or person [101]. 'Airborne' has been defined by Last as the "...dissemination of microbial aerosols to a portal of entry" [101]. However, Last [101] explicitly excludes the spread of droplets in his airborne definition. Instead, this is viewed as a *direct* way of transmission where droplets are projected "...onto the conjunctiva or onto the mucous membranes of the eyes, nose, or mouth" [101]. This is in accordance with the idea that droplet transmission is "...a form of contact transmission" [167] in which the microbial threat travels "directly from the respiratory tract of the infectious individual to [...] mucosal surfaces of the recipient..." [167]. Furthermore it is said to happen "generally over short distances..." [167]. Such transmission events have therefore also been termed 'droplet transmission' [134], 'droplet-borne transmission' [212] or 'droplet spread' transmission [101] as opposed to 'airborne transmission'. However, other authors such as Nakamura *et al.* [128], Madigan *et al.* [105] and Morawska [123] consider even 'droplets' as airborne.

A seemingly crucial characteristic in the foregoing paragraph is the distinction between direct and indirect transmission. What exactly is the difference between these two concepts within the context of droplet *vs.* airborne transmission? According to Last does the term 'direct' denote the "...essentially immediate transfer of infectious agents to a receptive portal of entry" [101]. This definition seems to be close to the notion that 'direct' transmission involves a close spatiotemporal contact of both the donor and the recipient individual in order to allow for transmission without intermediate objects or individuals [123] [167]. 'Indirect' transmission on the other hand should be significantly separated in time and/or space which is most likely to involve intermediate objects and/or individuals [101] [123] [167]. A first clue of the potential range of spatial distances which are involved in a 'direct' transmission event can be taken from the findings of Xie *et al.* These authors state that 'droplets' sized between [60, 100] μm can be directly expelled more than 6 m away by exhaled air at a velocity of 50 m/s (sneezing), more than 2 m away at a velocity of 10 m/s (coughing) and less than 1 m away at a velocity of 1 m/s (breathing)[2] [212]. This does obviously not exclude airborne transmission from 'short distances' [139] but it suggests potential limits for so-called 'droplet transmission'. The infectious distance of airborne transmission will be discussed further down in the text. Other authors neither assign the direct nor the indirect label to airborne transmission but rather list it as a separate item within the context of transmission ways [123] [167]. This suggests the potential ability of an airborne pathogen to be involved in both direct and indirect transmission events.

In hope of reaching a more satisfying conclusion on the issue of droplet *vs.* airborne transmission, I am now going to investigate the terms 'droplet' and 'droplet nucleus'. Last [101] makes a clear distinction between these terms. This author restricts airborne transmission to 'microbial aerosols' which are suspensions of particles in the air [101]. His definition of a particle includes 'droplet nuclei' and contaminated dust but explicitly does not include 'droplets'. According to Last [101] do hosts emit only 'droplets' which subsequently reduce their size via evaporation of fluid or any other atomizing process. The remaining residues are then called 'droplet nuclei'. This is in contrast to the earlier statement that both 'droplets' and 'droplet nuclei' are formed directly during the processes of exhalation [31] [212].

[2] All these figures represent most likely human respiratory activities [212].

particle description	⌀(μm)	reference	year of publication
droplet, smaller droplet, droplet residue	[1,2]	[123]	2006 (1945)
droplet nucleus	[1,3]	[155]	1974
droplet nucleus	<5	[134]	2005 (1945, 1994)
droplet nucleus	≤5	[167]	2007
droplet nucleus	<6	[31]	1998
(small, minute) droplet	<[100,200]	[199]	1934
droplet	[1,100]	[123]	2006
droplet	>5	[167]	2007
large droplet	>5	[212]	2007 (1996, 2005)
larger droplet	>[5,10]	[123]	2006
large droplet	>10	[212]	2007 (2004)
larger droplet	>[10,20]	[123]	2006
droplet	>25[5]	[134]	2005
large droplet	>100[6]	[212]	2007 (1934)
large drop, droplet	>[100,200]	[199]	1934

Table 2.2: Particle description and aerodynamic diameter ⌀: The table lists particle diameters ⌀(principally increasing from top to bottom) along with the applied particle description as encountered in the stated references. If a reference based their statements on other references, these secondary references are shown in parentheses.

The terms 'droplet' and 'droplet nucleus' are in fact used quite loosely throughout the literature which can obviously cause confusion and lead to misinterpretation. What exactly is the discriminating factor between a droplet and a droplet nucleus? Some authors state the size, i.e. the aerodynamic diameter, of a particle as the discriminating criterion [134]. Table 2.2 summarises a couple of circulating ideas on particle diameters along with the applied terminology. It quickly

[3]Nicas *et al.* actually state a droplet diameter of "...a few tens of μm or larger" [134] which I interpreted as >25 μm.

[4]Referring to [212] as stated in the next row.

[5]Nicas *et al.* actually state a droplet diameter of "...a few tens of μm or larger" [134] which I interpreted as >25 μm.

[6]Referring to [212] as stated in the next row.

becomes apparent through this table that the literature does not yet agree on a consistent system of well-defined terms and size thresholds. In fact, even a single publication can show some ambiguity in the use of terms.

$\varnothing_r$ (μm)	reference
<2	[176]
<3.3	[148]
$\leq$3.5	[119]
$\leq$4	[119]
$\leq$5	[119]
$\leq$10	[134]
$\leq$12	[134]

Table 2.3: Aerodynamic diameter of respirable particles $\varnothing_r$: The table lists a few proposed diameters (increasing from top to bottom) of particles which are considered to be sufficiently small for inhalation by (most likely) human beings.

Despite these ambiguities, there seems to be a shared view on a potential size limit for respirable particles with reference to the lower respiratory tractentry portallrespiratory tract. Table 2.3 lists a few suggestions on this matter. Given the information of this table one could conclude that in terms of inhalation towards the lower respiratory tract as a way of entry, the involved particles should roughly measure $\leq$12 μm in diameter. This figure might be different for host species other than humans. The literature furthermore states that particles >10 μm [134] or $\geq$6 μm [176] "generally" do not reach the alveolar region but might be trapped in the upper respiratory tract (here most likely referring to humans). Particles in the [2,10] μm "...range reach the alveolar region with variable efficiency" [134].

How can one be sure that inhalation is the exclusive way of entry associated with airborne transmission in the first place? As a matter of fact, many authors do regard inhalation as the sole way of entry for airborne transmission. Usually this involves transport to and infection of the lower respiratory tract, i.e. the alveolar, pulmonary, unciliated region of the deep lung [119] [134] [167] [176]. If the entry portal is instead the upper respiratory tract, like mucosal surfaces of mouth and nose or any other surfaces of the body, then it is often referred to as 'droplet transmission' [134] [167] [176] or 'droplet-borne transmission' [212]. Being already familiar with the problematic nature of the term droplet, this terminology appears to be rather debatable. Moreover, Nicas *et al.* acknowledge that "...it is invalid to broadly assert that respirable particles are required to transmit infection via inhalation, because for some pathogens the upper respiratory tract may be a target tissue" [134]. Last [101] actually makes no clear restrictions in terms of the involved entry portals at all but rather points to the respiratory tract as a whole as the usual, though not necessarily exclusive, portal of entry for airborne transmission. This view is supported and further specified by Wathes (in [33]) who considers airborne transmission not confined to respiratory disease and inhalation. Given his point of view, it may well involve other portals of entry such as ingestion, the conjunctiva or wounds. This idea seems to be shared by additional authors such as Oliveira *et al.* [139].

Following the entry portals I will now consider the exit materials and the sources for airborne microbial threats. Nicas *et al.* apparently focus on aerosols derived from exhaled air only [134].

Many other authors, however, also include various other materials resuspended in the air [31]. For instance, faeces [33] [123] [139] [176], urine splashes [176], vomit [123], skin flakes [123], contaminated dust particles from dry matter [128] [176] [193] such as 'soil' [101] [105] and 'bedding' [33] [101] [167] and even mechanical vectors such as 'animal foodstuff' [33] [101] [167] have been proposed in the literature. Madigan *et al.* [105] simply group all microbial threats that can be found in the air under the term 'airborne', independent of both the initial source and the involvement of intermediate resuspension processes.

Another important issue is the infectious distance of airborne transmission. Airflows act on a global scale[7] and gale-force wind speeds can reach up to 194 km/h. This doubtless bears the potential of disseminating microbial threats both quickly and globally. Air is luckily not considered to be a growth medium for microorganisms [105]. In fact, it provides a rather harsh environment for their survival [105] [187] and the volume of available air will also act on the infectious concentration. This is particularly true for outdoor environments which are also more exposed to the harmful ultraviolet spectrum of the sunlight [20]. Nevertheless, some airborne microbial threats such as the foot-and-mouth disease virus are known to travel long distances within the air and still remain infectious [33] [51] [119] [173] [176]. Infectious distances of ca. 10 km [33], >20 km [119], >100 km [176] and 250 km [92] have been reported for this virus. Evidence furthermore suggests that microbial threats, including bacteria and fungi, could possibly be transmitted over much larger - even intercontinental - distances like the transatlantic gap between Nothern Africa and the Americas [74] [148]. The concentration of microbial threats in the air might therefore be considerably higher in a contaminated indoor environment [20] [105]. However, some host-pathogens interactions have extremely low infectious doses [176]. A perfect example in this context is *Mycobacterium bovis* field strain AF2122/97 which is able to cause infection in cattle with an efficiency of 50% via just a single colony forming unit (cfu) [84]. Low concentrations of viable cells in the outdoor air might therefore still pose a serious risk of infection. This is even more worrisome when considering that certain microorganisms form survival stages such as spores etc. [193].

Flügge [62] [176] considered airborne transmission applicable to any microbial threat but rather viewed the actual incidence of airborne transmission as a function of probability. In other words it is more likely for certain microbial threats than it is for others. The *Centers for Disease Control and Prevention* (CDC) [167] [212], however, regards only the pathogens *Mycobacterium tuberculosis* (tuberculosis), measles virus and varicella-zoster virus (chickenpox) as truly airborne. Such variations might simply be due to the aforementioned discrepancies regarding the definition of airborne transmission.

The foregoing discussion on airborne transmission as well as associated terms and concepts casts a dark shadow on their current usefulness with respect to our approach. Given the findings, we have to consider *Mycobacterium bovis* to be capable of long distance transmission. However, the potential flexibility still might open the gates for individual interpretation which could result in significantly different sets of transmission routes. For example:

- Is airborne transmission a direct and/or an indirect way of transmission?
- If airborne transmission does involves indirect ways, how exactly does it take place?

[7] For example, jet streams which are fast flowing (ca. 92-398 km/h), narrow (ca.7-17km above sea level) air currents in the earth's atmosphere.

- Which exit materials and entry portals will be involved in airborne transmission?

To complete this discussion I should at least mention a few more terms like, for example, 'soilborne', 'dustborne', 'waterborne', 'foodborne' etc. Take the term 'dustborne', for example. Dust is an integral component of both soil and air. Does it therefore represent a combination or specification of the terms 'soilborne' and 'airborne'? It is not our intention to provide an answer on that issue. However, I will have a brief look at the term 'soilborne'. Soilborne diseases, according to Santamaría *et al.*, are restricted to enteric pathogens which contaminate soil only by means of faeces and which subsequently infect a susceptible host exclusively via the oral uptake of this very soil [161]. Santamaría *et al.* suggest the usage of distinct terms in order to refer to allegedly different situations which also involve soil. These terms are 'soil-associated', 'soil-based' and 'soil-related' [161] and the stated definitions appear somewhat confusing. Let us focus on the first two terms. 'Soil-associated' refers to diseases caused by "...opportunistic or emerging pathogens that belong to the normal soil microbiota" [161] whereas 'soil-based' diseases are "...caused by pathogens indigenous to soil" [161]. In other words, both of these categories seem to address pathogens that are considered to inhabit primarily soil. Only the former category assigns an additional specifications to the pathogens. Therefore the set of pathogens belonging to the former category is in fact a subset of those belonging to the latter. The specification restricts the pathogens to those which are considered to be opportunistic or emerging. The term 'opportunistic', which seems to be synonymous with the term 'facultative' in this context, would refer to pathogens which usually inhabit soil but under certain circumstances cause an infection (here disease) in a host. This should actually not be any different from those pathogens which are assigned to the soil-based category. The only discriminating factor left is consequently the concept of emergence. This concept is, as I have pointed out earlier, not unambiguous nor do the authors clarify their point of view. Moreover did I mention that the term 'emerging' is unlikely to refer to a genuinely novel pathogen but rather to a reported reemergence, i.e. an increase in the incidence of a host-pathogen interaction. Again, this should not be any different with those pathogens which have been assigned to the soil-based category. The suggested distinction between these catagories is therefore highly questionable thereby supporting the need for vigilance when confronted with new terms and concepts.

2.2.3.2 Example 2: *C.botulimum* - infection and/or intoxication ?

Another instructive example is provided by *Clostridium botulinum*. One might encounter the following information in the literature:

1. "Botulism is an intoxication [...] not an infection..." [88]. Compare also [96].
2. "Botulism [...] is caused by ingestion of the lethal neurotoxin of *Clostridium botulinum*..." [15].
3. *C.botulinum* was found in the intestines and/or faeces of various birds [15] [64] [141] and human infants [200].

Recalling our definitions of the terms host and infection, we would not regard *C. botulinum* as a relevant pathogen. However, the third point on this list raises the question how apparently vegetative cells of *C. botulinum* can show up in the intestines and faeces of hosts if botulism is

merely caused by ingestion of the neurotoxin. Perhaps some vegetative cells have been ingested along with the neurotoxin and these cells survived the passage through the intestines. Alternatively the neurotoxin could have been released by these cells after their ingestion, still surviving the intestinal passage. Another explanation would be that *C. botulinum* in fact does multiply within the host. In order to come to a conclusion one needs to have a closer look into the literature:

4. "Any animal eating..." [88] from an animal which died of botulism "...also ingests spores, which germinate in the intestine" [88]
5. "*Clostridium botulinum* appears to be a normal and innocuous inhabitant of the intestinal tract of horses, cattle and poultry where it multiplies and from which it is shed in large numbers in the faeces..." [96].
6. "Toxicoinfectious botulism is a rare form of the disease [...] which results when toxin is produced *in vivo* by *C. botulinum* growing in the body itself, rather then by ingestion of toxin..." [96].
7. "Phosphorus-deficient cattle chew any bones with accompanying bits of flesh that they find on the range..." [88].
8. Pica[8] and osteophagia[9] occur in cattle and sheep as a result of phosphorus deficiency, although in sheep more usually as a result of protein deficiency [96] [152].
9. "Spores can survive in the environment for over 30 years..." [152] but even the "toxin can persist in carrion for at least a year..." [152] "..., particularly in bones or if protected from leaching" [152].

Items four to six clearly suggest the incorporation of *C. botulinum* since the bacterium in fact does multiply within its hosts. This information was by the way derived from the exact same sources which also excluded botulism as an infection. This is even more surprising when considering that botulism is listed by one of these sources as an infectious disease of livestock which seems to require infection in the first place [96]. The reason for such confusion could be that the literature obviously distinguishes the illness or disease, which is called botulism or botulinus intoxication, from the way the toxin enters the host in the first place. Infection with *C.botulinum* in fact appears to be rather common, at least in certain host species. However, it is apparently rarely accompanied by the production of the toxin. Interestingly the production of the botulinum toxin in *C. botulinum* is dependent on "infection" of the bacterium with certain bacteriophages [25] [34].

The information so far suggests that, although the ingestion of botulinum toxin alone can cause botulism, under natural conditions ingestion of contaminated material is likely to include spores and/or vegetative cells [10]. Does this also hold true for toxin-containing bones? One could imagine

[8] An appetite or craving for something normally not regarded as nutritive.

[9] Chewing of bones by herbivores which is indicative of phosphorus deficiency.

[10] This could be very different with respect to other bacteria like, for example, *Vibrio cholerae*. Here an infection of the intestinal tract with *V. cholerae* might be required under natural conditions in order to cause disease via the cholera toxin.

two scenarios leading to toxin-containing bones. First, vegetative bacteria produced the toxin within the bones during bone marrow infection. Second, vegetative bacteria released the toxin in a bone-associated tissue leading to subsequent permeation of the toxin into the bones. It seems likely that spores will either remain within the bones with respect to the the former scenario or that spores will remain on the surface of the bones with respect to the latter scenario.

C.botulinum has now been identified as a relevant pathogen and animal carcasses, including the bones, as potential sources of infection. Nevertheless, one is still left with the task of formatting this information in terms of transmission routes. This requires a decision on how to deal with the mortal remains of hosts. With respect to our transmission framework, there are basically three options to choose from. Such remains could be regarded as:

1. A host still;
2. A mechanical vector;
3. Part of the environment.

Hosts have been defined as "living" creatures which are, or potentially can be, infected by a pathogen. A dead creature would therefore p.d. disqualify as a host. This point of view might be too strict. Imagine, for example, a predator-prey situation in which the predator is the susceptible and the prey the infected host. The pathogen might enter the susceptible host during consumption of the prey after the latter has been hunted down and killed. Such an event might be described most appropriately as a direct host-to-host interaction, even though the donor host already passed away. Moreover, the prey might not be consumed completely straight away. The predator might either hide the leftovers or leave them in the open. In the first case the predator might return to the leftovers at a later time whereas in the latter case scavengers might feed on them. Either way, as time proceeds it seems increasingly unreasonable to continue to view the consumption of these leftovers as a direct host-to-host interaction. It would also be inappropriate to assign these leftovers to any mechanical vector since they can still provide a rich growth medium for certain pathogens. More precisley, either the multiplication of pathogens temporarily carries on or spore-forming processes kick in. Hence, should we consider such remains as becoming and being part of the environment, in particular soil? Though this is a legitimate question, we decided to incorporate the consumption of dead body parts in such contexts as a direct host-to-host interaction. Bones, for example, have been assigned to the exit material 'flesh'.

2.3 Analysis

Calculations of the entire thesis were performed and plots were produced with Matlab version 7.6.0 R2008a, the corresponding Mathworks' "Statistics Toolbox" and functions from the Mathworks "File Exchange Server". Confidence intervals for the binomial proportions are presented in some of the figures and were calculated using method two as outlined in [156]. However, it should be kept in mind that the entire idea of a confidence interval is to draw general conclusions from a sample, which is assumed to consist of randomly and independently selected observations, about a large underlying population as a whole. In other words, the confidence intervals aim to accommodate the sampling error of the observed sample with respect to a given population size. It is questionable if

the sample of pathogens presented in this thesis meets these requirements and assumptions. First of all, though the reviewers most certainly did not include the entirety of bacteria and viruses known to affect the predefined livestock species, the sample will probably cover a large proportion of it. As a result would a confidence interval of the entire or perhaps almost entire population be of no use. Secondly, the applied search strategy and search procedure seems to render a random and independent selection of these livestock pathogens unlikely. Therefore, the applicability and/or the benefit of the confidence intervals should be viewed with care.

Chapter 3

Summary statistics

Table 3.1 lists the numbers of bacteria and viruses in the resulting dataset. Viruses have been classified according to the ICTV[1] classification up to the family level but also according to the Baltimore classification which is based on the different mechanisms of viral mRNA production.

	distinct pathogens	species	genera	families	orders	classes	phyla	Baltimore classes
bacteria	133	94	44	31	21	13	7	NaN
viruses	69	47	26	11	NaN	NaN	NaN	5

Table 3.1: Number of bacteria and viruses: NaN=Not a Number, i.e. no value available for the corresponding class.

[1] International Commitee on Taxonomy of Viruses.

3.1 Trends associated with host-pathogen interactions

3.1.1 Host range of pathogens

	this study	ref. [27]	ref. [143]
# pathogen spp.	141	1922	415
% multihost spp.	60	63	68

Table 3.2: Multihost pathogens (A)

Let us start the observation by looking at the proportion of multihost pathogens. Table 3.2 compares the rounded proportions of multihost pathogens discovered in different studies including the one at hand. The first row of the table states the number of incorporated "pathogen" species for each study. Reference [27] is a study of pathogens of humans and domestic mammals, i.e. livestock (cattle, sheep, goats, pigs and horses) as well as carnivores (dogs and cats). Publication [143] looked at "pathogens" of wild primate hosts, i.e. more precisely 119 distinct primate species in total. Both of these studies included viruses, bacteria, fungi, protozoa and helminths. Only reference [143] additionally included arthropods. The study at hand, however, focused on viruses and bacteria of seven livestock hosts including mammalian and avian species. Although all three studies looked both at different sets of host species as well as different sets of pathogens, the proportion of reported multihost pathogens seems relatively similar.

	this study	ref. [28], i.e. [27]	
taxonomic pathogen level	species	species	
included pathogen types	v,b	v,b,f,p,h	v,b
# distinct pathogens	109	504	248
% multihost pathogens	57	79	73

Table 3.3: Multihost pathogens (B)

As already mentioned the figures of table 3.2 reflect studies of fundamentally different nature. The authors of reference [27] kindly provided us with a slightly updated version of the underlying dataset[28]. This allowed for a more detailed comparison of our results with the those of reference [27] or rather [28]. All figures in table 3.3 have been calculated using restricted subsets of the original datasets, i.e. only the host species *Bos taurus* (cattle), *Ovis aries* (sheep) and *Sus scrofa domestica* (pigs) have been taken into account. Moreover, the difference with regard to the included pathogen types was also addressed. Row two of table 3.3 lists the incorporated pathogen types, i.e. v=viruses, b=bacteria, f=fungi, p=protozoa and h=helminths. It can be taken from this table that the proportions of multihost pathogens now exhibit a much stonger variation between these two studies.

Probably one of the most obvious next steps would be to look at the proportion of multihost and zoonotic bacteria and viruses. Table 3.4 summarises both the corresponding figures as observed in our study and figures published previously by other authors. It can be seen that bacteria seem to outperform viruses in terms of the proportion of zoonotic as well as multihost spp., though not necessarily with convincing significance as the example of reference [143] shows.

	% zoonotic			% multihost			# spp.			
	this study	ref. [28]	ref. [182]	this	ref. [28]	ref. [143]	this	[182]	[28]	[143]
bacterial spp.	18.1	17.2	31.0	61.7	66.9	90.0	94	538	632	32
viral spp.	14.9	9.9	19.0	55.3	50.4	87.0	47	217	314	82
							2.0	2.48	2.01	0.39

Table 3.4: Host range, bacteria *vs.* viruses (A): The table compares bacteria and viruses on a species level with respect to the zoonotic label and multihost label. The figures in the first two blocks state the corresponding proportions. For example, 18.1% of all bacterial spp. recorded in our study are zoonotic. The third block on the far right states the number of spp. recorded in each study. The additional row below shows the ratio of bacterial spp. to viral spp.

The proportion of zoonotic bacterial and viral spp. reported in reference [182] turned out to be higher than in the study at hand. The Venn diagram 3.1 addresses a difference between these two studies. The pool of zoonotic pathogens contributing to this very study can be expected to be incomplete since the human host was not supposed to be in the main focus in the first place whereas it was in study [182]. However, it is debatable if this caused the difference because study [28] also included the human host but still differs from the figures reported in study [182].

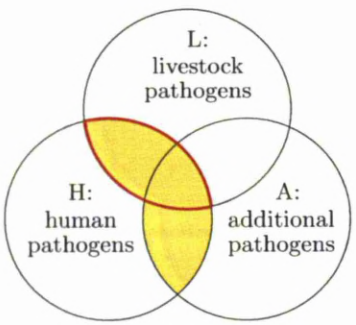

Figure 3.1: Comparing this study with study [182]: The red-framed area represents the pool of zoonotic pathogens that was included in our study, i.e. $L \cap H$ ($\cap$ = intersection). The area highlighted by a yellow background indicates the corresponding pool of zoonotic pathogens that was included in [182], i.e. $H \cap (L \wedge A)$ ($\wedge$ = union).

As already discussed in the introduction the multihost label is of questionable value when evaluating host-specificity. Table 3.5 therefore uses host-range categories as suggested in [145]. Bacteria seem to have a particular higher proportion of family-specific as well as phylum-specific species as compared to virus species. Viruses on the other hand show an increased proportion with respect to the multi-phylum category. Interestingly the genus-specific category is strikingly underrepresented regarding both bacteria and viruses.

	% species-specific	% genus-specific	% family-specific	% order-specific	% class-specific	% phylum-specific	% multi-phylum
94 bacterial spp.	38.3	0.0	17.0	6.4	11.7	17.0	9.6
47 viral spp.	44.7	2.1	4.3	10.6	14.9	0.0	23.4

Table 3.5: Host range, bacteria *vs.* viruses (B): For example, 38.3% of all bacterial spp. are species-specific, i.e. infect exactly one host species only. In other words 61.7% of the bacterial spp. are multihost pathogens which are assigned to the remaining host-range categories.

It might be worthwhile to examine the host range issue of bacteria and viruses from yet another point of view. In figure 3.2 every distinct pathogen (strain level) has been assigned to exactly one cell in the grid. The idea is to look at the host range by distinguishing the livestock market from any additional host species. On the other hand the figure also tells us something about the host range within each of these categories. The decision which host species should be regarded as part of the livestock market is clearly strongly influenced by the authors subjective point of view.

The numbers of both distinct bacteria and distinct viruses peak in the bottom far-left corners of the figure. These are the areas which refer to pathogens that are specific to any single host of the main seven livestock species. Keep in mind that only pathogens which affect this very group have been taken into consideration in the first place, hence there are no pathogens specific for any other host species. It can also be seen that a large number of pathogens spread around these peak areas, in fact in both directions. However, a general tendency towards a higher number of host species within the livestock market can be observed. This again is likely to be a reflection of the requirement for pathogen inclusion. Bacteria seem to have an inclination to a higher host range within the livestock market as compared to viruses. Some of them appear to be specific to the livestock market (see black arrows) whereas others are able to infect also a considerable number other host species (see red arrows).

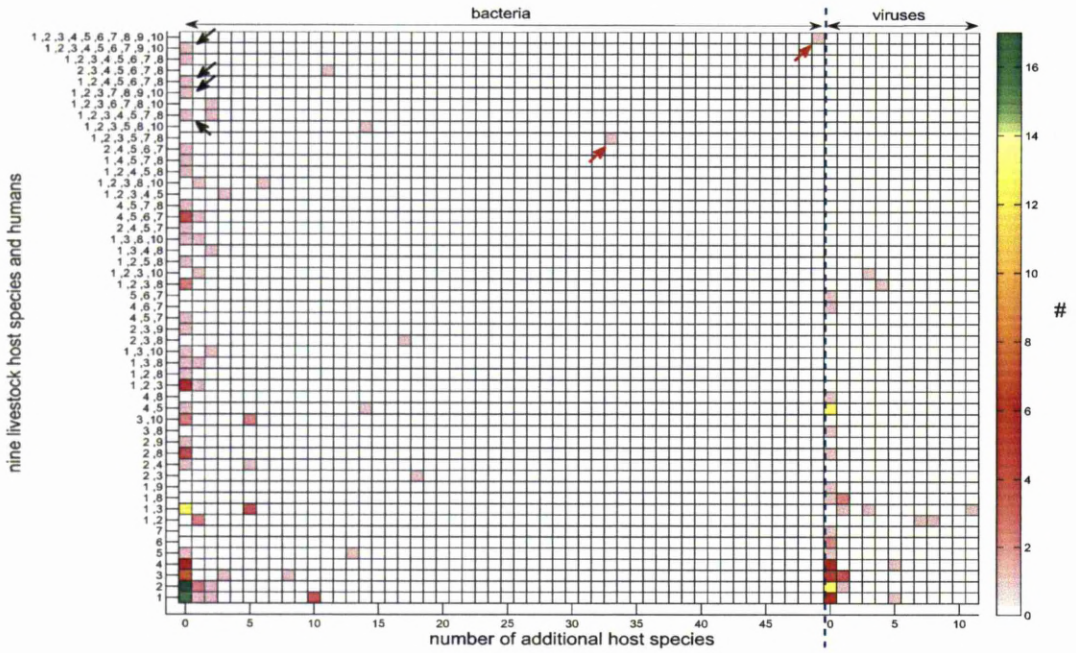

Figure 3.2: Host range of distinct bacterial and viral pathogens (strain level): The ordinate depicts a group of host species representing the livestock market, i.e. **1**=*Bos taurus*, **2**=*Sus scrofa domestica*, **3**=*Ovis aries*, **4**=*Gallus gallus domesticus*, **5**=*Meleagris gallopavo*, **6**=*Anser anser*, **7**=*Anas domesticus*, **8**=*Homo sapiens*, **9**=*Equus caballus*, **10**=*Capra aegagrus hircus*. The abscissa shows the number of any additional host species that can be infected. Each distinct pathogen is assigned to exactly one cell. Note that the figure is split into two sections, one for the bacteria (left)and one for the viruses (right). The two red arrows indicate *Mycobacterium avium* (top right) and *Clostridium botulinum*.

3.1.2 Pathogen richness of hosts

The next feature spotlights the hosts, namely their pathogen richness. This issue is approached in figure 3.3 simply by plotting the number of pathogens infecting certain host species. Given the large number of host species, a threshold regarding a minimum number of reported pathogens for each host species was applied. The figure is also divided into four separate subfigures to account for variations with respect to different levels of the pathogen taxonomy. It is obviously questionable if, for example, the species level of bacteria and viruses are comparable. Nevertheless, it is interesting to observe which other host species share a given minimal number of pathogens with the seven predefined livestock species. First of all, humans (*Homo sapiens*) clearly show the highest number of pathogens in common. Other prevalent livestock hosts such as goats (*Capra aegagrus hircus*),

horses (*Equus caballus*), even camels (*Struthio camelus*) as well as common pet animals like cats (*Felis catus*) and dogs (*Canis lupus familiaris*) also appear in the figure. The remaining host species on the figure are either ticks or they are known to occur frequently within or in the proximity of habitats shaped by humans including agricultural and farmland. In fact, because of this behaviour hosts such as rats (*Rattus spp.*), the Wild Boar (*Sus scrofa*), the Red Deer (*Cervus elaphus*), the Rock Pigeon (*Columba livia*), the House Sparrow (*Passer domesticus*), the Grey Partridge (*Perdix perdix*), the Mallard (*Anas platyrhynchos*), the Common Pheasant (*Phasianus colchicus*) and the Rook (*Corvus frugilegus*) are also referred to as synanthrops and/or hemerophillics[2].

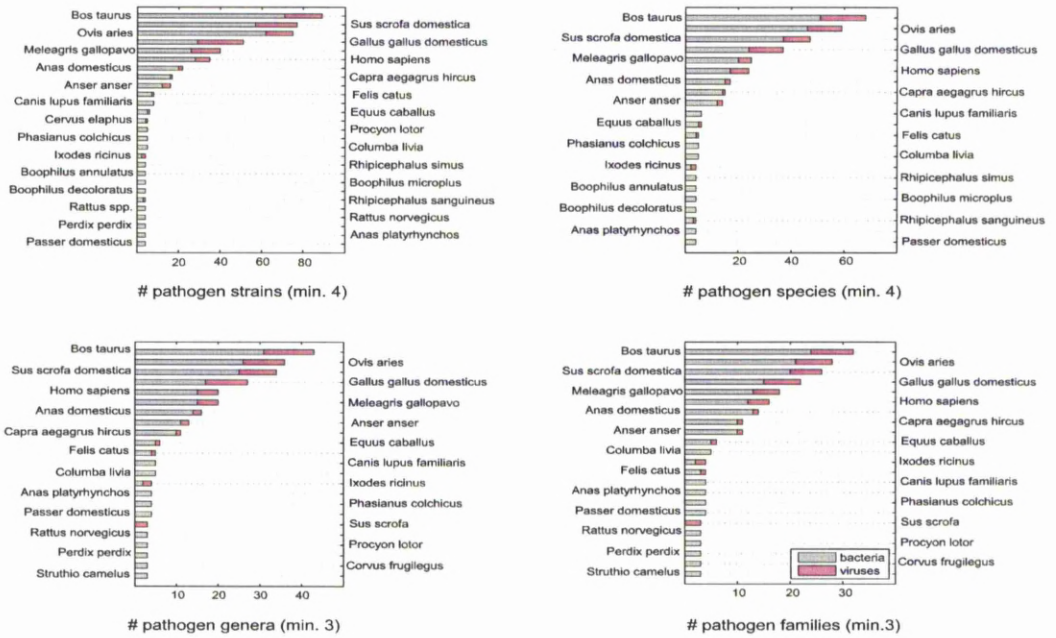

Figure 3.3: Pathogen richness of host species: The figure is separated into four subfigures referring to different taxonomic levels of the pathogens. The bars depict the total number of pathogens. Note that the part referring to viruses (red only) has been added to the bacterial part (grey only). The names of the host species have been attached to the ordinate in an alternating manner from left to right for better visibility.

[2]German (Germany): Kulturfolger.

The following comparisons are only sensible for the main seven livestock species because of the constraint for pathogen inclusion. Figure 3.4 compares the average body weight of the seven main livestock species with their associated number of pathogens. It can be seen that the number of observed pathogens does not simply increase with body weight. Chicken, for example, have been associated with a large number of pathogens, in particular viral pathogens. It is therefore questionable if these graphs reveal a trend of higher pathogen richness towards host species with larger body weight.

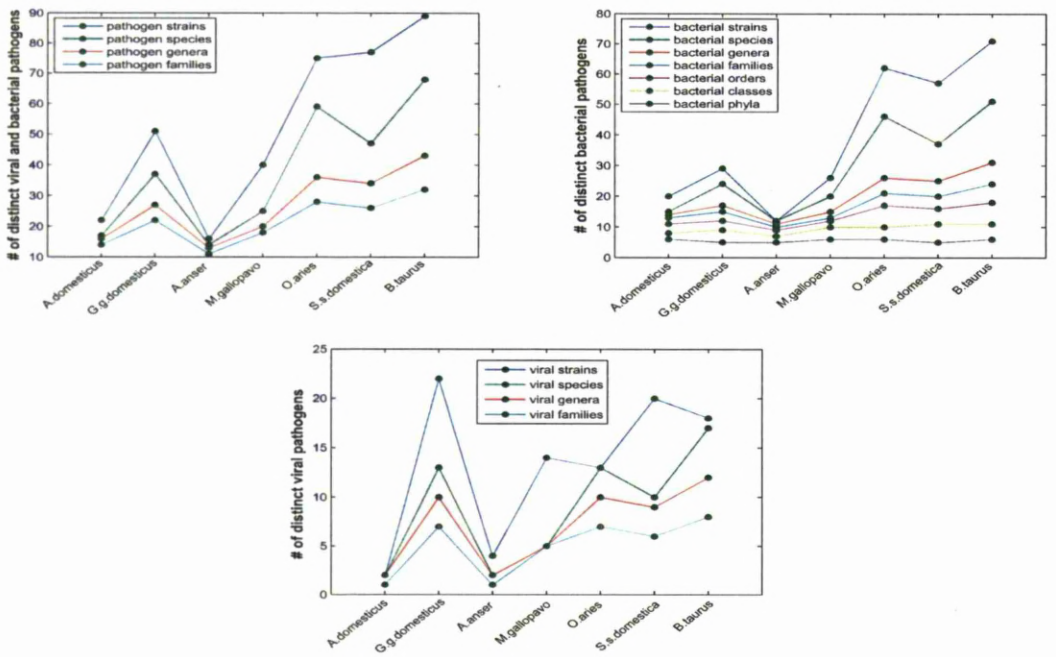

Figure 3.4: Pathogen richness *vs.* body weight: The figure is separated into three distinct subfigures, i.e top-left referring to all pathogens, top-right bacteria only and bottom to viruses only. The abscissa depicts the main seven livestock species ordered with increasing body weight from left to right whereas the ordinate refers to the number of pathogens discovered for each host. The weight can vary considerably between animal breeds and sexes of the same species. Here, the maximum reported weight for a species, i.e. any breed or sex, as stated in reference [160] was used as a rough indicator of the potential weight. Ducks (*A.domesticus*) can weigh up to 4 *kg*, chickens (*G.g.domesticus*) 6 *kg*, geese (*A.anser*) 12 *kg*, turkeys (*M.gallopavo*) 20 *kg*, sheep (*O.aries*) 140 *kg*, pigs (*S.s.domestica*) 350 *kg* and cattle (*B.taurus*) 1500 *kg*.

In order to complete the picture of pathogen richness, the concept of host-range categories for

pathogens has been incorporated into figure 3.5. The figure suggests differences in the host-range pattern of the pathogens between the main seven livestock species. Ducks and geese, both members of the family Anatidae and order Anseriformes, show a large gap regarding genus-, family- and order-specific pathogens. On the other hand their pathogens are particularly well represented by the phylum- and also class-specific categories. Pathogens affecting chickens and turkeys, which belong to the order Galliformes, although being less represented by the phylum- and class-specific categories as compared to ducks and geese, show a considerable proportion of order-specificity. Differences can even be observed regarding the three mammalian host species. The two members of the Bovidae family, i.e. cattle and sheep, exhibit comparatively balanced proportions over all but the genus-specific category. Pigs, however, belong to the family Suidae and their pathogens appear to fall into certain categories with greater preference. For example, pigs show a high proportion of host-specific (i.e. species-specific) pathogens but no family-specific-pathogens. Actually the proportion of host-specific pathogens in pigs constitutes the highest overall, at least on the strain level. At the species level, though, it roughly matches that of chickens. In fact, pigs and chickens seem to have a relatively comparable pattern.

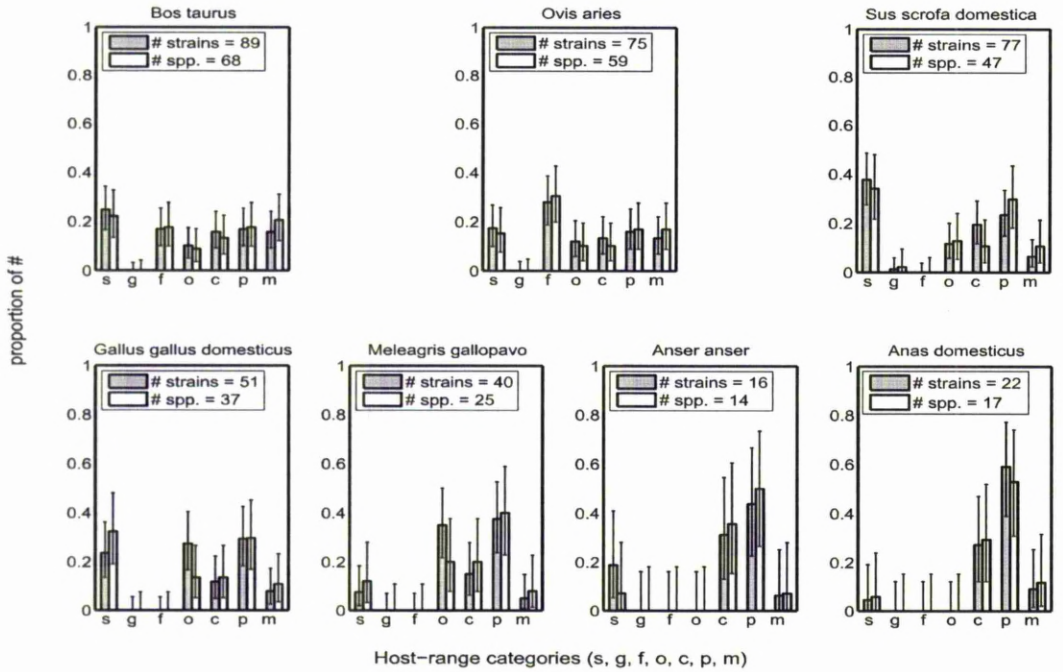

Figure 3.5: Pathogen richness with respect to host-range categories: The figure is subdivided into seven distinct figures, one for each main livestock species. The abscissa depicts the host-range categories, i.e. **s**=species-specific, **g**=genus-specific, **f**=family-specific, **o**=order-specific, **c**=class-specific, **p**=phylum-specific, **m**=multi-phylum. The ordinate shows the proportion of pathogens (bacteria and viruses together) that have been assigned to each of these categories given the total number of pathogens (both for the strain and species level) that infect the host species in focus.

3.1.3 Transmission

Both the host range of pathogens as well as the pathogen richness of hosts will depend on the dynamic contact structures between hosts and pathogens which are established via transmission in the first place. Rather than comparing and analysing individual transmission routes, it appears more prudent to start by searching for broader trends. Hence, the avoiding of the term transmission route.

Figure 3.6 aims at comparing several of the already encountered labels for pathogens with their tendency to be involved in certain transmission pathways. Overall it can be observed that multihost, bacterial and zoonotic pathogens outperform the corresponding singlehost, viral and non-zoonotic pathogens in terms of non-close/indirect transmission pathways. When focusing on pathogens

which solely transmit via close/direct host-to-host contact, however, this pattern reverses. But figure 3.6 offers far more detailed insights. For example, when considering the applied confidence intervals it can be found that the multihost pathogens outperform the singlehost ones particularly in the involvement of environments (E), E *AND* mechanical vectors (MV) and also anthropogenic mechanical vectors (MV-anthrop.). Roughly the same holds true when comparing zoonotic with non-zoonotic pathogens. Comparing bacteria with viruses, however, things are slightly different. Bacteria show a higher proportion with respect to the involvement of environments but also environments *AND* explicitly no anthropogenic MV's. Then, they are clearly outperformed by viruses in the group of close/direct host-to-host contact only.

The bottom-left subfigure adds another level of complexity to the bacteria and viruses. It can now be noticed that the differences between multihost-bacteria and viruses increased and that they are particularly severe with respect to the singlehost-viruses.

The zoonotic label is used in this context as a specific form of the multihost label. Observed differences between those pathogens bearing the zoonotic and those bearing the non-zoonotic label might therefore be due to the impact of singlehost pathogens. For this reason the bottom-right subfigure additionally depicts the proportions for non-zoonotic pathogens which are at the same time labelled as multihost. The impact of the singlehost pathogens appears to be negligibly small.

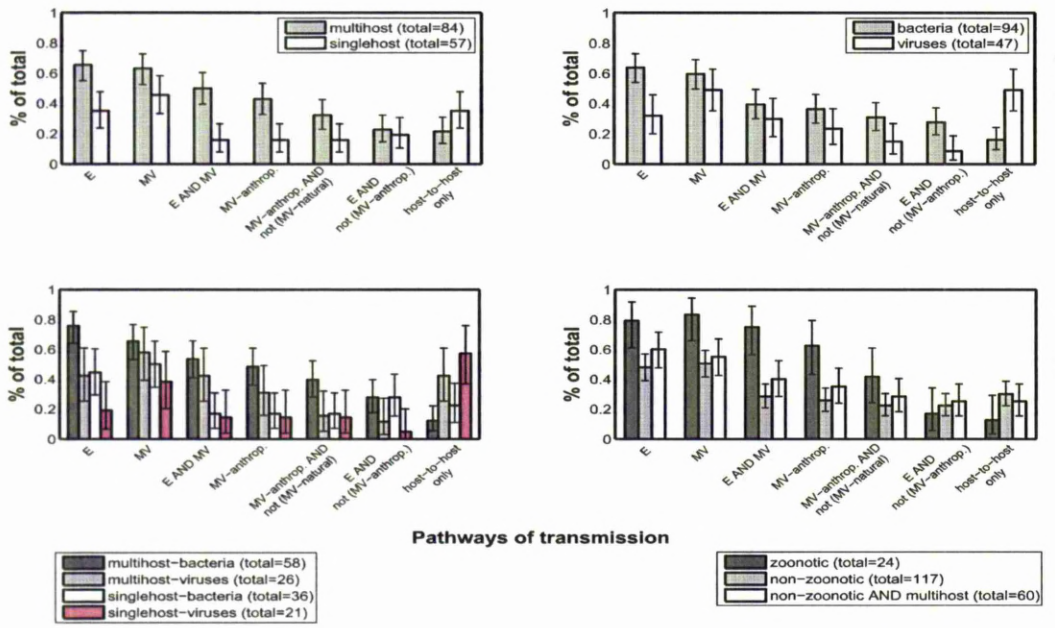

Figure 3.6: Pathogen labels *vs.* transmission pathways: The abscissas depict several items or rather group of items through which a host might receive an infection, i.e. **E**=environments, **MV**=mechanical vectors, **E AND MV**=both, **MV-anthrop.**=anthropogenic MV's, **MV-anthrop. AND not(MV-natural)**=anthrop. MV's but explicitly not natural MV's, **E AND not(MV-anthrop.)**=E but explicitly not anthrop. MV's, **host-to-host only**=explicitly host-to-host transmission via close/direct contact only. The ordinate states the proportion of the total number of pathogen spp. as show in the corresponding legends. For example, roughly 65% of all multihost pathogen spp. can reportedly be transmitted to, i.e. enter, at least one host species through any of the environment items. This does obviously not exclude these pathogens from being transmitted in other ways too.

It would be interesting to see if the dataset of transmission routes holds any evidence supporting a relation between the host range of pathogens and their transmission pathways. This issue is approached in figure 3.7. Since the dataset is potentially biased due to data collection and/or issues regarding the taxonomy of pathogens[3], it was deemed appropriate to look at several taxonomic levels of the pathogens. This is simply because such biases might potentially smooth out at a higher taxonomic level.

[3]It seems particularly questionable if bacterial and viral taxa can be treated equally.

One hypothesis could be that the host range increases with the pathogen's ability to be transmitted via non-close/indirect pathways. Figure 3.7 does in fact suggest such a trend. This can be observed in particular in the subfigures C1 to C3. The proportion of pathogen genera[4] that can be transmitted via environments, mechanical vectors or both increases with the host-range as represented by the host-range categories. One might consequently hypothesize that pathogens which are entirely reliant on close/direct contact transmission should exhibit narrower host ranges. Looking at subfigures A6, B6 and C6 many of these very pathogens have been assigned to the species-specific host-range category. However, a large number of these pathogens can also be encountered in the multiphylum category. Moreover, they also "peak"[5] in the class-specific categories of subfigures A6 and B6. These inconsistencies need more detailed study.

[4]That is at least one member of a genus has to fulfil the requirement.

[5]Given the magnitude of the error bars, it is debatable if the term 'peak' applies here.

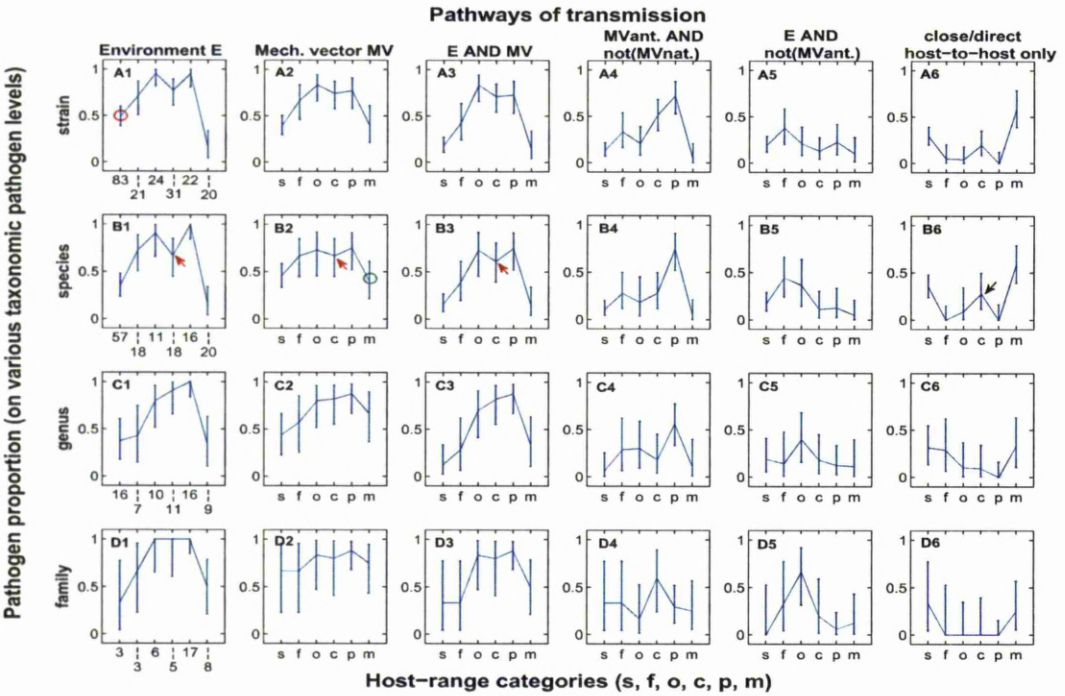

Figure 3.7: Host-range categories *vs.* transmission pathways: The abscissas hold the host-range categories **s**=species-specific, **f**=family-specific, **o**=order-specific, **c**=class-specific, **p**=phylum-specific and **m**=multiphylum. Note that with respect to the first column of subfigures these category labels have been replaced by the total number of pathogens assigned to each category. Each row of subfigures corresponds to one taxonomic pathogen level. The ordinates show the proportion of pathogens (both bacteria and viruses together) for each host-range category that involve the corresponding group of items through which a host might receive an infection. See text of figure 3.6 for further details on these groups. For example, ca. 50% out of 83 species-specific pathogens (strain level) can reportedly receive any infection via the group of environment items E (see red circle in subfigure A1). Note that E does not explicitly exclude MV and vice versa unless stated otherwise.

A striking characteristic in particular of subfigures A6 and B6 is a steep increase in the proportion of multiphylum pathogens. This contrasts with a steep drop of this proportion in most of the other subfigures. Table 3.6 reveals that a large number of the multiphylum pathogen spp. that are entirely dependent on close/direct contact transmission are in fact transmitted between hosts belonging to the phyla Arthropoda and Chordata, the only two phyla in the dataset. More precisely, they are transmitted between biting ticks or insects and vertebrates. The figures of table 3.6 furthermore suggest that 40% (eight out of twenty) multiphylum pathogen spp. can be transmitted by non-close/indirect pathways. Interestingly this proportion only appears in the multiphylum category of subfigure B2 (see green circle). This is simply due to the fact that these pathogen spp. can be

transmitted by arthropod bites even if the arthropod itself is not infected but rather carries the pathogen passively on its mouth parts, i.e. acts as a mechanical vector.

This does not explain the aforementioned peak in the class-specific category (see, for example, black arrow in subfigure B6). This peak, however, appears to have reverse counterparts in subfigures B1 to B3 (see red arrows). Looking at the entirety of host species which are affected by the class-specific pathogens of subfigure B6, only hosts that can be assigned to the livestock market can be discovered, namely cattle, pigs, sheep, horses, humans, chickens, turkeys, ducks and geese. The class-specific pathogens of subfigures B1 to B3, on the other hand, are able to infect a multitude of ca. 48 different host species including the ones just mentioned but also dogs, cats, alpacas, red deer, racoons, skunks, rats and various avian species.

Another interesting behaviour seen in the non-close/indirect subfigures is that the trends observed in the first three subfigures are partly disappearing in the fourth and even more so in the fifth subfigures. In these subfigures pathogens which could be transmitted via certain types of mechanical vectors have been explicitly excluded. Furthermore, the trends turn out to be most convincing in the third subfigures which refer to pathogens which can be transmitted both by environments as well as mechanical vectors. It can also be seen that the third, fourth and fifth subfigures show a somewhat smaller proportion in the "lower" host-range categories, namely species-specific and family-specific category, as compared to the first and second subfigures.

	species-specific	family-specific	order-specific	class-specific	phylum-specific	multi-phylum
# pathogen spp.	57	18	11	18	16	20
# close/direct host-to-host only	18	0	1	5	0	12
# Arthropoda ↔ Chordata only	0	0	0	0	0	10

Table 3.6: Direct host-to-host transmission and arthropod hosts: The table highlights the situation of subfigure B6 in figure 3.7. The first row states the number of pathogen spp. that have been assigned to each host-range category (outlined in the column headers) in total. The second row shows the number of pathogen spp. that are entirely reliant on direct host-to-host transmission within each category. The last row is a further specification of the previous one. It gives the number of pathogens that can be transmitted only via close/direct contact between Chordata and Arthropoda hosts.

The dataset of transmission routes also offers potential insights regarding the exit materials and entry portals. As can be taken from the lower part of figure 3.8 are there at least four particularly prevalent pathways with respect to close/direct contact transmission of both bacteria and viruses together. These are:

1. Inhalation of respiratory excretions;
2. Ingestion of faeces;
3. Mother-to-offspring either *in utero*/congenital or via breastfeeding;
4. Sexual contact.

Figure 3.8 also suggests differences in the usage of exit-entry pairs between bacteria and viruses with respect to close/direct contact transmissions. For example, whereas 30% of all bacterial spp. can be passed on from a mammalian female to its offspring transplacentally, i.e. *in utero*, only 9% of viral spp. where recorded to do so. Intriguingly, this seems to invert for *in utero* transmissions of avian, i.e. poultry, hosts. Here, only 8% of the bacterial but 24% of the viral spp. were found to be transmitted from the mother to unborn offspring, i.e. via the egg. Furthermore, the figure shows a number of exit-entry pairs which have been exclusively assigned to either the bacteria or the viruses. Also, many of these bacteria- and virus-specific exit-entry pairs appear in the upper part of the figure. In other words, they have hardly been suggested in the literature anyway.

As can be taken from figure 3.8 too, certain taxonomic groups of pathogens are particularly strongly represented in the dataset, e.g. Proteobacteria, Gram-positive Firmicutes and double-stranded DNA-viruses. The six capital letters A to F point at those pathogen spp. which have been found to exploit at least ten of the listed exit-entry pairs. Although the dataset contains only a small group of six single-stranded RNA-virus species, three of them have been found to make use of a considerable number of exit-entry pairs.

As a final remark on figure 3.8 it should be noted that although the respiratory tract, i.e. inhalation etc., appears to be an important entry portal, overall the entry portal alimentary tract could be more important both with respect to close/direct as well as non-close/indirect transmission pathways. Respiratory tract appears in only nine of the listed exit-entry pairs as compared to thirteen occurrences of alimentary tract.

Figure 3.9 takes a look at the entry portal issue from another perspective. Here six overall categories with respect to the actual source of infection were distinguished. The alimentary tract, i.e. the oral uptake of pathogens including both bacteria and viruses, now appears to be the most commonly used portal of entry. This is in fact the case for both close/direct contact and non-close/indirect sources of infection such as environments and mechanical vectors. However, this does not hold true for what could be called airborne infections. For such a scenario the respiratory tract turns out to be the main - for viruses apparently the sole - entry portal. Moreover, only a small percentage of bacterial spp. were found to enter any host via intact or damaged skin when derived from anthropogenic mechanical vectors. In contrast viruses show a remarkably high proportion of spp. which can enter any host via intact skin after being transmitted from anthropogenic mechanical vectors.

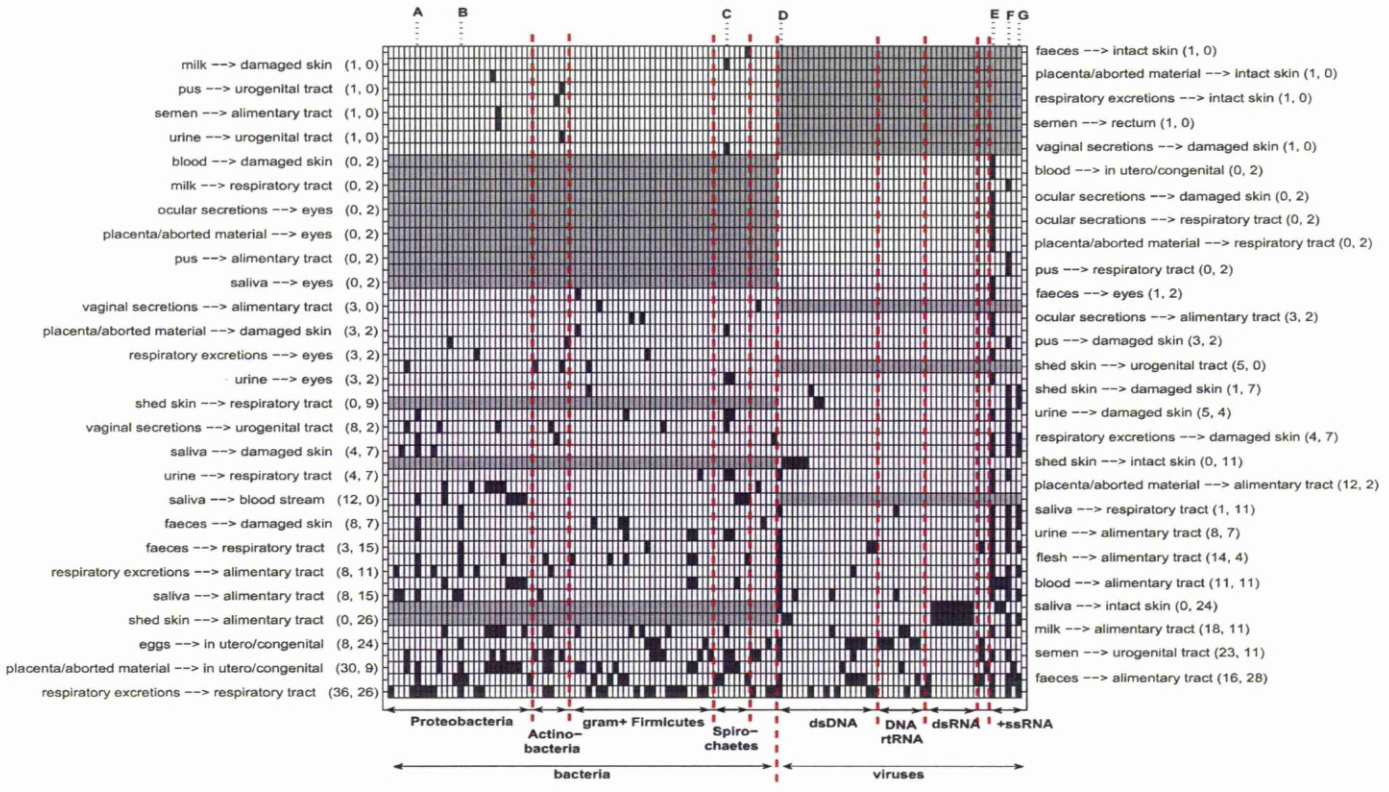

Figure 3.8: Exit materials and entry portals of close/direct host-to-host transmissions routes: The plot depicts the distribution of 'exit material'-'entry portal' combinations over pathogen spp. along the abscissa. The names for these exit-entry combinations or pairs have been placed along the ordinate in an alternating manner from left to right for better visibility. The figures in parentheses are the proportions of bacterial species and viral species which use the corresponding exit-entry pairs. The exit-entry pairs are ordered according to the sum of these percentages increasing form bottom to top. Cells are highlighted in dark grey if the corresponding pathogen spp. is reportedly transmitted via the exit-entry pair in focus. Light grey areas emphasise those exit-entry pairs which have not been reported at all either for bacteria or viruses. The letters A to G show the location of pathogen spp. which use at least ten of the exit-entry pairs (see following figures in parentheses). A=*Pasteurella multocida* (12), B=*Salmonella enterica* (10), C=*Leptospira interrogans* (11), D=*African swine fever virus* (13), E=*Classical swine fever virus* (30), F=*Foot-and-mouth disease virus* (23) and G=*Swine vesicular virus* (12).

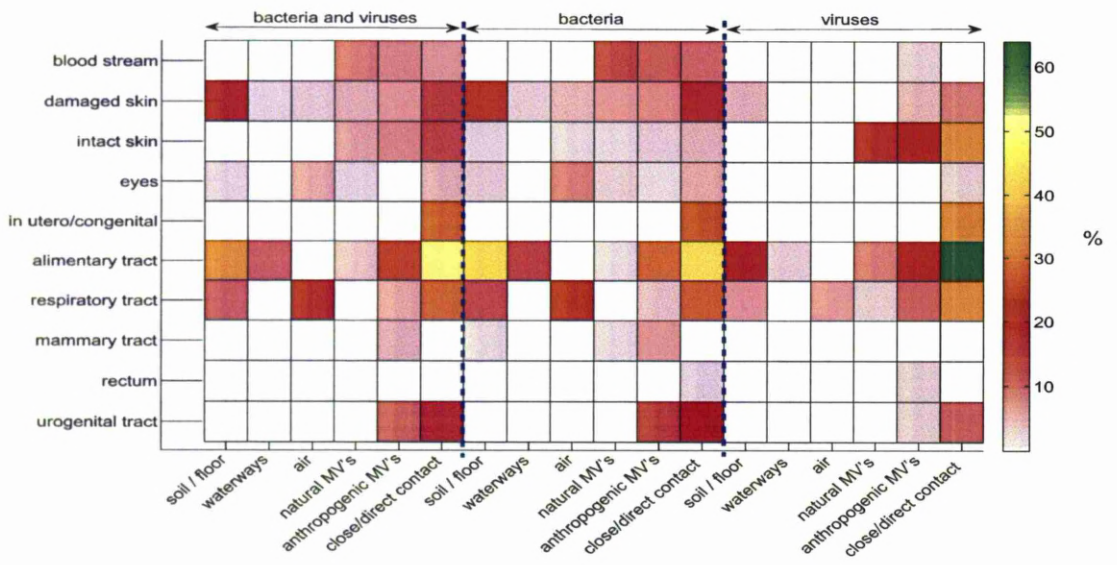

Figure 3.9: Entry portals *vs.* sources of infection – considering pathogen spp.: The grid displays colour-coded percentages regarding three broad groups of pathogens. For example, given all viral spp. in the dataset, roughly 65% have been reported to enter at least one host species through the alimentary tract during close/direct contact transmissions. Note that soil/floor is meant to include bedding. Note also that the percentages have been rounded and therefore in certain cases no colour appears in the group of bacteria and viruses whereas there is a minimal proportion in, for example, the bacterial group alone.

The examination of entry portals would not be complete without focusing the attention on the seven main livestock hosts and humans. Figure 3.10 applies the same six categories that we have encountered in the previous figure. The difference is that each segment of figure 3.10 takes those pathogen spp. (bacteria and viruses) into account that affect the corresponding host species only. The first eye-catching feature is that none of the pathogen spp. infecting ducks and geese gets transmitted by close/direct contact via the urogenital tract. In other words there seem to be no sexually transmitted pathogens for ducks and geese. Figure 3.10 uncovers another noteworthy aspect regarding the corresponding differences between these host species. Ducks and geese obtain a larger proportion of pathogen spp. via the oral uptake of environmental water as compared to the other host species.

Humans interestingly receive a high proportion of the recorded pathogen spp. through oral uptake via/of anthropogenic mechanical vectors (e.g. animal products such as meat, eggs and milk). They also get a large proportion of infections via the skin after close/direct contact with livestock animals. Though to a lesser degree, humans even seem to get infected with a certain proportion of pathogen spp. through the inhalation of contaminated air, ingestion of contaminated water and ingestion after being in close/direct contact with livestock hosts.

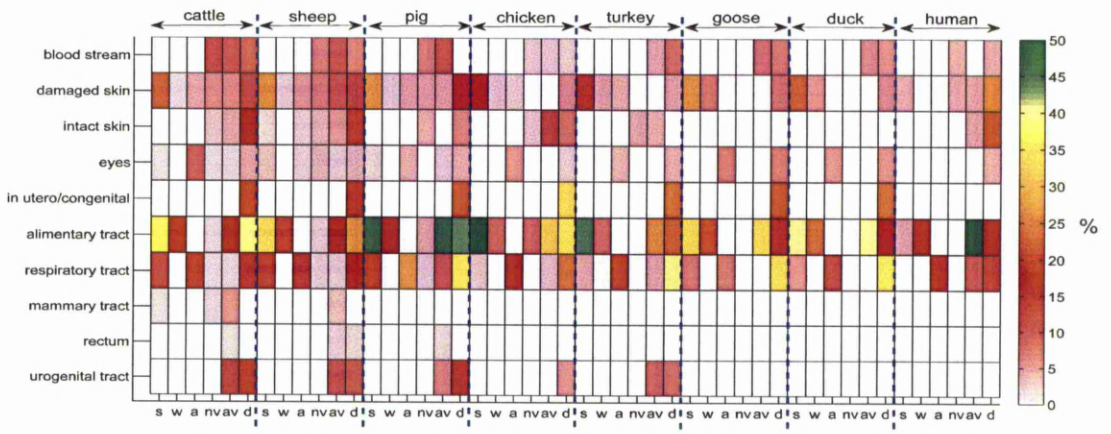

Figure 3.10: Entry portals *vs.* sources of infection – considering the hosts: The grid displays rounded, colour-coded percentages regarding the pathogens infecting the corresponding host species. For example, given all pathogen spp. in the dataset which are known to infect humans, ca. 50% enter the human host through the alimentary tract during transmissions involving any anthropogenic mechanical vector. **s**=soil/floor (includes bedding), **w**=waterways, **a**=air, **nv**=natural mechanical vectors, **av**=anthropogenic mechanical vectors, **d**=close/direct contact.

3.2 Discussion

Host range of pathogens:

The dataset of this study confirms previous findings with respect to the overall proportion of multihost pathogens. A large proportion of bacterial (55.3%) and viral (61.7%) spp. that affect livestock are multihost pathogens, contrary to the belief that evolution should favour singlehost (species-specific) pathogens [207]. Furthermore, it turned out that bacteria are more likely to be labelled both as multihost and zoonotic as compared to viruses, a finding which is also supported by other studies. This could suggest that bacteria are generally more sustainable outside a host and therefore more likely to make contact to new host species.

Bacteria and viruses also differed in their pattern of multihost-specificity with respect to the host-range categories as suggested in [145]. The usefulness of this categorisation is surely debatable but it offers at least a first step towards a better understanding of the multihost issue. For example, should a family-specific pathogen that can cause infection in one hundred hosts, belonging to ten different host genera be regarded as less host specific than a phylum-specific pathogen which merely affects two host species, even if these belong to different orders? It is questionable if the observed patterns, including the lack of genus-specific livestock pathogens reflect the true situation or if this is due to any form of bias. Bias might also explain the extreme number of host species discovered for certain bacteria.

It has also been emphasised that the host range can strongly vary depending both on the pathogen types as well as host types which are in focus of a study. Comparing different studies therefore needs to be carried out with due care.

Pathogen richness of hosts:

In terms of the pathogen richness of hosts the dataset highlights that particularly other livestock hosts, human beings which are part of the livestock environment, pet animals which frequently occur in close proximity to humans and farm areas and so-called synanthrops share the highest number of pathogens with the main seven livestock host species. This could support the assumption that a close [6] contact structure is a beneficial parameter in the spread of infections between different host species.

The results presented do not support the assumption that pathogen richness with respect to bacteria and/or viruses correlates with the host's average body weight. Especially chickens showed a strong discrepancy in this respect. It could be expected that the results are in fact superimposed by biases in the literature and research effort spent on each host. On the other hand, chickens might as well exhibit a more versatile diet due to their omnivorous feeding habits as compared to turkeys, ducks and geese. Broad diets have been suspected as contributing to increased "pathogen" (at least referring to helminths) richness [151].

Pathogen richness of hosts might also be viewed from a slightly different perspective. Figure 3.5 suggested characteristic patterns of hosts when considering the host ranges of their associated pathogens. Given the indicated confidence intervals, these results should be treated with care. Nevertheless, they might well reflect underlying differences regarding the hosts. For example, most pathogens affecting ducks and geese have been assigned to the class- and phylum-specific host-range categories and none of them were found restricted to the genus-, family- and order-specific categories. In other words, ducks and geese seem to have a tendency for infections caused by pathogens with a broader host range. Ducks and geese belong to a group of birds commonly known as waterfowl which refers both to their swimming skills and consequently to their preferred habitat, i.e. waterways. Water most doubtlessly has a major attraction to all living creatures thereby possibly increasing the chances for these pathogens to get in contact with a large number of hosts from various taxonomic backgrounds.

It is also striking that the corresponding pattern of pigs appears to be most similar to the pattern of chickens. Pigs and chickens are apparently kept in particularly intensive conditions as well as particularly huge populations sizes [11]. Moreover, these host species are also both omnivores, which feed on both plants and animals. However, it seems questionable if such farming or feeding habits can explain the higher proportion of host-specific pathogens in pigs and chickens. Perhaps the feeding habits increase the risk to get in contact with new pathogens or rather with pathogens which have not been noticed in other host species yet.

[6] Close referring to spatial proximity, i.e. non-close/indirect modes of transmission might be involved too.

Transmission:

The findings concerning the pathways of transmission revealed that multihost, zoonotic and bacterial pathogens are proportionally more likely to exploit non-close/indirect pathways than are singlehost, non-zoonotic and viral pathogens. The difference in the proportion of pathogens which have been reported to take advantage of close/indirect contact transmission is particularly significant between multihost-bacteria and singlehost-viruses. However, these trends inverted when looking at the close/direct contact transmission between hosts only.

Using the host-range categories as suggested in [145], it was shown that the ability to exploit non-close/indirect transmission pathways could potentially be linked to the actual host range of a pathogen. In fact, the trends observed in the first and second subfigures partly disintegrated after the explicit exclusion of mechanical vectors (see fourth and fifth subfigures) which suggests a key role of mechanical vectors in reaching broader host ranges for pathogens affecting livestock. This holds particularly true for the anthropogenic mechanical vectors (see fifth subfigures), a finding which seems in accordance with the idea that population migration, movement of goods and people, food production and new medical devices constitute important factors of (human) disease emergence [125].

To recapitulate, figure 3.7 suggested three strategies by which a pathogen might broaden its host range:

1. Exploiting non-close/indirect contact transmission pathways;
2. Getting access to close contact networks consisting of multiple host taxa, for example, the livestock environment;
3. Infecting arthropod hosts which are able to transmit to or between various host taxa.

Moreover, with respect to close/direct transmissions it has been discovered that the inhalation of respiratory excretions, the ingestion of faeces, the *in utero*/congenital pathway as well as sexual contact appeared to be of particular importance. Another potentially important pathway exists between blood-sucking and/or biting arthropod hosts and vertebrate hosts. This pathway also highlights an intrinsic shortcoming of the framework, i.e. the vulnerability to individual interpretation and subjective use of the framework. The corresponding transmission of bacterial pathogens has been modelled as follows: An arthropod host 'transmits to' a vertebrate host via the exit material 'saliva' and the entry portals 'blood stream' and/or 'damaged skin'. The people focusing on the viral pathogens, however, modelled this pathway in a different manner: An arthropod host 'transmits to' a vertebrate host via the exit material 'saliva' and the entry portal 'intact skin'[7]. Therefore, without this discrepancy this pathway would have shown up further down in plot, i.e. with a higher overall percentage. On the other hand could this also be due to real differences since many viruses apparently are not injected directly into the blood stream, but are taken up by dendrocytes in the skin [137] [113] [135].

[7]It is questionable if 'damaged skin' has also been applied for viruses.

Overall, i.e. when considering also non-close/indirect pathways, ingestion turned out to be the most important portal of entry. Differences regarding the entry portals and sources of infection were also detected between host species.

The highest number of reported exit-entry pairs with respect to close/direct contact transmissions have been found for positive single-stranded RNA viruses. This might be caused by their higher mutation rate and therefore potentially increased adaptability. On the other hand these pathogens do not necessarily show broad host ranges. For example, the *Classical swine fever virus* exhibits the highest number of exit-entry pairs but also a very limited host range (data not shown). This could either be interpreted as a contradiction to the adaptability concept or it might suggest particularly poor sustainability outside a host thereby preventing the virus to get in touch with other host species.

It should be noted that the differences in the usage of exit-entry pairs as suggested by the light-grey bars in figure 3.8 could be partly due to interpretative ambiguities of the framework and/or literature information. Furthermore, there is also always the possibility that the literature information is simply wrong. For example does one of the main textbooks state that "infection of an animal ..." [152] with *Corynebacterium pseudotuberculosis* "... is facilitated by the presence of skin wounds but the organism can invade through intact skin" [152] too. The cited reference [126] that allegedly supports the claim of invasion through intact skin does in fact not mention anything alike at all.

Chapter 4

Host centrality and reemergence frequency

The term "transmission network" is commonly used in the study of the dynamics and spread of infectious diseases between individuals of a specific population or between individuals of distinct populations, distinct species etc. The transmission framework outlined in section 2.1 enables us to categorise modes or pathways of transmission as suggested in the literature. This leads to a set of transmission routes for each pathogen which finally form pathogen-specific "networks of transmission routes". These networks are fundamentally different from the transmission networks mentioned earlier. First of all, they look at a different scale, i.e. they are not an individual-based concept. Secondly they are merely static descriptions or summaries of suggested, assumed and perceived transmission modes. Nevertheless, they might still bear crucial information about a pathogen's potential to emerge and reemerge. In fact, it might even be possible to differentiate host-pathogen interactions based on the information stored in the networks of transmission routes. My hypothesis is that the position of a host vertex within its pathogen-specific network of transmission routes may reveal important insights into its potential for corresponding outbreaks and reemerging host-pathogen interactions. Henceforth the term 'transmission network' will be used with reference to 'networks of transmission routes'.

In order to approach the hypothesis, information on the frequency of occurrence of host-pathogen interactions is required. Comprehensive, well-recorded outbreak statistics, however, are a very rare commodity. Freely available datasets of adequate quality focus on serovars of the bacterium *Salmonella enterica* subspecies *enterica*. *Salmonella enterica* is therefore the organism of choice for this chapter.

4.1 Materials and methods

4.1.1 The transmission framework

It was decided to adjust the established transmission framework of section 2.1 for this chapter only. It is hoped that these adjustments will improve the quality of the data but also facilitate the application of the framework and speed up the process of data collection by neglecting redundant transmission routes. For example, paying equal attention to all affected host species with respect to the transmission routes should clearly improve the comprehensiveness and informational value of the resulting transmission networks. On the other hand, omitting 'natural mechanical vectors' seems justified because their contribution for disseminating pathogens will either be covered by close/direct host-to-host contacts or environmental factors. On top of that, it is extremely difficult and time-consuming to find evidence for this kind of mechanical vector since actually each host species also constitutes a natural mechanical vector itself. The adjustments included:

1. **Mechanical vectors:** Only three categories were applied, i.e. 'animal foodstuff', 'animal products' and 'people', whereby the last one also included 'medical equipment', 'animal equipment' and 'vehicles'. Animal drinkers have been assigned to the item 'animal foodstuff'.
2. **Environments:** The item 'bedding' was merged into the item 'man made'.
3. **Entry portals:** The item 'intact skin' was also used for serovars which are the cause of eczemas, skin rashes, pustules or infective dermatitis.
4. **Host species:** All affected host species were treated equally, i.e. there was no distinction between different host groups as depicted in figure 2.3.
5. **Transmission routes:** Transmission routes of the type 'environment-to-environment', 'mechanical vector-to-mechanical vector', 'environment-to-mechanical vector' and vice versa were taken into consideration (see figure 2.3).

Moreover, the data collection was carried out in two steps resulting in two datasets , i.e. a basic version and an extended version. In the first step a transmission route was either supported by very specific evidence from the literature or it resulted from a set of basic assumptions that were thought to be particularly relevant or obvious:

1. Every host species (except the human host) contaminates 'soil' with its associated serovar(s) via 'faeces'.
2. Individuals of the same host species are *per se* capable of transmitting their associated serovar(s) between each other via the faecal-oral route, i.e. 'faeces' and 'alimentary tract'.
3. Every host species contaminates 'soil' with its associated serovar(s) via any of the following exit materials if the exit material has been reported for the host-serovar interaction: 'urine', 'placenta/aborted material' and 'vaginal secretion'.
4. The exit material 'placenta/aborted material' leads to infection of individuals of the same host species via 'alimentary tract' and '*in utero*/congenital'.

5. True ovarian transmission has been viewed as evidence for '*in utero*/congenital' transmission.
6. The exit materials 'milk' and 'eggs' will contaminate the mechanical vector 'animal products'.
7. The exit material 'milk' will cause infection of the offspring via 'alimentary tract'.
8. If a host species has been reported to receive infection by drinking from contaminated water courses like rivers or lakes, then it will also contaminate these water courses via 'faeces'.

In the second step the list of assumptions was merely extended to account for additional scenarios. Both the previous as well as the following assumptions have been validated by an expert on *Salmonella*:

1. Every host species gets infected with its associated serovar(s) via ingestion of 'soil'.
2. Humans contaminate 'soil' with their associated serovar(s) via 'faeces'.
3. Every host species (except insects and arthropods) contaminates 'rivers and streams' and 'lakes and ponds' via 'faeces'.
4. Livestock hosts and humans contaminate 'man made' structures with their associated serovar(s) via 'faeces'.
5. Gulls, crows, rats, mice and flies get infected by 'man made' structures via 'alimentary tract'.
6. Gulls and crows get infected via ingestion of contaminated 'placenta/aborted material'.
7. Livestock hosts contaminate the item 'animal product' with their associated serovars via 'faeces'.
8. Songbirds and pigeons contaminate the item 'people' (e.g. birds feeders) with their associated serovars via 'faeces'.
9. 'Saliva' and 'faeces' of pet lizards and pet turtles infects humans via 'damaged skin'. Only valid for serovars which are reported both in humans and in lizards and/or turtles.
10. The exit material 'milk' contaminates the mechanical vector 'people' (e.g. milking equipment).
11. Predators get infected via ingestion of contaminated 'flesh' of the prey. Cats feed on songbirds, mice and rabbits. Birds of prey feed on mice, rats, rabbits. Foxes feed on mice, rabbits, insects, poultry.

4.1.2 Literature search

The genus *Salmonella* is generally divided into two species [16]. One species is *Salmonella enterica* which is subdivided into six subspecies comprising a total of approximately > 2500 serovars [171] [22]. Some serovars can be further differentiated via phage and/or molecular typing. More than 1400 of these serovars have been assigned to *Salmonella enterica* subspecies *enterica* [16].

All serovars of the genus *Salmonella* are regarded as a hazard to public health [56]. However, detailed data on transmission related issues are available for a comparaitively few serovars only.

Twelve serovars of *Salmonella enterica* subspecies *enterica* were selected that had to infect at least one of the previously defined seven livestock host species. These serovar are Typhimurium, Enteritidis, Dublin, Abortusovis, Agona, Brandenburg, Choleraesuis, Menston, Montevideo, Newport, Thompson and Virchow, have been included in the following analysis. The basic search strategy was organised as described in section 2.2. A total of 103 additional references including *Salmonella*-specific textbooks, review articles, case reports etc. have provided information. The final datasets to be used in the analysis consisted of 110 distinct host species, 809 transmission routes for the basic dataset and 1716 transmission routes for the extended dataset.

4.1.3 Derived features: Measures of centrality

A major interest in network analysis is the discovery of vertices that are of central importance to the network. The importance of a vertex within the network is often approached by quantifying the vertex centrality. Measuring the vertex centrality is a very active field in its own right. Many different indices for its measurement have been suggested reflecting the many aspects of network topologies [133] [211] [111].

First of all, none of the following centrality measures paid attention to the exit materials and/or entry portals. More precisely, two or more transmission routes that differed merely in the exit material and/or entry portal counted only once.

It was decided to incorporate various centrality measures based on the indegree concept. The first measure shall be called (close/direct contact) host-to-host indegree. This is the number of transmission routes within a given network which are adjacent from a host and which are adjacent to the host species in focus. In order to account for report bias in the literature this feature was calculated on the species, genus as well as family level with respect to the donor hosts.

The next feature has been termed (non-close/indirect) mechanical vector (MV) indegree and is basically an indegree of indegrees. It takes all mechanical vectors into account which are adjacent to the host species in focus. Subsequently it sums up the number of hosts (either species, genera or families) that are able to contaminate each one of these mechanical vectors. The same type of indegree has been calculated for the environments too. Figure 4.1 exemplifies the calculation for mechanical vectors indegrees using three host species A, B, and C. Host species A and C both receive a MV-indegree of five (species level) or four (family level), respectively. Host B, instead, yields a MV-indegree of eight (species level) but again only four on the family level.

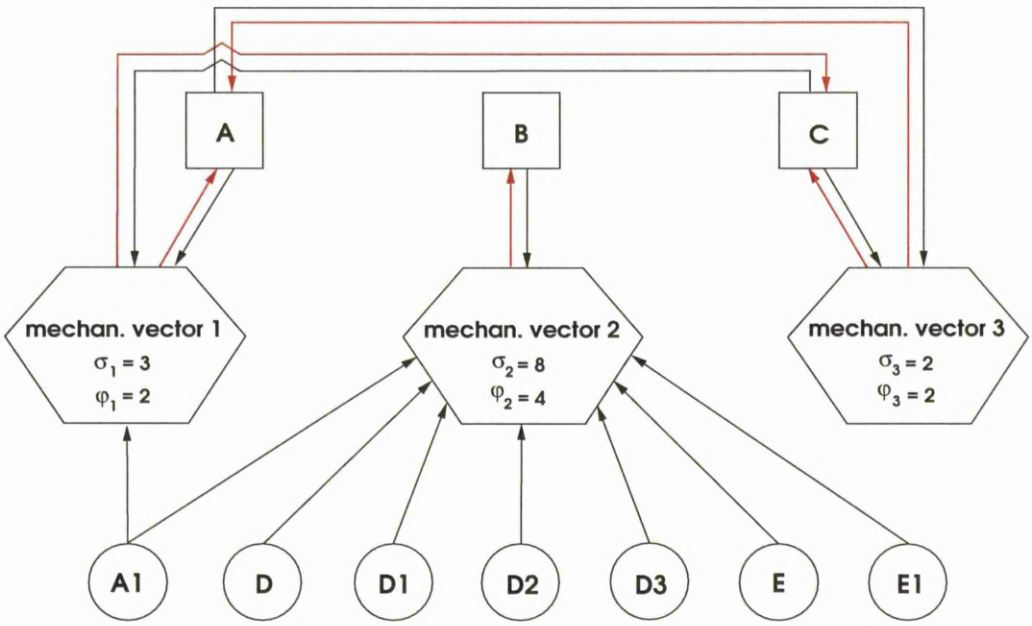

Figure 4.1: Mechanical vector indegree: The figure depicts a fictitious transmission network. A square corresponds to host species in focus and circles to any additional host species. Capital letters symbolise the host's family and the numbers denote the host species. No number refers to host species 0. Red arrows highlight transmission routes that are adjacent to a host species in focus. The σ's show the host-to-mechanical vector indegree for every single mechanical vector on a species and φ on the family level.

The final centrality feature that was included is the size of the largest *strongly connected* subgraph of a digraph (*di*rected *graph*) [87] for a host-serovar interaction. Within a strongly connected subgraph there is a directed path or walk from every vertex to every other vertex. In other words, every vertex is accessible from every other vertex. Hosts which are part of large strongly connected subgraphs might be expected to be at greater risk in terms of reinfection.

4.1.4 Analysis: PCA and *k*-means cluster algorithm

The dataset consisted of twelve objects (i.e. the host-serovar interactions) and ten centrality features and was rescaled via the Z-transformation in order to increase the comparability of the variables [37]. It centres the values of each variable around an arithmetic mean of zero with a standard deviation of one and is used for interval-scaled data, i.e. discrete numerical or continuous.

First of all, I used principal component analysis (PCA) to look at the spread of variation between the host-serovar interactions or rather their variables. Subsequently I applied the k-means algorithm

in order to discover groupings (i.e. clusters) in set of host-serovar interactions. Here, I tested the 'Squared Eucledian' distance measure and the 'Cosine' similarity measure to calculate the proximity between the objects. To account for potential local minima, the algorithm was repeated 200 times every time. An introductory tutorial on PCA and a list of useful references can be found in [166]. For an introduction to cluster analysis the reader is referred to Kaufman and Rousseeuw [90], Xu *et al.* [213], Zhao *et al.* [214] and Handl *et al.* [76].

4.1.4.1 Validation of clusterings

Cluster algorithms always generate a clustering. In fact, the number of possible clusterings for any given dataset might be large. The validation and assessment of the clustering's quality, usefulness or meaning therefore is a pivotal task in the process of cluster analysis. The literature generally distinguishes two major types of validation criteria, i.e. *internal* and *external* criteria [213] [76] [214]. The former criteria base their validation entirely on the clustering and the original data. Many of them look at, for example, the compactness of clusters (intra-cluster distances) and/or the separation of clusters (inter-cluster distances).

The second type of validation criteria make use of external or additional information on the objects. This results in a labelling of the objects which can be understood as a desired or optimal clustering itself. The idea is to validate a clustering by its agreement or overlap with the grouping as claimed by the object labels. Such an approach would require the external information to be of appropriate quality. Measuring the overlap can be achieved in various ways. The result of these measures might show significant differences depending on the data at hand. For example, the Rand-index apparently oversimplifies clusterings with relatively small clusters [76]. It has therefore been decided to test three distinct indices, namely the mutual information, the Rand-index and the Jaccard-index. Moreover, since both internal as well as external criteria suffer from intrinsic biases when evaluating the correct number of clusters [76], it has been decided to incorporate an internal criterion too.

Silhouette value:

The *silhouette* [157] value is a representative of the internal criteria. It takes both compactness and separation into account. Each object of a clustering receives a value in the range [-1,+1]. A value close to +1 means that an object is on average closer to the objects in its current cluster as compared to the objects of any other cluster. Consequently, it would be considered being assigned to the "correct" cluster. The opposite is the case for a value of -1. A clustering receiving an average value of approximately +1 is therefore characterised by compact and well separated clusters.

$$\mathrm{sil}(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

with:

- i: Object i.
- $a(i)$: Average proximity of object i with all objects of its cluster.
- $b(i)$: Average proximity of object i with all objects of next closest cluster[1].

However, compactness and separation might not be the most appropriate criteria for a particular dataset in the first place. Moreover, this type of validation provides no further interpretation of a clusterings meaning.

Mutual information and entropy:

The *mutual information* I measures the interdependence between two random variables. The unit is usually the *bit*[2] or rather bit per symbol. The random variables would correspond to the clusterings C on the one hand and to the external annotation or information A on the other hand:

$$I(C, A) = H(C) + H(A) - H(C, A) \tag{4.1}$$

H denotes the *entropy*. Consequently $H(C, A)$ is the joint entropy of C and A. Shannon [162] defined entropy is an additive measure of uncertainty[3], also referred to as the average surprisal. Given a sequence containing n distinct symbols, entropy yields a numerical value for the sequence's informational content:

$$H = -\sum_{i=1}^{n} p_i log_2 p_i \tag{4.2}$$

with:

H:	Entropy (unit: bits per symbol, i.e. logarithmic base 2).
M:	Number of distinct symbols.
p_i:	Probability to encounter the i^{th} symbol.
$-log_2 p_i$:	Surprisal [186].

Figure 4.2 illustrates the entropy function for two distinct symbols. The probability to encounter the second symbol calculates as p_2=1 $- p_1$. H reaches its maximal value - in this case one - when both symbols exhibit the equal probabilities, i.e. $p_1 = p_2 = 0.5$. However, H declines whenever the probability of any one of the two symbols predominates. In other words, p_1=0 leaves no room for uncertainty. The surprisal to encounter the second symbol would be zero whereas the surprisal to obtain the first symbol would be infinitely high.

[2]Not to be confused with its usage as in *binary digit*.

[3]The opposite of uncertainty would be referred to as *information* since a decrease of uncertainty is accompanied with an increase of information [162].

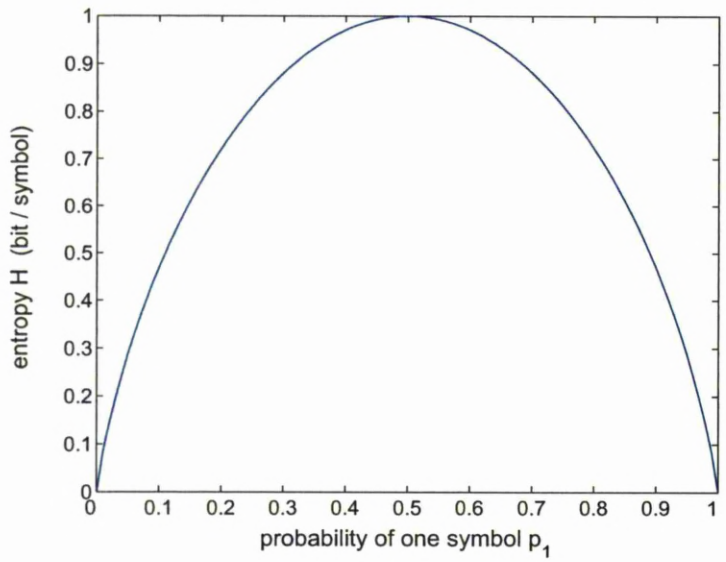

Figure 4.2: Entropy function exemplified by two distinct symbols.

Jaccard- and Rand-index:

Both indices categorise all possible object-pairs based on their cluster membership and external annotation. The indices are simple relations between the sum of object-pairs assigned to each category of interest:

$$\text{Jaccard} = \frac{\sum SS}{\sum SS + \sum SD + \sum DS}$$

$$\text{Rand} = \frac{\sum SS + \sum DD}{\sum SS + \sum SD + \sum DS + \sum DD}$$

with:

- SS: ***S***ame cluster and ***s***ame annotation.
- SD: ***S***ame cluster but ***d***ifferent annotation.
- DS: ***D***ifferent cluster but ***s***ame annotation.
- DD: ***D***ifferent cluster and ***d***ifferent annotation.

The maximal value to be reached is obviously one. This is the case, iff the clustering perfectly agrees with the external annotation. Note that the Rand-index also pays attention to object-pairs which differ in both the cluster membership and annotation.

Null hypothesis and statistical significance:

Let the null hypothesis H_0 be:

> **H_0:** The overlap-index of a particular clustering is not better than the overlap-index after randomising this clustering.

In order to generate a randomised basic population of clusterings, a shuffle procedure was applied. The shuffle step kept the cluster-sizes of the clustering in focus constant and randomly swapped the objects along with their annotations. In other words, objects are randomly assigned to any of the generated clusters (see table 4.1).

cluster	1	1	1	2	2	2	3	3	⇒	1	1	1	2	2	2	3	3
object	1	2	3	4	5	6	7	8		3	4	6	8	1	5	2	7
annotation	X	X	X	Y	Y	Y	Z	Z		X	Y	Y	Z	X	Y	X	Z

Table 4.1: Randomisation of a clustering - the shuffle procedure.

The significance of a clustering is then calculated as:

$$\text{significance} = \frac{\text{clustering} - \overline{\text{shuffle}}}{\text{shuffle}_\sigma}$$

with:

significance:	The higher the more likely to reject H_0.
clustering:	Index of a particular clustering.
$\overline{\text{shuffle}}$:	Arithmetic mean of index based on 50,000 random shuffles.
shuffle_σ:	Standard deviation of index based on 50,000 random shuffles.

4.2 Results

4.2.1 Summary statistics: Host range, entry portals and exit materials

In order to provide a general overview of the twelve serovars, they are now compared with respect to their host range, their entry portals and also exit materials. Table 4.2 reveals large variation between the serovars in terms of host ranges.

	# mammalian spp.	# avian spp.	# reptilian spp.	# insect spp.	# amphibian spp.	# gastropod spp.	total
Typhimurium	21	47	5	5	0	0	78
Enteritidis	12	5	3	2	1	1	24
Newport	7	2	10	2	0	0	21
Dublin	12	6	0	0	0	0	18
Montevideo	7	5	6	0	0	0	18
Virchow	10	4	1	0	1	0	16
Thompson	9	3	2	0	1	0	15
Brandenburg	8	4	0	0	0	0	12
Agona	5	5	1	0	0	0	11
Abortusovis	5	0	0	0	0	0	5
Choleraesuis	3	0	1	0	0	0	4
Menston	1	1	0	0	0	0	2

Table 4.2: Host range of the included *Salmonella enterica* subsp. *enterica* serovars: All serovars turned out to be zoonotic, i.e. infect humans. It is debatable whether the slug *Arion rufus* (yellow box) should be viewed as a host of the serovar Enteritidis.

Variation between the serovars can also be detected when looking at the reported entry portals and exit materials (see table 4.3). Note that the figures are only presenting a rough overview and that the entry portals and exit material are not used by every reported host. In fact, entry portals and exit materials might even exhibit varying specificity. For example, an entry portals could be 'host class'-specific etc., i.e. be reported for hosts belonging to different host orders of exactly one host class.

		Typhimurium	Dublin	Enteritidis	Abortusovis	Montevideo	Newport	Brandenburg	Virchow	Thompson	Menston	Choleraesuis	Agona
entry portals	blood stream												
	damaged skin	✓			✓		✓	✓					
	intact skin	✓	✓			✓					✓		
	eyes	✓	✓	✓									
	in utero/congenital	✓	✓	✓	✓	✓	✓			✓			
	alimentary tract	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	respiratory tract	✓	✓	✓	✓				✓				
	mammary tract												
	rectum												
	urogenital tract			✓									
exit materials	blood												
	eggs	✓		✓						✓	✓		
	faeces	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	flesh	✓	✓	✓		✓	✓			✓			
	milk	✓	✓	✓	✓		✓		✓	✓			
	ocular secretion												
	placenta/aborted material	✓	✓	✓	✓	✓		✓	✓				
	pus												
	respiratory excretions				✓								
	saliva	✓		✓		✓	✓						
	semen				✓								
	shed skin												
	urine		✓										
	vaginal secretion		✓		✓								

Table 4.3: Entry portals and exit materials of *Salmonella enterica* subsp. *enterica* serovars: Ticked boxes = reported for at least one host species.

4.2.2 External annotation

The external annotation applied in the following analysis was derived from data provided by the *Veterinary Laboratories Agency* (VLA), the *Health Protection Agency* (HPA) and the *Health Protection Scotland* (HPS). The VLA holds records on incidences and isolations for certain livestock hosts whereas both the HPA and HPS are concerned with isolation data of the human host only. The VLA defined the terms 'isolation' and 'incidence' as follows:

- "An **isolation** is defined as the report of the first isolate[4] of a given *Salmonella* (defined by serovar, and/or phage type, if available) from the same group of animals on the given occasion. If two submissions from the same group of animals on different dates give the same serovar, this is reported as two isolations ..." [189].
- "An **incidence** comprises the first isolation and all subsequent isolations of the same serovar or serovar and phage/definite type combination of a particular *Salmonella* from an animal, group of animals or their environment on a single premises, within a defined time period (usually 30 days) ..." [189].

Apparently incidents "... afford a truer picture of the amount of *Salmonella* in the animal population as they do not include repeat isolations of a serovar that may result from a number of samplings during the course of an investigation, or monitoring activities on a particular premises" [189]. From this point of view this chapter should focus on the incidence rather than isolation data. Incidence data is not available for interactions with the human host. Calculating a Pearson's linear correlation coefficient between the entirety of the VLA's incidence and isolation data over the years from 1999 to 2008 yields $\rho = 0.99$ (p-value$=6 * 10^{-68}$). This justifies omission of the incidence data in favour of the isolation data. It also casts doubt on the notion of a truer picture provided by incidence data.

Figure 4.3 shows the average annual number of isolations with respect to the included 60 host-serovar interactions. Note that the arithmetic mean is very similar to the median for all host-pathogen interactions, suggesting a dataset containing few outliers, i.e. the annual number of isolations over the period of ten years is relatively stable. Three reemergence categories were distinguished based on figure 4.3:

1. Low reemergence-frequency: on average (arithmetic mean) < 20 isolations/year,
2. Medium reemergence-frequency: on average (arithmetic mean) $[20, 49]$ isolations/year,
3. High reemergence-frequency: on average (arithmetic mean) ≥ 50 isolations/year.

In order to get an idea of the research effort spent on each host-serovar interaction, i.e. potential literature bias, NCBI's PubMed database was searched for publication numbers. The search included various possible names for the host species[5] and was done using five distinct search criteria:

[4] "An isolate is a single culture of a particular *Salmonella*, and results from a single sample ..." [189].

[5] For example, *Bos taurus*, cattle, cow, calve, calf etc.

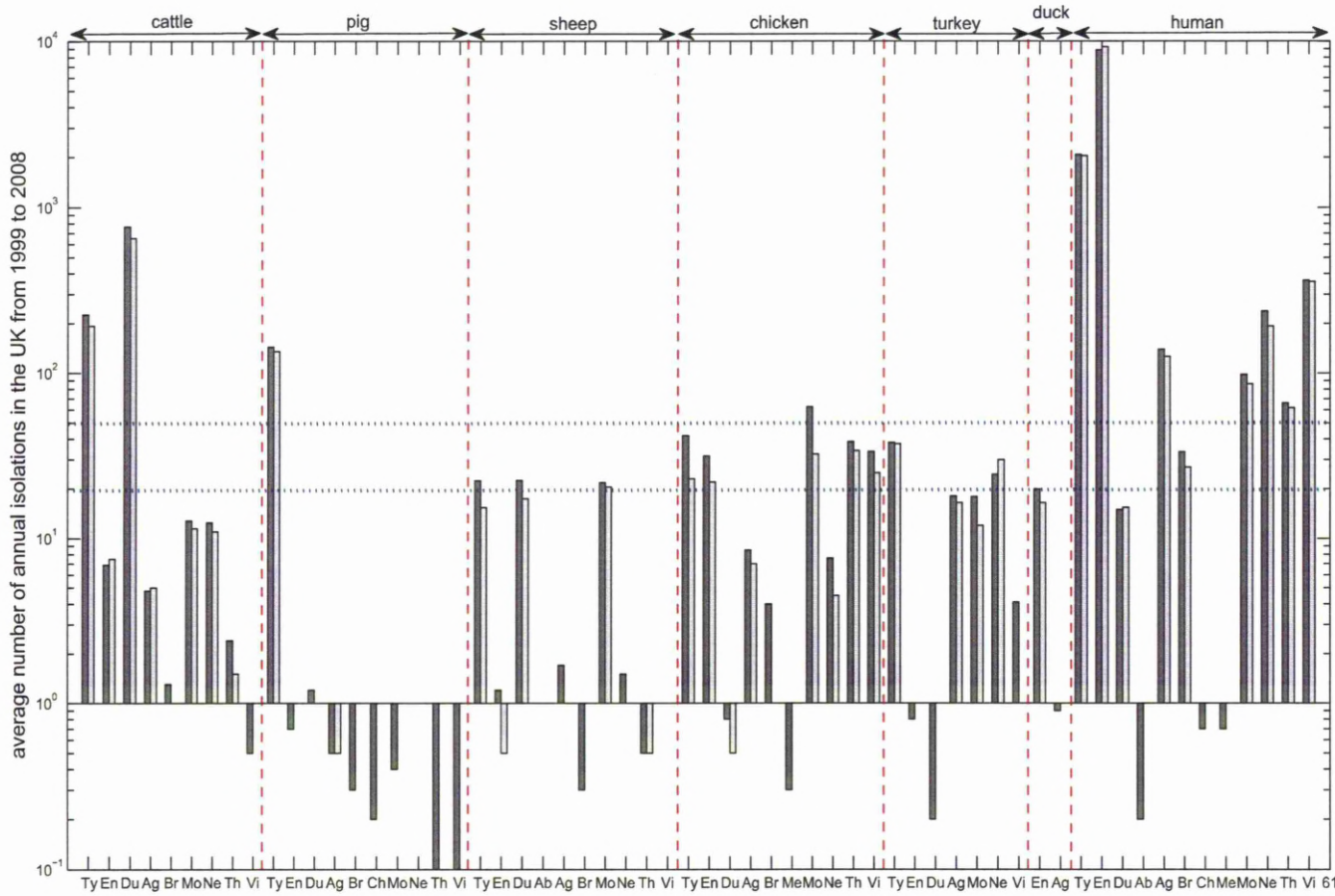

Figure 4.3: Average annual number of isolations for the twelve serovars: The plot depicts the data provided by the VLA for livestock hosts as well as the combined data of the HPA and HPS for the human host. The blue, horizontal, dotted lines refer to the applied thresholds in order to categorise the isolation data into groups of reemergence-frequency. The lower of these lines represents an average of 20 isolations/year whereas the top line was set at an average of 50 isolations/year. The serovars are **Ty**=Typhimurium, **En**=Enteritidis, **Du**=Dublin, **Ab**=Abortusovis, **Ag**=Agona, **Br**=Brandenburg, **Ch**=Choleraesuis, **Me**=Menston, **Mo**=Montevideo, **Ne**=Newport, **Th**=Thompson and **Vi**=Virchow. Note the logarithmic scale applied to the y-axis.

1. Search "All fields" for the host *AND* the serovar.
2. Search "Title/Abstract" for the host *AND* the serovar.
3. Search "Title/Abstract" for the host *AND* the serovar *AND* the term 'infection'[6].
4. Search "Title/Abstract" for the host *AND* the serovar *AND* the term 'transmission'.
5. Search "Title/Abstract" for the host *AND* the serovar *AND* both of the above terms.

The results of this search as compared to the average annual number of isolations is shown in table 4.4. It can be seen that the correlation between the isolation data and the publication numbers increases from left to right. The p-values also suggest a significant correlation between the number of isolations and the number of publications.

	Isolation vs. PubMed 1	Isolation vs. PubMed 2	Isolation vs. PubMed 3	Isolation vs. PubMed 4	Isolation vs. PubMed 5
ρ	0.27	0.42	0.46	0.53	0.60
p-value	0.04	$0.009 * 10^{-1}$	$0.002 * 10^{-1}$	$0.002 * 10^{-2}$	$0.005 * 10^{-4}$

Table 4.4: Isolation data compared to number of publications: PubMed 1-5 refers to the five applied search criteria (see text). ρ is the Pearson's linear correlation coefficient. Note that a p-value of < 0.05 can be regarded as ρ being significantly different from zero.

4.2.3 PCA

Though this particular dataset comprises only ten variables, it still seems reasonable to apply principal component analysis in order to get a first impression on the internal data-structure. Table 4.5 compares the result of a principal component analysis for the basic version with the result for the extended version (extended list of assumptions) of the dataset. The first two principal components account for 92% or 93%, respectively, of the total variance in the datasets. The loadings of the indegree measures point to strong correlations between related types of indegree measures. The only major difference between the loadings of both datasets can be observed for the strongly connected subgraph feature. This leads us to the interpretation of the PC's meaning. The loadings are nothing else but the coefficients of the linear combination that forms a PC. Given

[6]Note that PubMed currently does not support the use of wildcards, e.g. the asterisk 'infect*'. Searching an alternative database which provides such a feature appears to be useful.

the size of the loadings it becomes apparent that the first PC of the basic dataset is formed by the impact of the (non-close/indirect contact) transmission from environments. The second PC of the basic dataset, however, combines close/direct contact transmission with a negative influence of the strongly connected feature. Since negative subgraph sizes are not possible, it would mean that any host-serovar interaction performing well on this PC would be closely connected to comparatively many hosts but these hosts are not part of a large connected transmission subnetwork, i.e. the chain of transmission will be cut off between most of the hosts, unless they are part of a relatively small close/direct host-to-host contact network. Interestingly, looking at the third PC of the basic dataset it can be seen that the impact of mechanical vectors might be accompanied by larger strongly connected subgraphs.

	basic dataset			extended dataset		
	PC 1 (72%)	PC 2 (20%)	PC 3 (6%)	PC 1 (71%)	PC 2 (22%)	PC 3 (5%)
host-to-host indegree (species)	0.03	0.46	0.11	0.00	0.58	0.00
host-to-host indegree (genus)	0.03	0.46	0.11	0.00	0.58	0.00
host-to-host indegree (family)	0.04	0.45	0.12	0.00	0.57	0.01
environment-to-host indegree (species)	0.57	0.03	-0.03	0.51	0.00	0.00
environment-to-host indegree (genus)	0.57	0.02	-0.02	0.51	0.00	0.00
environment-to-host indegree (family)	0.56	0.02	-0.01	0.52	0.01	-0.02
MV-to-host indegree (species)	0.00	0.11	0.46	0.01	-0.05	0.60
MV-to-host indegree (genus)	0.00	0.09	0.47	0.01	-0.01	0.58
MV-to-host indegree (family)	-0.09	0.12	0.50	-0.02	0.06	0.55
size of strongly connected subgraph	0.13	-0.57	0.53	0.40	-0.01	0.02

Table 4.5: Loadings of the principal components: Those loadings which determine the meaning of a PC are highlighted by a yellow background. The percentages in parentheses state the explained variance. Shown are the loadings after rotation of the PC's in order to improve the loadings' pattern, i.e. their interpretability (see 'rotatefactors' function, MATLAB).

Looking at figure 4.4 most host-serovar interactions can be detected within a small area below the average (see red, dashed lines) of both PC's. Many human-serovar interactions, though performing below average with respect to the first PC (i.e. environmental impact), perform better on the second PC (i.e. impact of close/direct contact). The opposite is the case for interactions with livestock hosts, in particular for grazing hosts such as cattle and sheep. Moreover, certain interactions of the serovar Typhimurium are remarkably separated from the rest of the interactions. Typhimurium seems to affect the human host with respect to both environmental factors as well as close/direct contact with infected hosts.

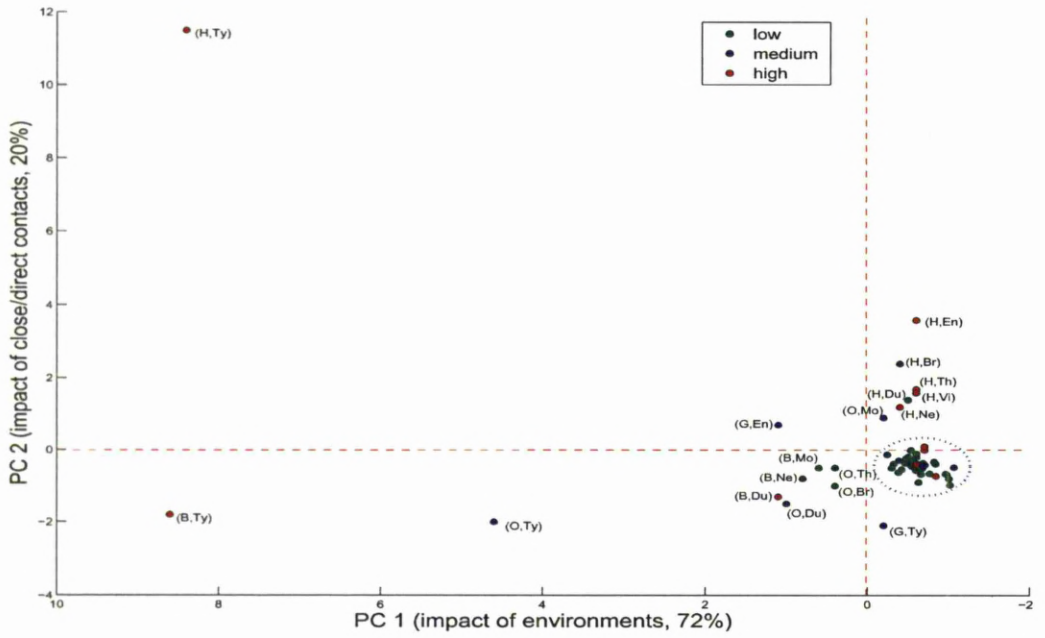

Figure 4.4: PCA of the basic dataset (first and second PC): Note that the negative values on the abscissa have been assigned to the right of the zero value. The legend refers to the reemergence-frequency categories. The host-serovar interactions are stated in parentheses. The first capital letter stands for the host, i.e. **B**=*Bos taurus*, **S**=*Sus scrofa domestica*, **O**=*Ovis aries*, **G**=*Gallus gallus domesticus*, **M**=*Meleagris gallopavo*, **H**=*Homo sapiens*. The following two letters indicate the serovar, i.e. **Ty**=Typhimurium, **En**=Enteritidis, **Du**=Dublin, **Ag**=Agona, **Br**=Brandenburg, **Mo**=Montevideo, **Ne**=Newport, **Th**=Thompson and **Vi**=Virchow. Most interactions are located within the blue, dotted circle/ellipse. Random noise has been added in to this area to increase the visibility. The red interactions in this area refer to (G,Mo), (S,Ty), (H,Mo) and (H,Ag). The blue interactions are (G,Vi), (G,Th), (M,Ty) and (M,Ne).

The right block of table 4.5 shows the result for the extended dataset. It appears as if the additional assumptions have caused a slightly new network topology. The first PC is now defined by the environmental impact and strongly connected subnetworks whereas the second PC merely by a contribution of close/direct contact. The pattern of figure 4.5 does not significantly change from the trends already observed in figure 4.4. However, the additional assumption made, appear to have improved the pattern. Half of the host-serovar interactions with either a high or medium reemergence-frequency moved out of the dotted, blue area whereas some of those with a low reemergence frequency have moved into this area.

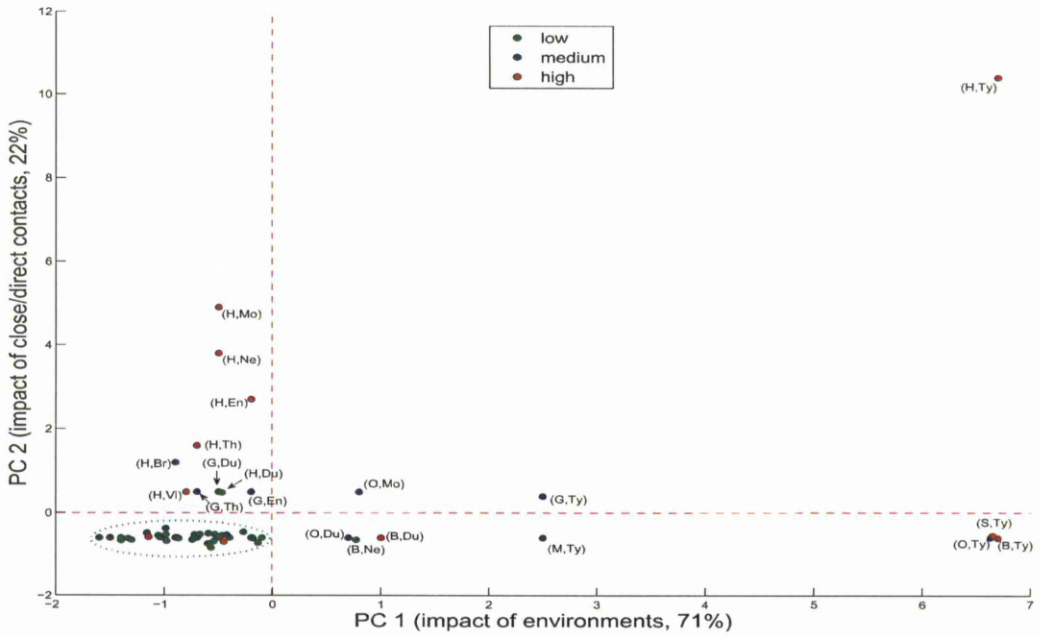

Figure 4.5: PCA of the extended dataset (first and second PC): The legend refers to the reemergence-frequency categories. The host-serovar interactions are stated in parentheses. The first capital letter stands for the host, i.e. **B**=*Bos taurus*, **O**=*Ovis aries*, **G**=*Gallus gallus domesticus*, **M**=*Meleagris gallopavo*, **H**=*Homo sapiens*. The following two letters indicate the serovar, i.e. **Ty**=Typhimurium, **En**=Enteritidis, **Ag**=Agona, **Du**=Dublin, **Br**=Brandenburg, **Mo**=Montevideo, **Ne**=Newport, **Th**=Thompson and **Vi**=Virchow. Most interactions are located within the blue, dotted ellipse. Random noise has been added in to this area to increase the visibility. The red interactions in this area refer to (G,Mo) and (H,Ag). The blue interactions are (G,Vi) and (M,Ne).

Figure 4.6 plots the first principal component of the extended dataset against the third. Though the third component accounts for only 5% of the variance , it can be seen that a few livestock-serovar interactions have been particularly linked to influence of mechanical vectors. Many interactions with humans "cluster" in the lower part of the top left square as indicated by the red dashed lines.

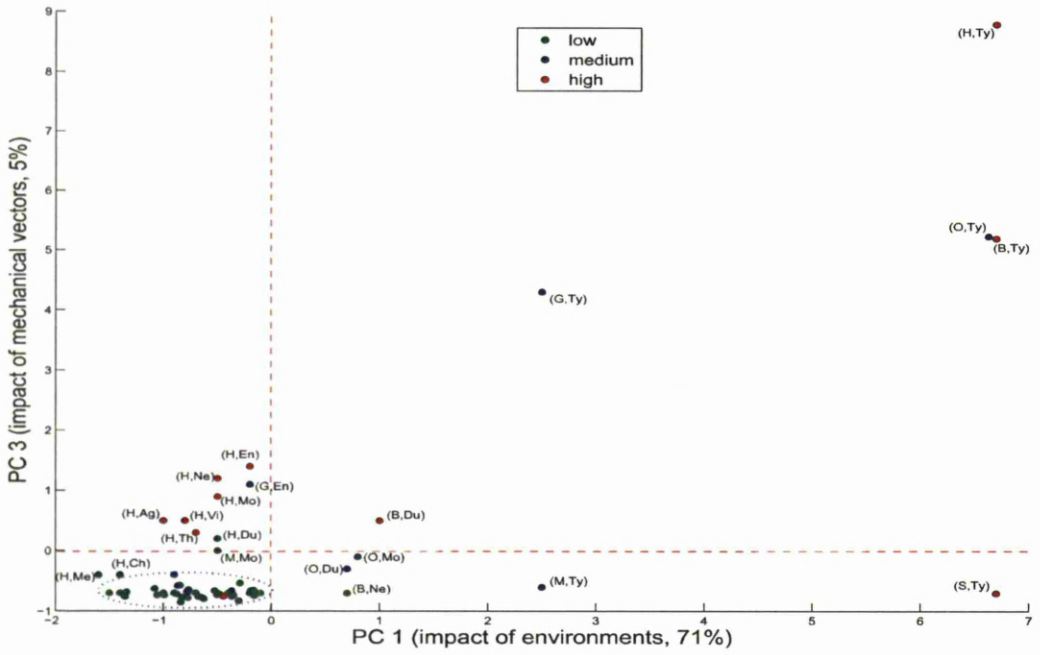

Figure 4.6: PCA of the extended dataset (first and third PC): The legend refers to the reemergence-frequency categories. The host-serovar interactions are stated in parentheses. The first capital letter stands for the host, i.e. **B**=*Bos taurus*, **S**=*Sus scrofa domestica*, **O**=*Ovis aries*, **G**=*Gallus gallus domesticus*, **M**=*Meleagris gallopavo*, **H**=*Homo sapiens*. The following two letters indicate the serovar, i.e. **Ty**=Typhimurium, **En**=Enteritidis, **Du**=Dublin, **Ag**=Agona, **Br**=Brandenburg, **Ch**=Choleraesuis, **Me**=Menston, **Mo**=Montevideo, **Ne**=Newport, **Th**=Thompson and **Vi**=Virchow. Most interactions are located within the blue, dotted, ellipse. Random noise has been added in to this area to increase the visibility. The red interaction in this area refers to (G,Mo). The blue interactions are (G,Vi), (G,Th), (M,Ne) and (H,Br).

4.2.4 *k*-means

Given the plots of the principal component analysis, the potential cluster number could be expected to be relatively low. Nevertheless, since the number cannot be inferred from these plots with absolute certainty, it was decided to observe a cluster-range from two to twenty clusters. Figure 4.7 shows the results of the clustering's significance with respect to the applied reemergence labels for both versions of the dataset. It can be seen that:

1. The related Jaccard- and Rand-indices exhibit very similar results. The mutual information, however, disagrees with Rand- and Jaccard with respect to a cosine-clustering of the basic dataset at seven clusters.
2. The number of clusters of the significant clusterings deviate between the basic and the extended version of the dataset.
3. The number of clusters of the significant clusterings are also seem to be dependent on the applied proximity measures.
4. The number of clusters is suggested to be relatively low ranging from two to maximally eight depending on the proximity measure and dataset.

The calculations for figure 4.7 have been repeated leading to figure 4.8. Comparing these two figures reveals inconsistencies in the bottom left subfigure which is based on the 'squared Euclidean' distance measure. Using this proximity measure the algorithms seems to be indecisive about the correct clustering. Accounting for empty clusters and drastically increasing the number of reruns as well as the random basic population did not lead to a stabilisation of this subfigure.

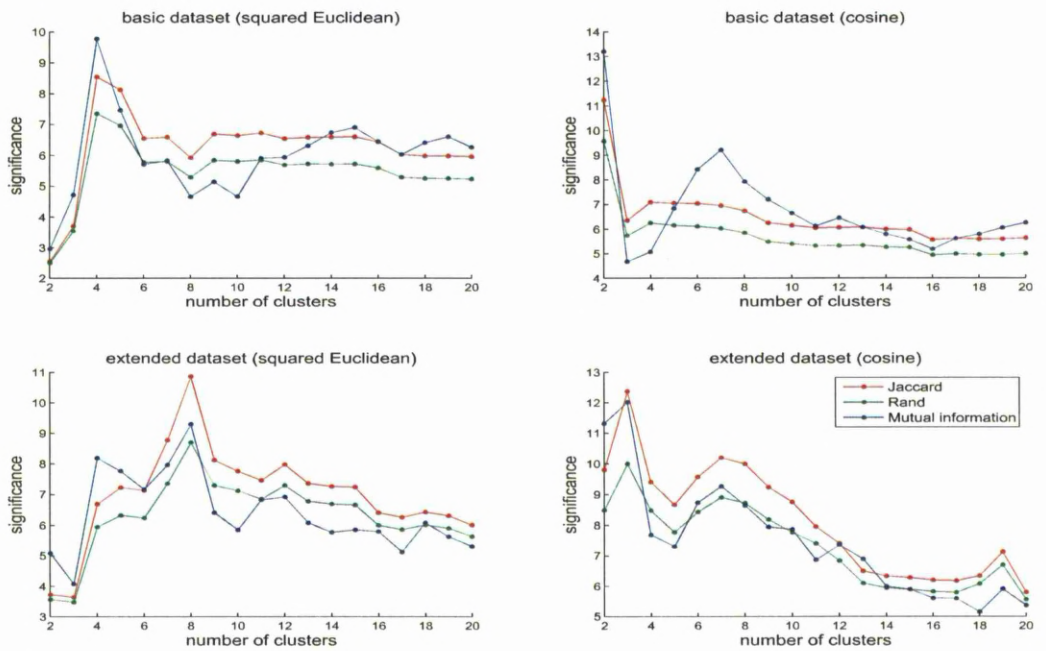

Figure 4.7: External criterion: k-means (settings: 200 replicates). Overlap between clusterings and external annotation compared with a random basic population of 50,000 individual shuffles.

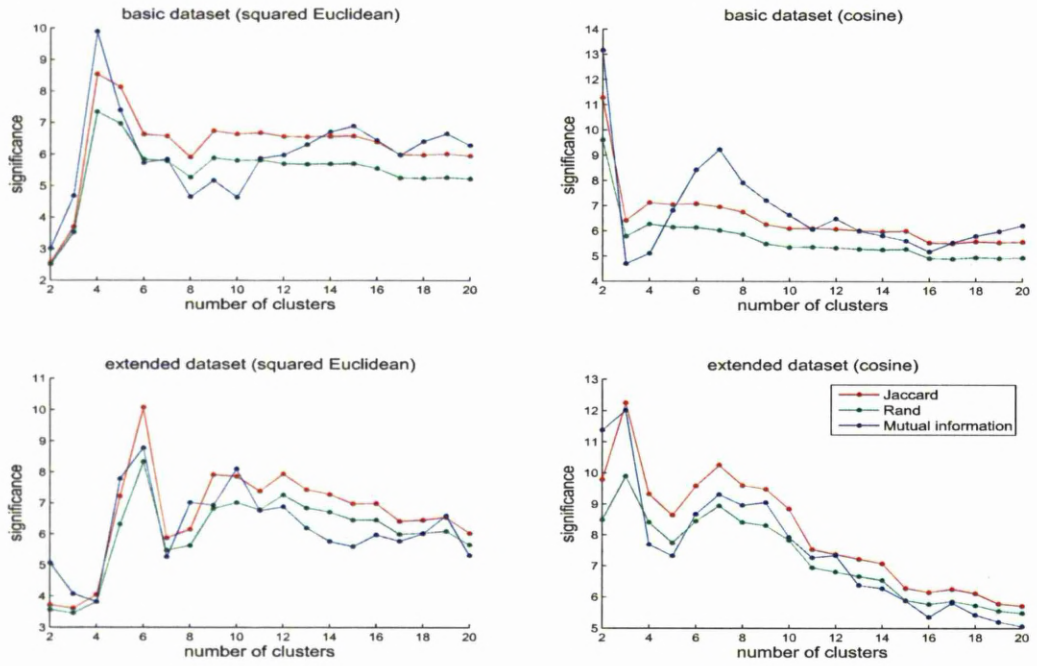

Figure 4.8: External criterion: k-means (settings: 200 replicates). Overlap between clusterings and external annotation compared with a random basic population of 50,000 individual shuffles.

Figure 4.9 depicts the average silhouette value for a range of cluster numbers in order to evaluate the clusterings additionally with an internal criterion. The silhouette measure also suggests a rather small number of clusters, i.e. three clusters. Repeated calculations using the 'squared Euclidean' distance measure again show strong fluctuations at low cluster numbers whereas the results based on the cosine similarity remain stable. This suggests that the cosine similarity might be better suited for this dataset. Therefore, the cosine similarity will be the measurement of choice. Note that the average silhouette value will approach +1 for higher cluster numbers, simply as a result of the increasing number of clusters containing just one object. For the same reason the significance in figures 4.7 and 4.8 will decrease with increasing cluster numbers.

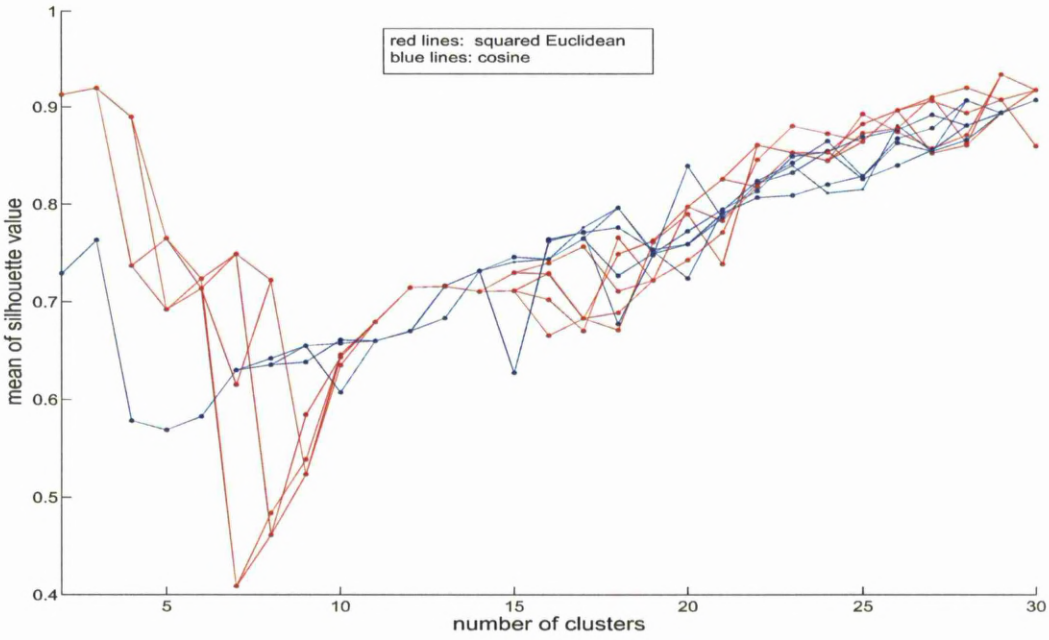

Figure 4.9: Internal criteriion, mean silhouette value: Red denotes the 'squared Euclidean' distance blue the cosine similarity. k-means (settings: squared Euclidean or cosine, 200 replicates). The proximity measure of the silhouette value was fixed in agreement with the k-means algorithm.

So far nothing has been said about the usefulness or meaning of a significant clustering. The focus will be on the extended dataset since it can be viewed as more comprehensive and showed improved results with respect to the PCA plot. The null hypothesis is rejected for the k-means clustering with the cosine measure at k=3. Figures 4.10, 4.11 and 4.10 look at the contents of the corresponding clusters. All three figures show that cluster three is dominating the clustering. Most of the host-serovar interactions assigned to that cluster are of low reemergence-frequency. The cluster contains balanced proportions of host species with one exception, it contains fewer human-serovar interactions. The same is true for the serovars. The only serovar that does not appear in cluster three is Typhimurium.

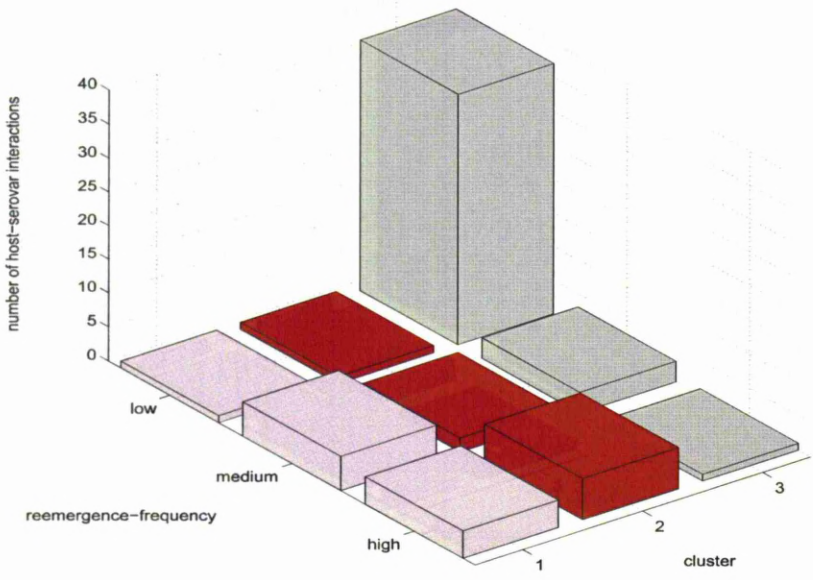

Figure 4.10: Characteristics of clusters by reemergence categories and number of host-serovar interactions: k-means (settings: cosine similarity, k=3, 200 replicates).

All interactions with the serovar Typhimurium have been assigned to cluster one, a cluster that is mainly consisting of medium and highly reemerging interactions. It contains the interactions of Typhimurium with humans, sheep, cattle, chickens, pigs and turkeys. Furthermore it includes cattle-Dublin, sheep-Dublin, cattle-Newport and sheep-Montevideo.

Cluster two is clearly characterised by the human host, containing most of the highly reemerging human interactions - namely with Agona, Brandenburg, Dublin, Enteritidis, Montevideo, Newport, Thompson and Virchow - and only one additional non-human interaction which is chicken-Enteritidis.

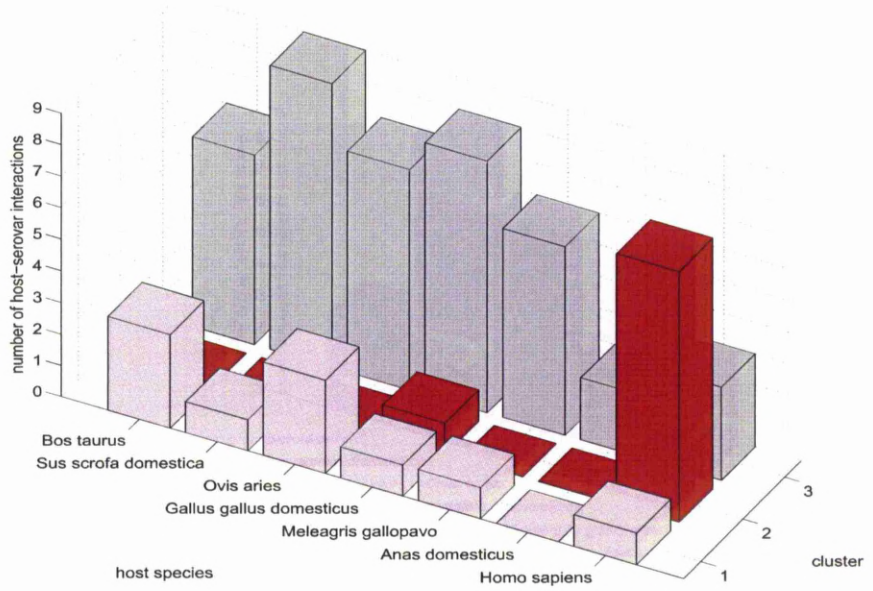

Figure 4.11: Characteristics of clusters by hosts species and number of host-serovar interactions: k-means (settings: cosine similarity, k=3, 200 replicates).

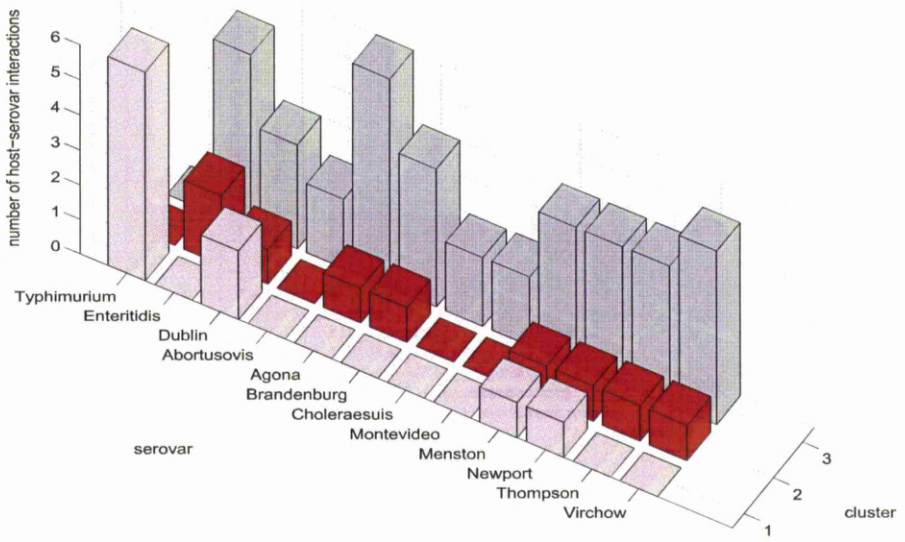

Figure 4.12: Characteristics of clusters by serovars and number of host-serovar interactions: k-means (settings: cosine similarity, k=3, 200 replicates).

4.3 Discussion

It appears useful to start by recapitulating the results of this chapter:

- The additional assumptions influenced the results, i.e. they affected the network topologies.
- The PC-plots revealed the existence of outliers.
- Livestock-serovar interactions separated in the PC-plots along the environmental component and partly along the mechanical vector component as well.
- Human-serovar interactions separated in the PC-plots along the close/direct contact component and were also influenced by the impact of mechanical vectors.
- According to the PC loadings of the extended dataset, the environments appear to be the connecting factor for a particular kind of transmission networks, i.e. those that are strongly connected.
- The cosine similarity appeared to be better suited for the extended dataset.
- The clustering (cosine, k=3) is dominated by cluster three which contains the majority of host-serovar interactions, nearly all of which are considered to be of low reemergence-frequency. Though it contains not a single interaction with Typhimurium, it shows relatively balanced proportions for the host species and serovars.
- Clusters one and two are both comparatively small. The second cluster is nearly exclusively characterised by interactions with the human host, most of which have been assigned to the high reemergence category. Cluster one contains interactions with most host species which are mostly of medium and high reemergence-frequency. Its main characteristic is that it contains, amongst a few other interactions (e.g. with Dublin), the entirety of Typhimurium interactions.
- The strong correlation between the number of publications and the number of annual isolations for each host-serovar interaction could suggest a research and surveillance bias.

The k-means algorithm is known for its sensitivity to outliers and noise [213]. Outliers will be forced into a cluster, thereby distorting the cluster's shape. The question is how to account for these outliers since they are not based on technical or physical measurements but rather on qualitative literature statements. Simply accounting for the number of publications on a given host-serovar interaction neglects the detailed information that was required for transmission routes. A single reliable report could have led to the inclusion of a transmission route and shaped the network topology significantly. A potential way to overcome this problem would be the application of an algorithm that is better suited for dealing with outliers. Such cluster algorithms have been proposed, for example, in [58] and [5].

It is difficult to make a definite decision on the number and thresholds of reemergence-frequency categories merely by looking at the isolation data. Certain host-serovar interaction lie very close to any of the thresholds. However, given the strong over- as well as underrepresentation of many of the annual number of isolations - which is likely to be due to bias itself - it was deemed justified to test an admittedly subjective categorisation. In the future it would be interesting to test the approach's

sensitivity towards various threshold settings. The isolation data might also be categorised by more sophisticated means, for example, using k-means itself to group the data.

Another disadvantage of the k-means algorithm is its behaviour towards certain data structures. For example, it struggles with closely neighbouring, elongated clusters such as ellipses or hyperellipsoids (cigar-shaped structures in higher dimensions). It would be interesting to compare k-means to other cluster algorithms such as agglomerative hierarchical approaches based on the single linkage algorithm [76] and also more sophisticated algorithms such as ensemble approaches etc. [180] [65] [89]. In fact, even the k-means algorithm has more to offer, for example, alternative proximity measures and a variety of algorithmic improvements [213]. The potential impact of the proximity measure was demonstrated too. The behaviour of the 'squared Euclidean' distance measure might be explained by a potentially higher sensitivity towards both outliers and insuffciently separated clusters, at least with respect to the dataset at hand.

On a more positive note, the k-means algorithm also offers a major advantage. It is possible to assign new cases or objects to existing clusters. Since just 60 out of 224 generated host-serovar interactions have been used - based on the availability of external information - it now would be feasible to assign all the remaining host-serovar interactions to any of the detected clusters. This way one might get a first impression on their potential reemergence-frequency. Moreover, the VLA provides information on incidences and isolations for a range of additional hosts such as geese, guinea fowl, pheasants, partridges, pigeons and quail [189]. The problem is that the data is presented in a format that does not allow the assignment of the incidence and isolation to a particular host species, i.e. the figures represent groups of host species.

One also needs to think about expanding and adjusting the list of centrality features since the topology of the transmission networks might be described more appropriately with alternative measures of centrality. The literature, in fact, provides a multitude of possible centrality indices [133] [211] [111]. However, the main insufficiency of these networks is the lack of weights, i.e. values that rate the importance of each individual transmission route. The weights would need to account for both host and serovar characteristics with respect to a specific transmission route.

Addressing the biological meaning of the clusters, one needs to consider the principal components too. All host-serovar interactions of cluster one showed an above average "performance" with respect to the first principal component which was mostly described by environmental impact. In turn, all interactions that appear on the positive, i.e. above average, side of the first principal component belong to cluster one. The suggested environmental impact or the risk associated with contaminated environments might be linked to the survival of serovars and the maintenance of infectivity outside a host. Typhimurium has been reported to be sustainable in soil and on pastures for long periods of time, e.g. 251 days in garden soil [201] or 28 weeks on pastures [188]. Human infections with Typhimurium were suggested to have originated from gardens contaminated by hedgehogs [75] or even via the ingestion of contaminated snow, especially under bird feeders [170]. Given that 47 distinct avian species, including many garden birds, were detected as hosts for Typhimurium (see table 4.2), it makes the risk for human Typhimurium infection through contaminated garden environments quite realistic, especially for children playing in the garden. In fact, Typhimurium can even survive in air sufficiently long to pose a significant airborne risk [194] [152]. Dublin was able to survive 13 to 24 weeks on pastures and constituted a risk for grazing animals [114]. For chickens, however, it was shown that the greatest contribution to infection came from close/direct contact with other chickens and not their environment[159]. Typhimurium might form an exception because of its potentially longer sustainability outside a host. Therefore, environments could act both as a

reservoir and as a distributor of these serovars. Moreover, they appear to be the crucial factor of the strongly connected subnetworks which bear the potential to disseminate an infection from any of its associated host species to every other host species that is part of this subnetwork.

The interactions of cluster two separated in the PC-plots either through the impact of close/direct contact (PC 2) and/or via contribution of mechanical vectors (PC 3). Cluster two consists almost entirely of human interactions, the interaction of chickens with Enteritidis being the sole exception. A biological or rather sociological interpretation of this cluster could involve, apart from research bias on the human host of course, human activities as compared to non-human animals. Humans can be expected to come into close contact with many distinct host species as a result of their various roles in society, for example, as farmers [21], animal caretakers, veterinary surgeons [191] [202], pet owners [149] [192] [198] etc. This might increase the risk for humans to receive a *Salmonella* infection via direct contact considerably. In fact, human *Salmonella* infections are increasingly linked to close/direct contact transmission from various common as well as exotic pet animals such as avian species, dogs, cats, iguanas, turtles, snakes and hedgehogs [94] [192] [198].

Serovars of *Salmonella enterica* subspecies *enterica* have been recognised as the second most frequent cause of foodborne outbreaks in the human population within the European Union [57]. This could partly explain the fact that most human interactions showed positive values on the third PC and that the majority of these, though Typhimurium assigned to cluster one, were members of cluster two. Relatively strong evidence for human foodborne infections with Typhimurium [210] [8] [185] [149] [73] [106] [198], Enteritidis [130] [30] [210] [159] [103] [107], Virchow [14][121], Thompson [21] [107], Dublin [202], Agona [14] and to a lesser degree Montevideo [107] could be found in the literature. However, apparently no human cases of foodborne infections with Abortusovis were reported according to [114]. No clear evidence was found for Newport as well as Menston and only implications were stated for Choleraesuis [39] and Brandenburg [26].

Foodborne infections were also identified for non-human animals. Enteritidis, for example, has been shown to survive for at least 26 weeks in artificially contaminated poultry food [40]. Paratyphoid *Salmonella* were in fact able to survive for two years in artificially contaminated feeds [159]. Typhimurium was isolated from cattle feed which can be heavily contaminated and cause outbreaks [118] [201] [30], whereas Dublin is said to be a rare source of foodborne infection of cattle [202] which is in agreement with its position in figure 4.6. Food was also reported as a risk factor for outbreaks of Typhimurium, Dublin, Montevideo and Brandenburg in sheep [110]. However, figure 4.6 supports this idea only for the sheep-Typhimurium interaction.

Note that the references cited to support the findings could be part of the existing reasearch bias. For example, the survival times for most serovars simply have not been investigated yet. Nevertheless, the clustering suggests the existence of three distinct groups of host-serovar interactions:

1. A large group consisting of host vertices with low centrality values in their associated transmission networks. The corresponding host-serovar interactions are predominantly of low reemergence-frequency. Characteristically, Typhimurium is not part of this group.
2. A smaller group containing mainly human interactions of high reemergence-frequency. The interactions of this group are characterised both by intensive inter-species close/direct contact structures as well as an inclination towards infections from mechanical vectors which likely involve food or feeds.

3. Another small group which includes mainly interactions with livestock hosts, the interaction of humans with Typhimurium being the only exception. The interactions of this group are characterised by infections stemming from mechanical vectors (food or feeds likely) and/or the contact with contaminated environments.

As was pointed out earlier are the thresholds defining the reemergence categories based on subjective judgements. It is even more surprising to find certain host-serovar interactions within a given cluster. For example, the sheep-Typhimurium interaction was assigned to the medium reemergence-frequency category. Its corresponding median value, however, lies below the threshold of the medium-frequency category (see figure 4.3). In other words, the average number of annual isolations does not support the idea that this interaction is to be expected to reemerge particularly frequently. The cluster assignment and the PC-plots, instead, suggest a different picture. Typhimurium appears to have a much higher potential to reemerge in sheep. A similar conclusion can be drawn for the interaction of cattle with Newport and in the opposite direction for chicken with Montevideo.

As a final remark it should be mentioned that the nomenclature of the genus *Salmonella* is not standardised [22] [16]. There are at least two different systems of nomenclature in use which apparently are even often combined [183] [16]. This clearly leads to confusion and could have inflicted bias on the dataset. Note also that differences occur even in the report and surveillance effort of livestock hosts. The surveillance of chickens, for example, is apparently carried out more thoroughly as compared to other livestock hosts.

Chapter 5

Discussion

The thesis aimed at developing a framework that supports the detailed and so far unprecedented categorisation of information on transmission pathways. It paid attention to hosts, their environments, various mechanical vectors, the entry portals of pathogens into hosts and the exit materials by which a pathogen is released from a host. The framework was then used in practice to generate two datasets of transmission routes. The first dataset included bacteria and viruses and was intended to answer general questions about, for example, host ranges, the impact of close/direct and non-close/indirect contact transmissions and differences between viruses and bacteria. The second dataset was concerned with twelve serovars of the bacterium *Salmonella enterica*. The analysis of the second dataset aimed at testing the hypothesis that the centrality or connectedness of a host vertex within its network of transmission routes is linked to its potential to acquire infections. The following list focuses the key results as I see it. Note that the host range is a reflection of previously emerging host-pathogen interactions:

1. The results of the analysis of the first dataset suggested that:
   (a) Multihost, zoonotic and bacterial pathogens are more likely to make use of non-close/indirect contact transmisson.
   (b) The host range of a pathogen is mainly influenced by its ability to be transmitted via non-close/indirect transmission pathways involving environments and mechanical vectors. Alternatively a pathogen might broaden its host range via close/direct pathways if it gets access to a network of directly interacting host species[1].
   (c) Anthropogenic mechanical vectors seem to play a particularly important role in broadening the host range of a pathogen that affects livestock.
   (d) Host species existing in close proximity share an increased number of pathogens.
2. The results of the analysis of the second dataset suggested that:

[1] The suggested option to rely on transmission via arthropod hosts will not be considered here.

(a) The reemergence-frequency of a host-serovar interaction is linked to the centrality or connectedness of the host vertex within its network of transmission routes.

(b) Anthropogenic mechanical vectors - food and feeds in particular - are major drivers for the frequency of reemerging infections in both humans and livestock animals.

(c) Environments can pose a risk[2] for infections with serovars that are sustainable outside the host.

(d) Close/direct contact transmission can pose a risk of infection for host species that cultivate close connections to additional host species.

Considering these results together, it seems that the importance or impact of a given transmission pathway is influenced by both host and pathogen characteristics. For example, host species that share grazing areas are more likely to pass on infections to other host species if the pathogen is able to maintain its infectivity outside a host for a prolonged time period. This requirement for pathogens becomes less important for scenarios in which a host species maintains close/direct contacts to other host species. A similar situation might arise for a particular kind of mechanical vectors such as fresh food products which have to be distributed, sold and consumed in a short period of time. This by the way could also explain the fact that all twelve *Salmonella enterica* serovars of the second dataset turned out to affect the human host even though the serovars were initially selected to affect livestock hosts. None of the livestock hosts, however, is reported to be affected by all twelve serovars. Humans consume fresh meat, milk or eggs of all seven livestock hosts thereby increasing both the potential of a serovar/pathogen to emerge in humans for the first time, i.e. to broaden its host range, and to reemerge frequently once humans have been established as a host. This supports the idea that the emergence of host-pathogen interactions "...often results from [...] human activity favouring the amplification of risk conditions" [150]. This could also shed a light on reports of feedborne outbreaks in poultry, cattle and sheep. The difference could be that animal feedstuffs are more likely to be processed and as a result less likely to classify as fresh. Therefore, animal feedstuffs could be more likely to select for serovars/pathogens that are sustainable outside the host such as Typhimurium, Dublin or Enteritidis. Moreover, the previous statements also implicate a link from feedborne livestock infections to foodborne human infections. Human infections with the serovar Agona, for example, apparently increased after the introduction of the serovar into animal feed [35]. Interestingly the analysis emphasised the human-Agona interaction but none of the livestock-Agona interactions, which demonstrates the potential for biases in the approach.

[2]Note that the term 'risk' is both ambiguous and controversial. It has been argued that "...whoever controls the definition of risk controls the rational solution to the problem at hand. [...] Defining risk is thus an exercise in power [169].

5.1 Biases and limitations

> *"... It was an abstract model, and like any abstract model, it's not really intended to be descriptively accurate in detail, it's intended to sort of pull out some crucial features and study those. And you have to ask in the case of an abstract model, how much of the complex reality does it really capture?"*
>
> Noam Chomsky [24]

Given the nature of the framework and consequently of the resulting data it appears to be an impossible task to control for biases. The following lists simply summarises the main sources bias associated with this project. Note that the proposed categories are neither mutually exclusive nor independent. In fact, certain sources of bias could have unpredictable effects on each other:

1. **The scientific effort:** It has already been recognized elsewhere in the literature that the scientific effort spent on a host and/or pathogen is likely to be proportional to the amount and quality of associated information [143] [27] [2] [151]. Such bias has been termed *hot stuff bias*, but the heat of a scientific field might also reduce the quality of the research findings [83]. However, it has been argued that the pathogen richness of hosts is likely to increase as more individuals and populations of a host species are examined [151]. This is nicely exemplified by the "... inherent bias of humans in studying themselves in preference to other species" [27]. In fact, study [27] identified 1415 pathogen species for the human host, 616 pathogen species for livestock hosts and only 374 pathogen species for domestic carnivores[3]. The pathogen richness will not only be determined by the number of screened host individuals and populations but also by the pathogens that form the focus of a study [143], let alone the applied screening techniques. These issues will inevitable influence the host range of pathogens too. It has been found that well studied pathogens have a broader host range [143]. Another source of bias are the taxonomies of hosts and pathogens, in particular the latter one. Inaccurate knowledge of pathogen taxonomy could, for example, render truly singlehost pathogens as multihost pathogens if the distinction between significantly different pathogens is not recognised [143] [2]. This problem has been emphasised in particular for viruses [143]. The taxonomy of pathogens is in fact a very active and highly debated issue in its own right. Efforts to resolve or simply to alert to such taxonomic ambiguities have been carried out already [77] [55]. Even "... what is meant by "species" may differ from one group to another; some pathogens have complex subspecific taxonomies (e.g. *Salmonella enterica*, *Listeria monocytogenes*, [...]), making direct comparisons of different "species" potentially problematic" [204]. The literature can also be biased between certain research fields. For example, the literature on human medicine might view many pathogens as singlehost pathogens whereas the same pathogens could be labelled as multihost and/or zoonotic in the field of veterinary medicine [27].

[3] Pathogens included viruses, bacteria, protozoa, helminths and fungi [27].

2. **The transmission framework:** The framework aims at conceptualising the field of pathogen transmission. McGray "...argues that all conceptualisations are biased, both because they depend on the purposes for which they have been created, and because they are closely tied to the world view of their designers" [112] which in turn "...depends on the state of general knowledge at the time, as well as on the personal knowledge of the designer" [112]. One should therefore expect some degree of bias based on the design of the transmission framework itself. Moreover, how can we be confident that all relevant items have been indentified or that no redundant or even irrelevant items were incorporated etc.? This is clearly very much dependent on interpretation. For example, Santamaría *et al.* pointed out that "...any microorganism present in soil, either allochthonous or autochthonous will eventually end up in the water or air"[161]. This might challenge the approach's high granularity with respect to the number of distinct environments. On the other hand, the framework did not include transmissions between environments but rather transmission from environments into hosts. Details on the environmental source of an infection could constitute an important fact. The potential flexibility in the framework's design is admittedly almost endless, a situation that is a recognised cause of bias [83]. This bias also accounts for the following issue.

3. **Interpretation of terms and concepts:** The transmission framework cannot be expected to account for all terms and concepts provided by the literature. The search for information might reveal additional terms and concepts which are not directly covered by the design of the framework. Any attempt to adjust or reconcile new concepts with the framework might be largely open to interpretation, in particular if authors fail to make clear and unambiguous statements. In fact, any room left for interpretation is likely to cause bias [83]. Bias caused by the interpretation of results and contradictory evidence was also termed *cognitive dissonance bias* [158]. Moreover, different individuals could favour distinct interpretations. Since viruses and bacteria were covered by different people, these issues could have inflicted bias on the first dataset. Even the same individual - knowingly or unknowingly - could modify his/her point of view over time since the "...perception of the literature will be better informed as you become more involved with the synthesis" [43] of data.

4. **Reading-up on the field:**

   (a) **Distinct individual contribution:** Even with the framework and literature search strategy at hand, the contribution of distinct individuals is still a potential cause for bias due to varying individual knowledge, individual differences in the interpretation of data and information and individual approaches to the search strategy etc. For example, it was not explicitly agreed to search by pathogens rather than by hosts. An individual reviewer might have focused on pathogens of a particular host species in the first place, thereby potentially covering certain host species more comprehensively than others. The possible variations in conducting the actual search are in fact manifold and will not be discussed here. However, the viruses and bacteria have been reviewed by distinct individuals which could have caused differences between viruses and bacteria that are not due to biological variations but rather based on individual contributions of the reviewers. Then again, continual meetings and discussions between the reviewers should have minimised this form of bias.

   (b) **Positive result bias:** It has been recognised that positive results are more likely to be published [158] [83].

5. **Analysis of the data:** Biases caused by the applied analytical methods and the applied additional information have been outlined in the discussion of chapter 4.

6. **Interpretation of results:** The interpretation of the analysis needs to consider the design of the framework. For example, one might ask for the proportion of zoonotic pathogens in the first dataset. The resulting value is unlikely to reflect a true proportion of zoonotic pathogens with respect to pathogens at large. It would in fact only reflect the zoonotic proportion of pathogens which are able to infect the main seven livestock host species in the first place. This could make a significant difference when it comes to the interpretation of the result since humans and livestock could be expected to coexist in a close spatial contact structure.

5.2 Outlook and conclusion

This thesis dealt with a complex subject. The complexity not only was due to actual biological elements but to a plethora of loosely defined and applied terms and concepts in the literature. Therefore, it was particularly difficult to distinguish the meaningful from the meaningless, which hindered the process of making informed decisions. However, the application of the framework and the subsequent analysis of the resulting datasets revealed both highly expected trends but also interesting new insights and perspectives. It appears therefore promising to think about new applications and future extensions of this framework. For example, the data was initially stored in spreadsheet tables for convenience and afterwards uploaded (at least partly) onto an online relational database system [129]. In fact, it would be possible to grant access to a broader research community in order to increase the quantity and quality of the database. This way we should also consider other pathogenic groups such as fungi, helminths and protozoa and various host species perhaps even plants since "kingdom jumping" [72] of plant pathogens to vertebrate hosts has been hypothesised [122].

It would be essential to investigate the framework's sensitivity towards changes of the framework's design and also towards human perception. In the latter case one could present the framework and given list of literature information on a few selected pathogens to a group of people (preferably researchers working in the field of infectious diseases), asking them to independently generate transmission routes based on the information they have been given. Alternatively only the framework and the list of pathogens could be provided and free access to information sources would be granted. Subsequently the transmission networks would be analysed for differences.

The need for rating the importance of transmission routes was discussed in chapter 4. Weighted or without weights, the networks of transmission routes should also be analysed in a different way, i.e. on the overall network rather than vertex level. It has been shown that complex systems from diverse areas like, for example, technology, social science and biology exhibit universal architectural features when represented and analysed on a network level. This has led to the suggestion that "...similar laws may govern most complex networks in nature" [6]. The application of network analysis to epidemiology and infectious processes is currently a vivid field of research [196] [153] [111]. It therefore seems worthwhile to look for some well-known features such as the scale-free and small-world properties in these transmission networks. The small world property would suggest that an infections of a host species can reach another host species by only a few transmission routes. The scale-free property could identify so-called hubs which are vertices with a very high degree as compared to most other vertices in the network. They might constitute prime candidates for strategies to control the spread to other host species. One might also compare viruses and bacteria with respect to certain network features such as the degree distribution, mean path length or average clustering coefficient. A higher average cluster coefficient and therefore a potentially higher

connectivity within, for example, bacterial networks could indicate that bacteria are more likely to emerge or to cause outbreaks than viruses. Alternatively so-called network motifs can be used to compare pathogen networks, i.e. small subgraph or subnetwork structures that are over- or underrepresented when compared to similar randomised networks [116]. This could be used to characterise and thereby distinguish networks of bacteria from those of viruses. Depending on the definition and design of the motives, the motif patterns could also reveal patterns of pathogen transmission.

The results of this thesis strongly suggest the continuous screening and surveillance of livestock animals, their commercial products, their terrain, their equipment and perhaps even of people working in close/direct contact with livestock animals for a wide range of pathogens. In fact, the results also suggest including synanthrops into any livestock surveillance strategy since they tend to share an increased number of pathogens with livestock animals. The latter holds also true for arthropods such as ticks which might become even more important in the future due to climate changes and consequently modified geographic ranges of arthropod hosts.

A major contribution to the field of infectious diseases would involve the develpoment of a common language with well-defined and standardised terms and concepts. In case more than one concept or definition is circulating - addressing opposing views or not - researchers should clearly state the definition applied and/or refer to the corresponding source. Publications on findings for transmission pathways would best be published, at least partly or in addition to the actual publication, in a standardised fashion stating all the information required for a transmission route in a specified format. Such publications would make the literature far more accessible to computerised text mining methods thereby facilitating future research and reducing cost for extensive manual literature searches.

Having reached the end of my investigations, the two quotations at the very beginning of the general introduction now clearly must have originated from a period that already has been termed by the U.S. Institute of Medicine as the "era of complacency" [72][4]. Such statements resulted from an overconfidence in existing antibiotics and vaccines [172]. Surprisingly, it was known since a long time that viruses (i.e. oncoviruses) can be the cause of tumours. At least nowadays, the link between heart diseases and microorganisms is not a novelty anymore, e.g. infective endocarditis caused by bacteria of the *Streptococcus* genus.

[4]Note that an automatic text search of reference [38] did not find the expression "era of complacency" as claimed by the author of [72].

Bibliography

[1] ACHA, P. N., AND SZYFRES, B. *Zoonoses and communicable diseases common to man and animals: Bacteriosis and Mycoses*, 3rd ed., vol. 1. 2003. Science and Technical Publication No. 580.

[2] ALTIZER, S., NUNN, C. L., AND LINDENFORS, P. Do threatened hosts have fewer parasites? A comparative study in primates. *Journal of Animal Ecology 76*, 2 (Mar 2007), 304–314.

[3] ANTONIOU, G., AND VAN HARMELEN, F. *A Semantic Web Primer*. The MIT Press, 2004.

[4] AVILA-PIRES, F. D. The Use and Mis-use of some Ecological Terms and Concepts in Epidemiology. *Memorias do Instituto Oswaldo Cruz 90*, 5 (Sep-Oct 1995), 561–564.

[5] BALL, G. H., AND HALL, D. J. A clustering tecnique for summarizing multivariate data. *Behavioral Science 12*, 2 (Mar 1967), 153–155.

[6] BARABÁSI, A.-L., AND OLTVAI, Z. L. Network Biology: Understanding The Cell's Functional Organisation.

[7] BARD, J. B., AND RHEE, S. Y. Ontologies In Biology: Design, Applications And Future Challenges. *Nature Reviews. Genetics 5*, 3 (Mar 2004), 213–222.

[8] BEGG, A. P., JOHNSTON, K. G., HUTCHINS, D. R., AND EDWARDS, D. J. Some aspects of the epidemiology of equine salmonellosis. *Australian Veterinary Journal 65*, 7 (Jul 1988), 221–223.

[9] BEGON, M., BENNETT, M., BOWERS, R. G., FRENCH, N. P., HAZEL, S. M., AND TURNER, J. A clarification of transmission terms in host-microparasite models: numbers, densities and areas. *Epidemiology and Infection 129*, 1 (Aug 2002), 147–153.

[10] BEGON, M., TOWNSEND, C. R., AND HARPER, J. L. *Ecology: From Individuals to Ecosystems*, 4th ed. Blackwell Publishing Ltd, 2006.

[11] BENNETT, M. *Personal communication* (Apr 2011).

[12] BILL & MELINDA GATES FOUNDATION. *Global Health Program*, 2010. `http://www.gatesfoundation.org/global-health/Pages/overview.aspx`.

[13] BODENREIDER, O., AND STEVENS, R. Bio-ontologies: current trends and future directions. *Briefings In Bioinformatics 7*, 3 (Sep 2006), 256–274.

[14] Borland, E. D. Salmonella infection in poultry. *The Veterinary Journal 97*, 21 (Nov 1975), 406–408.

[15] Borland, E. D., Moryson, C. J., and Smith, G. R. Avian botulism and the high prevalence of Clostridium botulinum in the Norflok Broads. *The Veterinary Record 100*, 6 (Feb 1977), 106–109.

[16] Brenner, F. W., Villar, R. G., Angulo, F. J., Tauxe, R., and Swaminathan, B. *Salmonella* Nomenclature (Guest Commentary). *Journal of Clinical Microbiology 38*, 7 (Jul 2000), 2465–2467.

[17] Brown, C. Emerging zoonoses and pathogens of public health significance - an overview. *Revue scientifique et technique (International Office of Epizootics) 23*, 2 (Aug 2004), 435–442.

[18] Brown, I. H. The epidemiology and evolution of influenza viruses in pigs. *Veterinary Microbiology 74*, 1-2 (May 2000), 29–46.

[19] Brown, S. P. Do all parasites manipulate their hosts? *Behavioural Processses 68*, 3 (Mar 2005), 237–240.

[20] Burge, H. A. *Bioaerosols in the Residential Environment.* In Cox, C. S., Wathes, C. M. Bioaerosols handbook, CRC Press, Inc., 1995, ch. 21, pp. 579-597.

[21] Buxton, A., and Gordon, R. F. The Epidemiology And Control Of *Salmonella Thompson* Of Fowls. *The Journal of Hygiene (Lond) 45*, 3 (Aug 1947), 265-281.

[22] Center for Food Security and Public Health, Iowa State University. *Salmonellosis: Paratyphoid, Non-typhoidal Salmonellosis*, May 1 2005. `http://www.cfsph.iastate.edu/factsheets/pdfs/nontyphoidal_salmonellosis.pdf`.

[23] Cézilly, F. Host Manipulation by Parasites. *Behavioural Processses 68*, 3 (Mar 2005), 185–295.

[24] Chomsky, A. N. *Understanding Power: The Indispesable Chomsky.* 2003.

[25] Ciganovich, E., Friend, M., and Franson, J. C. *Field Manual of Wildlife Diseases: General Field Procedures and Diseases of Birds.* U.S. Dept. of the Interior; U.S. Geological Survey, 1999. Information and Technology Report 1999-001.

[26] Clark, R. G., Fenwick, S. G., Marchant, R. M., Swanney, S., Gill, J. M., Holmes, J. D., Leyland, M., and Davies, P. R. *Salmonella* Brandenburg - emergence of a new strain affecting stock and humans in the South Island of New Zealand. *New Zealand Veterinary Journal 52*, 1 (Feb 2004), 26–36.

[27] Cleaveland, S., Laurenson, M. K., and Taylor, L. H. Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences 356*, 1411 (Jul 2001), 991–999.

[28] Cleaveland, S., and Taylor, L. H. Personal communication regarding: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences 356*, 1411 (Jul 2001), 991–999.

[29] Clover, C. Someone must take control before foot and mouth turns into a disaster. *The Telegraph* (Sep 15 2007).

[30] COETZER, J. A. W., AND TUSTIN, R. C. *Infectious Diseases of Livestock*, 2nd ed., vol. 1-3. Oxford University Press, 2004.

[31] COLE, E. C., AND COOK, C. E. Characterization of infectious aerosols in health care facilities: An aid to effective engineering controls and preventive strategies. *American Journal of Infection Control 26*, 4 (Aug 1998), 453–464.

[32] COLVILLE, J. L., AND BERRYHILL, D. L. *Handbook Of Zoonoses: Identification And Prevention*. Mosby Elsevier, 2007.

[33] COX, C. S., AND WATHES, C. M. *Bioaerosols handbook*. CRC Press, Inc., 1995.

[34] CROSS, D. H., FREDRICKSON, L. H., AND REID, F. A. *Waterfowl Management Handbook*. U.S. Dept. of the Interior, Fish and Wildlife Service, 1988. Fish and Wildlife Leaflet 13. `http://www.nwrc.usgs.gov/wdb/pub/wmh/contents.html;~http://digitalcommons.unl.edu/icwdmwfm/42/`.

[35] CRUMP, J. A., GRIFFIN, P. M., AND ANGULO, F. J. Bacterial Contamination of Animal Feed and Its Relationship to Human Foodborne Illness. *Food Safety 35* (Oct 2002), 859–865.

[36] DASZAK, P., CUNNINGHAM, A. A., AND HYATT, A. D. Emerging infectious diseases of wildlife–threats to biodiversity and human health. *Science (New York, N.Y.) 287*, 5452 (Jan 2000), 443–449.

[37] DAVIDSON, I. Understanding K-Means Non-hierarchical Clustering. *SUNY Albany - Technical Report 02-2*.

[38] DAVIES, J. R., AND LEDERBERG, J. *Emerging Infectious Diseases From the Global to the Local Perspective*. U.S. Institute of Medicine, National Academy Press, Washington, D.C., 2001. Workshop Summary of the Forum on Emerging Infections.

[39] DAVIES, R. H., DALZIEL, R., GIBBENS, J. C., WILESMITH, J. W., RYAN, J. M. B., EVANS, S. J., BYRNE, C., PAIBA, G. A., PASCOE, S. J. S., AND TEALE, C. J. National survey for *Salmonella* in pigs, cattle and sheep at slaughter in Great Britain (1999-2000). *Journal of Applied Microbiology 96*, 4 (2004), 750–760.

[40] DAVIES, R. H., AND WRAY, C. Persistence of Salmonella enteritidis in poultry units and poultry food. *British Poultry Science 37* (Dec 1996), 589–596.

[41] DE JONG, M. C. M., STEGEMAN, A., VAN DER GOOT, J., AND KOCH, G. Intra- and interspecies transmission of H7N7 highly pathogenic avian influenzavirus during the avian influenza epidemic in the Netherlands in 2003. *Revue scientifique et technique (International Office of Epizootics) 28*, 1 (Sep 2009), 333–340.

[42] DEACON, T. W. *Memes as Signs* in the Dynamic Logic of Semiosis: Beyond Molecular Science and Computation Theory. In Wolff, K. E., Pfeiffer, H. D., Delugach, H. S. Conceptual Structures at Work: 12th International Conference on Conceptual Structures, Springer, July 2004, pp. 17-30.

[43] DECOSTER, J. *Meta-Analysis Notes*. Department of Psychology, University of Alabama, Sep 19 2004. `http://www.stat-help.com`.

[44] DEFELIPE, J. From the Connectome to the Synaptome: An Epic Love Story. *Science 330*, 6008 (Nov 2010), 1198–1201.

[45] Department for Environment, Food and Rural Affairs (DEFRA). *Zoonoses Report: United Kingdom 2005*, August 2006.

[46] Department for Environment, Food and Rural Affairs (DEFRA). *FMD 2007 Epidemiology Report - Situation at 12:00 Friday 21 September 2007. Day 49*, Sep 21 2007. http://www.defra.gov.uk/animalh/diseases/fmd/pdf/epireport240907.pdf.

[47] Department for Environment, Food and Rural Affairs (DEFRA). *FMD 2007. Summary Epidemiology Report - Situation as at 10:00 Thursday 09 August 2007. Day 6*, Aug 9 2007. http://www.defra.gov.uk/animalh/diseases/fmd/pdf/fmd-epireport090807.pdf.

[48] Department for Environment, Food and Rural Affairs (DEFRA). *Initial Epidemiological Report on the Outbreak of Bluetongue in East Anglia and South East England from Investigastions completed to 19 October 2007*, Oct 19 2007. http://www.defra.gov.uk/animalh/diseases/notifiable/bluetongue/pdf/epidemiology-report301007.pdf.

[49] Department for Environment, Food and Rural Affairs (DEFRA). *Outbreak of Highly Pathogenic H5N1 Avian Influenza Virus in Suffolk in January 2007*, Apr 5 2007. http://www.defra.gov.uk/animalh/diseases/notifiable/disease/ai/pdf/epid_findings050407.pdf.

[50] Dobson, A. Population dynamics of pathogens with multiple host species. *The American Naturalist 164 Suppl 5* (Nov 2004), S64–78.

[51] Donaldson, A. I. Quantitative data on airborne foot-and-mouth disease virus: its production, carriage and deposition. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences 302*, 1111 (Aug 1983), 529–534.

[52] Drake, J. W. Rates of spontaneous mutation among RNA viruses. *Proceedings of the National Academy of Sciences of the United States of America 90*, 9 (May 1993), 4171–4175.

[53] Drake, J. W. The Distribution of Rates of Spontaneous Mutation over Viruses, Prokaryotes and Eukaryotes. *Annals of the New York Academy of Sciences 870*, Molecular Strategies in Biological Evolution (May 1999), 100–107.

[54] Drake, J. W., Charelesworth, B., Charlesworth, D., and Crow, J. F. Rates of Spontaneous Mutation. *Genetics 148*, 4 (Apr 1998), 1667–1686.

[55] Ecker, D. J., Sampath, R., Willet, P., Wyatt, J. R., Massire, V. S. C., Hall, T. A., Hari, K., McNeil, J. A., Büchen-Osmond, C., and Budowle, B. The Microbial Rosetta Stone Database: A compilation of global and emerging infectious microorganisms and bioterrorist threat agents. *BMC Microbiology 5*, 19 (Apr 2005), 1–17.

[56] EFSA Journal, 341, 1-131. *Opinion of the Scientific Panel on Biological Hazards on the request from the Commission related to "Risk assessment and mitigation options of Salmonella in pig production"*, 2006. http://www.efsa.europa.eu/en/scdocs/doc/341.pdf.

[57] EFSA Journal, 8(1):1496. *Trends and sources of zoonoses and zoonotic agents and foodborne outbreaks in the European Union in 2008*, 2010. http://www.moredun.ac.uk/webfm_send/97.

[58] Estivill-Castro, V., and Yang, J. Fast and Robust General Purpose Custering Algorithm. *Data Mining and Knowledge Discovery 8*, 2 (Mar 2004), 127–150.

[59] EUZÉBY, J. P. LPSN: List of Procaryotic names with Standing in Nomenclature. http://www.bacterio.cict.fr;~http://www.bacterio.net.

[60] FENTON, A., AND PEDERSEN, A. B. Community Epidemiology Framework for Classifying Disease Threats. *Emerging Infectious Diseases 11*, 12 (Dec 2005), 1815–1821.

[61] FLORET, N., VIEL, J. F., MAUNY, F., HOEN, B., AND PIARROUX, R. Negligible Risk for Epidemics after Geophysical Disasters. *Emerging Infectious Diseases 12*, 4 (Apr 2006), 543–548.

[62] FLÜGGE, C. Über Luftinfektion. *Zeitschrift für Hygiene und Infektionskrankheiten 25* (1897), 179–193.

[63] FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. *Emergency Prevention System for Transboundary Animal Diseases.* http://www.fao.org/ag/AGAinfo/programmes/en/empres/diseases.asp.

[64] FRANCIOSA, G., FENICIA, L., CALDIANI, C., AND AURELI, P. PCR for Detection of *Clostridium botulinum* Type C in Avian and Environmental Samples. *Journal of Clinical Microbiology 34*, 4 (Apr 1996), 882–885.

[65] FRED, A. L. N., AND JAIN, A. K. Combining Multiple Clusterings Using Evidence Accumulation. *IEEE Transactions On Pattern Analysis And Machine Intelligence 27*, 6 (Jun 2005), 835–850.

[66] FRIEDMAN, C., AND PETERSEN, K. H. *Infection Control in Ambulatory Care.* Jones & Bartlett Publishers, 2003. An official publication of the Association for Professionals in Infection Control and Epidemiology Inc. (APIC).

[67] GANNON, B. W., HAYES, C. M., AND ROE, J. M. Survival rate of airborne *Mycobacterium bovis. Research in veterinary science 82*, 2 (Apr 2007), 169–172.

[68] GOGUEN, J. A. What is a Concept? *Conceptual Structures: Common Semantics for Sharing Knowledge (Series: Lecture Notes in Computer Science) 3596/2005* (Jul 2005), 52–77. http://www.springerlink.com/content/0d29925luepmvc48/fulltext.pdf.

[69] GOLDACRE, B. Bad science: Testing the plausibility effect. *The Guardian* (Jul 12 2008).

[70] GOOGLE.ORG. *Predict and Prevent*, 2010. http://www.google.org/predict.html.

[71] GRAY, R. An airborne, fast spreading virus: The disease is not just a threat to farmers but also the country's economy. *The Telegraph* (Aug 5 2007).

[72] GREGER, M. The Human/Animal Interface: Emergence and Resurgence of Zoonotic Infectious Diseases. *Critical Reviews in Microbiology 33*, 4 (Aug 2007), 243–299.

[73] GRESHAM, A. C. J. An Old Disease Of Increasing Significance - Salmonella Infection In Pigs In Great Britain. *The Pig Journal 37* (May 1996), 59–64.

[74] GRIFFIN, D. W., WESTPHAL, D. L., AND GRAY, M. A. Airborne microorganisms in the African desert dust corridor over the mid-Atlantic ridge, Ocean Drilling Program, Leg 209. *Aerobiologia 22*, 3 (2006), 211–226.

[75] HANDELAND, K., REFSUM, T., JOHANSEN, B. S., HOLSTAD, G., KNUTSEN, G., SOLBERG, I., SCHULZE, J., AND KAPPERUD, G. Prevalence of *Salmonella* Typhimurium infection in Norwegian hedgehog populations associated with two human disease outbreaks. *Epidemiology and Infection 128*, 3 (Jun 2002), 523–527.

[76] HANDL, J., KNOWELS, J., AND KELL, D. B. Computational cluster validation in post-genomic data analysis. *Bioinformatics 21*, 15 (May 2005), 3201–3212.

[77] HARI, K. L., GOATES, A. T., JAIN, R., TOWERS, A., HARPIN, V. S., ROBERTSON, J. M., WILSON, M. R., SAMANT, V. S., ECKER, D. J., MCNEIL, J. A., AND BUDOWLE, B. The Microbial Rosetta Stone: a database system for tracking infectious microorganisms. *International Journal of Legal Medicine 123*, 1 (Jan 2009), 65–69.

[78] HEESTERBEEK, J. A. P. A Brief History Of R_0 And A Recipe For Its Calculation. *Acta Biotheoretica 50*, 3 (Sep 2002), 189–204.

[79] HEFFERNAN, J. M., SMITH, R. J., AND WAHL, L. M. Perspectives on the basic reproductive ratio. *Journal of the Royal Society, Interface 2*, 4 (Sep 2005), 281–293.

[80] HENDERSON, M. Junk Medicine - Bioethics: Experts of no import. *The Times* (Jan 26 2008).

[81] HLAVSA, M. C., MOONAN, P. K., COWAN, L. S., NAVIN, T. R., KAMMERER, J. S., MORLOCK, G. P., CRAWFORD, J. T., AND LOBUE, P. A. Human tuberculosis due to *Mycobacterium bovis* in the United States, 1995-2005. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 47*, 2 (Jul 2008), 168–175.

[82] HOFWEBER, T. Logic and Ontology. *Stanford Encyclopedia of Philosophy* (Oct 2004). `http://plato.stanford.edu/entries/logic-ontology/#3.1`.

[83] IONNIDIS, J. P. A. Why Most Published Research Findings Are False. *PLoS Medicine 2*, 8 (Aug 2005), 0696–0701.

[84] JOHNSON, L., DEAN, G., RHODES, S., HEWINSON, G., VORDERMEIER, M., AND WANGOO, A. Low-dose *Mycobacterium bovis* infection in cattle results in pathology indistinguishable from that of high-dose infection. *Tuberculosis 87*, 1 (Jan 2007), 71–76.

[85] JONES, K. E., PATEL, N. G., LEVY, M. A., STOREYGARD, A., BALK, D., GITTLEMAN, J. L., AND DASZAK, P. Global trends in emerging infectious diseases. *Nature 451* (Feb 2008), 990–993.

[86] JORDAN, F., PATTISON, M., ALEXANDER, D., AND FARAGHAR, T. *Poultry Diseases*, 5th ed. W.B. Saunders of Harcourt Publishers Limited, 2001.

[87] JUNGNICKEL, D. *Graphs, Networks And Algorithms*, 3rd ed., vol. 5 of *Algorithms and Computation in Mathematics*. Springer, Heidelberg, 2008.

[88] KAHN, C. M., LINE, S., AND AIELLO, S. E. *The Merck Veterinary Manual*, 9th ed. MERCK & CO., INC., 2005.

[89] KARYPIS, G., HAN, E.-H. S., AND KUMAR, V. Cameleon: Hierarchical Custering Using Dynamic Modeling. *IEEE Computer 32*, 8 (Aug 1999), 68–75.

[90] KAUFMAN, L., AND ROUSSEEUW, P. J. *Finding Groups In Data: An Introduction To Custer Analysis*. Wiley & Sons, Inc., 1990.

[91] KITCHING, P., THOMSON, G., ALBRECHT, S., SABARA, M., AND WERLING, D. Transboundary disease management: the theory and the practice–the science and the politics. *Transboundary and emerging diseases 55*, 1 (2008), 1–2.

[92] KITCHING, R. P., THRUSFIELD, M. V., AND TAYLOR, N. M. Use and abuse of mathematical models: an illustration from the 2001 foot and mouth disease epidemic in the United Kingdom. *Revue scientifique et technique (International Office of Epizootics) 25*, 1 (Apr 2006), 293–311.

[93] KOPERSKI, J. Models. *The Internet Encyclopedia of Philosophy* (2006). `http://www.iep.utm.edu/m/models.htm`.

[94] KRAUS, H., WEBER, A., APPEL, M., ENDERS, B., ISENBERG, H. D., SCHIEFER, H. G., SLENCZKA, W., VON GRAEVENITZ, A., AND ZAHNER, H. *Zoonoses: Infectious Diseases Transmissible from Animals to Humans*, 3rd ed. ASM Press, Washington, D.C., 2003.

[95] KRAUSE, M. *Emerging Infections*. Biomedical Research Reports. Academic Press, 1998.

[96] KRIEK, N. P. J., AND ODENDAAL, M. W. *Botulism*. In Coetzer, J. A. W., Tustin, R. C. Infectious Diseases of Livestock, Oxford University Press, 2004, 2nd ed., vol. 3, pp. 1885-1902.

[97] KÜHNE, T. What is a Model? *Dagstuhl Seminar Proceedings 04101: Language Engineering for Model-Driven Software Development* (Feb 2005). `http://drops.dagstuhl.de/opus/volltexte/2005/23;http://www.informatik.uni-trier.de/~ley/db/conf/dagstuhl/P4101.html`.

[98] KURIS, A. M., BLAUSTEIN, A. R., AND ALIÓ, J. J. HOSTS AS ISLANDS. *The American Naturalist 116*, 4 (Oct 1980), 570–586.

[99] LAFFERTY, K. D. The Evolution of Trophic Transmission. *Parasitology Today 15*, 3 (Mar 1999), 111–115.

[100] LAKOFF, G., AND JOHNSON, M. *Metaphors We Live By*. The University of Chicago Press, 1980.

[101] LAST, J. M. *A Dictionary of Epidemiology*, 4th ed. Oxford University Press, 2001.

[102] LEDERBERG, J., SHOPE, R. E., AND OAKS JR., S. C. *Emerging Infections: Microbial Threats to Health in the United States*. U.S. Institute of Medicine, National Academy Press, Washington, D.C., 1992.

[103] LEVER, M. S., AND WILLIAMS, A. Cross-infection of chicks by airborne transmission of *Salmonella enteritidis* PT4. *Letters in Applied Microbiology 23*, 5 (Nov 1996), 347–349.

[104] LEVINS, R. The Challange of New Diseases. *Annals of the New York Academy of Science 740*, 1 (Dec 1994), xvii–xix.

[105] MADIGAN, M. J., AND MARTINKO, J. M. *Brock: Biology of Microorganisms*, 11th ed. Pearson Prentice Hall, 2006.

[106] MADSEN, M. Prevalence and serovar distribution of *Salmonella* in fresh and frozen meat from captive Nile crocodiles (*Crocodylus niloticus*). *International Journal of Food Microbiology 29*, 1 (Feb 1996), 111–118.

[107] Magnino, S., Colin, P., Dei-Cas, E., Madsen, M., McLauchlin, J., Nöckler, K., Maradona, M. P., Tsigarida, E., Vanopdenbosch, E., and Peteghem, C. V. Biological risks associated with consumption of reptile products. *International Journal of Food Microbiology 134*, 3 (Sep 2009), 163–175.

[108] Magnus, M. *Essentials of Infectious Disease Epidemiology.* Jones and Bartlett Publishers, Inc., 2008.

[109] Margolis, E., and Laurence, S. Concepts. *Stanford Encyclopedia of Philosophy* (Feb 2006). `http://plato.stanford.edu/entries/concepts/`.

[110] Martin, W. B., and Aitken, I. D. *Diseases of Sheep*, 3rd ed. Blackwell Science Ltd., 2000.

[111] Mason, O., and Verwoerd, M. Graph theory and networks in Biology. *IET Systems Biology 1*, 2 (Mar 2007), 89–119.

[112] McCray, A. T. Conceptualizing the world: Lessons from history. *Journal of Biomedical Informatics 39*, 3 (Jun 2006), 267–273.

[113] McGee, C. E., Schneider, B. S., Girard, Y. A., Vanlandingham, D. L., and Higgs, S. Nonviremic Transmission Of West Nile Virus: Evaluation Of The Effects Of Space, Time And Mosquito Species. *The American Journal of Tropical Medicine and Hygiene 76*, 3 (2007), 442–430.

[114] Meara, P. J. Salmonellosis In Slaughter Animals As A Source Of Human Food Poisoning. *Journal of the South African Veterinary Association 44*, 3 (1973), 215–233.

[115] Merianos, A. Surveillance and Response to Disease Emergence. *Current Topics in Microbiology and Immunology 315* (2007), 477–509.

[116] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. Network Motivs: Simple Building Blocks of Complex Networks. *Science 298*, 5594 (Oct 2002), 824–827.

[117] Minati, G., and Pessa, E. *Collective Beings: Contemporary Systems Thinking.* Springer, 2006.

[118] Minga, U. M., Licht, H. H., and Shlundt, J. Four Outbreaks Of Salmonellosis Due To *Salmonella Typhimurium* Among Cattle In One District In Denmark: Case Reports. *British Veterinary Journal 141*, 5 (Sep-Oct 1985), 490–497.

[119] Ministry of Agriculture, Fisheries and Food. *Atmospheric emissions of particulates from agriculture: a scoping study*, 2000. Project code WA 0802.

[120] Minsky, M. L. *THE SOCIETY OF MIND.* Touchstone, Simon & Schuster, Inc., Rockefeller Center, New York, 1988.

[121] Molla, B., Alemayehu, D., and Salah, W. Sources and distribution of *Salmonella* serotypes isolated from food animals, slaughterhouse personnel and retail meat products in Ethopia: 1997-2002. *Ethiopian Journal of Health Development 17*, 1 (2003), 63–70.

[122] Monath, T. P. Ecology of Marburg and Ebola Viruses: Speculations and Directions for Future Research. *Journal of Infectious Diseases 179*, Suppl 1 (Feb 1999), S127–S138.

[123] MORAWSKA, L. Droplet fate in indoor environments, or can we prevent the spread of infection? *Indoor Air 16*, 5 (Oct 2006), 335–347.

[124] MORSE, S. S. Global microbial traffic and the interchange of disease. *American Journal of Public Health 82*, 10 (Oct 1992), 1326–1327.

[125] MORSE, S. S. Factors and determinants of disease emergence. *Revue scientifique et technique (International Office of Epizootics) 23*, 2 (Aug 2004), 443–451.

[126] MUCKLE, C. A., MENZIES, P. I., LI, Y., AND HWANG, Y. T. Analysis of the immunodominant antigens of *Corynebacterium pseudotuberculosis*. *Veterinary Microbiology 30*, 1 (Jan 1992), 47–58.

[127] MUMFORD, J. A. Vaccines and viral antigenic diversity. *Revue scientifique et technique (International Office of Epizootics) 26*, 1 (Apr 2007), 69–90.

[128] NAKAMURA, M., TAKAGI, M., TAKAHASHI, T., SUZUKI, S., SATO, S., AND TAKEHARA, K. The effect of the flow of air on horizontal transmission of *Salmonella enteritidis* in chickens. *Avian Diseases 41*, 2 (Apr-Jun 1997), 354–360.

[129] NATIONAL CENTRE FOR ZOONOSIS RESEARCH. `http://www.zoonosis.ac.uk/zoonosis/login.aspx`.

[130] NAUERBY, B., PEDERSEN, K., DIETZ, H. H., AND MADSEN, M. Comparison of Danish Isolates of *Salmonella enterica* Serovar Enteritidis PT9a and PT11 from Hedgehogs (*Erinaceus europaeus*) and Humans by Plasmid Profiling and Pulsed-Field Gel Electrophoresis. *Journal of Clinical Microbiology 38*, 10 (Oct 2000), 3631–3635.

[131] NCBI TAXONOMY HOMEPAGE. `http://www.ncbi.nlm.nih.gov/Taxonomy`.

[132] NELSON, G. S. Schistosome Infections As Zoonoses In Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene 54*, 4 (May 1960), 301–316.

[133] NEWMAN, M. E. J. A measure of betweenness centrality based on random walks. *Social Networks 27*, 1 (Jan 2005), 39–54.

[134] NICAS, M., NAZAROFF, W. W., AND HUBBARD, A. Toward understanding the risk of secondary airborne infection: emission of respirable pathogens. *Journal of Occupational and Environmental Hygiene 2*, 3 (Mar 2005), 143–154.

[135] NORMAN, R., ROSS, D., LAURENSON, M. K., AND HUDSON, P. J. The role of non-viraemic transmission on the persistence and dynamics of a tick borne virus - louping ill in red grouse (lagopus lagopus scoticus) and mountain hares (lepus timidus). *Journal of Mathematical Biology 48*, 2 (Feb 2004), 119–134.

[136] NUNN, C. L., AND ALTIZER, S. M. The Global Mammal Parasite Database: An Online Resource for Infectious Disease Records in Wild Primates. *Evolutionary Anthropology 14*, 1 (Feb 2005), 1–2.

[137] NUTTALL, P. A. Displaced tick-parasite interactions at the host interface. *Parasitology 116*, Supplement S1 (1998), S65–S72.

[138] OJOSNEGROS, S., BEERENWINKEL, N., NOVAK, M. A., ESCARMIS, C., AND DOMINGO, E. Competition-colonization dynamics in an RNA virus. *Proceedings of the National Acedemy of Sciences (PNAS) 107*, 5 (Feb 2010), 2108–2112.

[139] Oliveira, C. J., Carvalho, L. F., and Garcia, T. B. Experimental airborne transmission of *Salmonella* Agona and *Salmonella* Typhimurium in weaned pigs. *Epidemiology and infection 134*, 1 (Feb 2006), 199–209.

[140] O'Loughlin, T. Can an ant eat australia's prime pest, or will nature win again? *The Guardian* (Apr 4 2009).

[141] Ortiz, N. E., and Smith, G. R. *Clostridium botulinum* type A in the gut contents of a British gull. *The Veterinary Record 135*, 3 (Jul 1994), 68.

[142] Palmer (Lord Soulsby), S. R., and Simpson, D. I. H. *Zoonoses: Biology, Clinical Practice, and Public Health Control.* Oxford University Press, 1998.

[143] Pedersen, A. B., Altizer, S., Poss, M., Cunningham, A. A., and Nunn, C. L. Patterns of host specificity and transmission among parasites of wild primates. *International journal for parasitology 35*, 6 (May 2005), 647–657.

[144] Pedersen, A. B., and Fenton, A. Emphasizing the ecology in parasite community ecology. *Trends in Ecology and Evolution (Personal edition) 22*, 3 (Mar 2007).

[145] Pedersen, A. B., Jones, K. E., Nunn, C. L., and Altizer, S. Infectious diseases and extinction risk in wild mammals. *Conservation biology : the journal of the Society for Conservation Biology 21*, 5 (Oct 2007), 1269–1279.

[146] Perkins, S. E., and Fenton, A. Helminths as vectors of pathogens in vertebrate hosts: A theoretical approach. *International Journal for Parasitology 36*, 8 (Jul 2006), 887–894.

[147] Polley, L. Navigating parasite webs and parasite flow: emerging and re-emerging parasitic zoonoses of wildlife origin. *International Journal for Parasitology 35*, 11-12 (Oct 2005), 1279–1294.

[148] Polymenakou, P. N., Mandalakis, M., Stephanou, E. G., and Tselepides, A. Particle Size Distribution of Airborne Microorganisms and Pathogens during an Intense African Dust Event in the Eastern Mediterranean. *Environmental Health Perspectives 116*, 3 (Mar 2008), 292–296.

[149] Poppe, C., Smart, N., Khakhria, R., Johnson, W., Spika, J., and Prescott, J. *Salmonella typhimurium* DT104: A virulent and drug-resistent pathogen. *The Canadian Veterinary Journal 39*, 9 (Sep 1998), 559–565.

[150] Possas, C. A. Social ecosystem health: confronting the complexity and emergence of infectious diseases. *Cadernos de saude publica / Ministerio da Saude, Fundacao Oswaldo Cruz, Escola Nacional de Saude Publica 17*, 1 (Jan-Feb 2001), 31–41.

[151] Poulin, R. Macroecological patterns of species richness in parasite assemblages. *Basic and Applied Ecology 5*, 5 (Nov 2004), 423–434.

[152] Radostits, O. M., Gay, C. C., Hinchcliff, K. W., and Constable, P. D. *Veterinary Medicine: A textbook of the diseases of cattle, horses, sheep, pigs, and goats*, 10th ed. Saunders Elsevier, Elsevier Limited, 2007.

[153] Read, J. M., and Keeling, M. J. Disease evolution on networks: the role of contact structure. *Proceedings of the Royal Society (London). Series B. Biological sciences 270*, 1516 (Apr 2003), 699–708.

[154] Reperant, L. A. Applying the Theory of Island Biogeography to Emerging Pathogens: Towards Predicting the Sources of Future Emerging Zoonotic and Vector-Borne Diseases. *Vector-Borne and Zoonotic Diseases* (Jul 2009). Online Ahead of Print.

[155] Riley, R. L. Airborne Iinfection. *The American Journal of Medicine 57* (Sep 1974), 466–475.

[156] Ross, T. D. Accurate confidence intervals for binomial proportion and Poisson rate estimation. *Computers in Biology and Medicine 33*, 6 (Nov 2003), 509–531.

[157] Rousseeuw, P. J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Custer Analysis. *Journal of Computational and Applied Mathematics 20* (Mar 1987), 53–65.

[158] Sackett, D. L. Bias In Analytic Research. *Journal of Chronic Diseases 32* (1979), 51–63.

[159] Saif, Y. M., Barnes, H. J., Fadly, A. M., Glisson, J. R., McDougald, L. R., and Swayne, D. E. *Diseases Of Poultry*, 11th ed. Blackwell Publishing, 2003.

[160] Sambraus, H. H. *A Colour Atlas of Livestock Breeds: 220 breeds in words and picture*, 3rd ed. Wolfe Publishing Ltd., 1992.

[161] Santamaría, J., and Toranzos, G. A. Enteric pathogens and soil: a short review. *International Microbiology: The official Journal of the Spanish Society for Microbiology 6*, 1 (Mar 2003), 5–9.

[162] Schneider, T. D. Information Theory Primer, version 2.64. `http://www-lmmb.ncifcrf.gov/~tom/paper/primer`.

[163] Schnitzler, S. U., and Schnitzler, P. An update on swine-origin infuenza virus A/H1N1: a review. *Virus Genes 39*, 3 (Dec 2009), 279–292.

[164] Shakespeare, M. *Zoonoses*, 1st ed. Pharmaceutical Press, 2002.

[165] Sharma, S., Sachdeva, P., and Virdi, J. S. Emerging water-borne pathogens. *Applied Microbiology and Biotechnology 61*, 5-6 (Jun 2003), 424–428.

[166] Shlens, J. A Tutorial on Principal Component Analysis - version 2. *Institute for Nonlinear Science, University of California, San Diego* (Dec 2005), 1–13.

[167] Siegel, J. D., Jackson, M., Chiarello, L., and the Healthcare Infection Control Practices Advisory Committee. *Guideline for Isolation Precautions: Preventing Transmission of Infectious Agents in Healthcare Settings 2007*. Centers for Disease Control and Prevention (CDC), June 2007. `http://www.cdc.gov/ncidod/dhqp/pdf/guidelines/Isolation2007.pdf`.

[168] Skovgaard, N. New trends in emerging pathogens. *International journal of food microbiology 120*, 3 (Dec 2007), 217–224.

[169] Slovic, P. Trust, Emotion, Sex, Politics, and Science: Surveying the Risk-Assessment Battlefield. *Risk Analysis 19*, 4 (1999), 689–701.

[170] Smith, J. M. B., and Robinson, R. A. *Salmonella Typhimurium* In New Zealand Hedgehogs. *New Zealand Veterinary Journal 12*, 5 (Oct 1964), 111–112.

[171] Smith-Palmer, A., Stewart, W. C., Mather, H., Greig, A., Cowden, J. M., and Reilly, W. J. Epidemiology of *Salmonella enterica* serovars Enteritidis and Typhimurium in animals and people in Scotland between 1990 and 2001. *The Veterinary Record 153*, 17 (Oct 2003), 517–520.

[172] Smolinski, M. S., Hamburg, M. A., and Lederberg, J. *Microbial Threats to Health: Emergence, Detection, and Response.* U.S. Institute of Medicine, The National Academies Press, Washington, D.C., 2003.

[173] Spratt, B. G. Independent Review of the safety of UK facilities handling foot-and-mouth disease virus, Aug 31 2007. Presented to the Secretary of State for Environment, Food and Rural Affairs and the Chief Veterinary Officer.

[174] Staab, S., and Studer, R. *Handbook on Ontologies (Series: International Handbooks on Information Systems).* Springer-Verlag, Berlin, Heidelberg, 2004.

[175] Stallknecht, D. E. Impediments to Wildlife Disease Surveillance, Research, and Diagnostics. *Current Topics in Microbiology and Immunology 315* (2007), 445–461.

[176] Stärk, K. D. C. The Role of Infectious Aerosols in Disease Transmission in Pigs. *The Veterinary Journal 158*, 3 (11 1999), 164–181.

[177] Stephen, C., Artsob, H., Bowie, W. R., Drebot, M., Fraser, E., Leighton, T., Morshed, M., Ong, C., and Patrick, D. Perspectives on emerging zoonotic disease research and capacity building in Canada. *The Canadian Journal of Infectious Diseases & Medical Microbiology = Journal canadien des maladies infectieuses et de la microbiologie médicale / AMMI Canada 15*, 6 (Nov 2004), 339–344.

[178] Stephens, D. S., Moxon, E. R., Adams, J., Altizer, S., Antonovics, J., Aral, S., Berkelman, R., Bond, E., Bull, J., Cauthen, G., Farley, M. M., Glasgow, A., Glasser, J. W., Katner, H. P., Kelley, S., Mittler, J., Nahmias, A. J., Nichol, S., Perrot, V., Pinner, R. W., Schrag, S., Small, P., and Thrall, P. H. Emerging and reemerging infectious diseases: a multidisciplinary perspective. *The American Journal of the Medical Sciences 315*, 2 (Feb 1998), 64–75.

[179] Straw, B. E., Zimmerman, J. J., D'Allaire, S., and Taylor, D. J. *Diseases Of Swine*, 9th ed. Blackwell Publishing, 2006.

[180] Strehl, A., and Ghosh, J. Cluster Ensembles - A Knowledge Reuse Framework for Combining Multiple Partitions. *Journal of Machine Learning Research 3* (Dec 2002), 583–617.

[181] Taylor, D. J. *PIG DISEASES*, 6th ed. Farming Press, 1995.

[182] Taylor, L. H., Latham, S. M., and Woolhouse, M. E. J. Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences 356*, 1411 (Jul 2001), 983–989.

[183] Tindall, B. J., Grimont, P. A. D., Garrity, G. M., and Euzéby, J. P. Nomenclature and taxonomy of the genus *Salmonella*. *International Journal of Systematic and Evolutionary Microbiology 55*, Pt 1 (Jan 2005), 521–524.

[184] Tramper, J. Modern biotechnology: a new box for Pandora? *Tijdschrift voor Diergeneeskunde 127*, 4 (Feb 2002), 116–119.

[185] TRAUB-DAGATZ, J. L., SALMAN, D., AND JONES, R. L. Epidemiologic study of salmonellae shedding in the feces of horses and potential risk factors for development of the infection in hospitalised horses. *JAVMA 196*, 10 (May 1990), 1617–1622.

[186] TRIBUS, M. *Thermostatics and Thermodynamics.* D. van Nostrand Company, Inc., Princeton, New York, 1961.

[187] TRINGE, S. G., ZHANG, T., LIU, X., YU, Y., LEE, W. H., YAP, J., YAO, F., SUAN, S. T., ING, S. K., HAYNES, M., ROHWER, F., WEI, C. L., TAN, P., BRISTOW, J., RUBIN, E. M., AND RUAN, Y. The Airborne Metagenome in an Indoor Urban Environment. *PLoS ONE 3*, 4 (Apr 2008), e1862.

[188] VANSELOW, B. A., HUM, S., HORNITZKI, M. A., EAMENS, G. J., AND QUINN, K. *Salmonella* Typhimurium persistence in a Hunter Valley dairy herd. *Australian Veterinary Journal 85*, 11 (Nov 2007), 446–450.

[189] VETERINARY LABORATORIES AGENCY (VLA). *Salmonella in Livestock Production in GB - 2009 report*, 2009. `http://www.defra.gov.uk/vla/reports/rep_salm_rep09.htm`.

[190] VIDAL, J. Big outbreak feared as bluetongue spreads. *The Guardian* (Sep 29 2007).

[191] VISSER, I. J. R. Cutaneous salmonellosis in veterinarians. *The Veterinary Record 129* (1991), 364.

[192] WARWICK, C., LAMBIRIS, A. G. L., WESTWOOD, D., AND STEEDMAN, C. Reptile-related salmonellosis. *Journal of the Royal Society of Medicine 94*, 3 (Mar 2001), 124–126.

[193] WATHES, C. M. *Bioaerosols in Animal Houses.* In Cox, C. S., Wathes, C. M. Bioaerosols handbook, CRC Press, Inc., 1995, ch. 20, pp. 547-577.

[194] WATHES, C. M., ZAIDAN, W. A. R., PEARSON, G. R., HINTON, M., AND TODD, N. Aerosol infection of calves and mice with *Salmonella typhimurium*. *The Veterinary Record 123*, 23 (Dec 1988), 590–594.

[195] WATSON, J. T., GAYER, M., AND CONNOLLY, M. A. Epidemics after Natural Disasters. *Emerging Infectious Diseases 13*, 1 (Jan 2007), 1–5.

[196] WATTS, D. J., AND STROGATZ, S. H. Collective dynamics of 'small world' networks. *Nature 393*, 6684 (Jun 1998), 440–442.

[197] WEBSTER, J. P. Parasitic manipulation: where else should we go? *Behavioural Processses 68*, 3 (Mar 2005), 275–277.

[198] WEIL, B. J., MARTENS, P. B., AND HARTE, J. S. Iguana-Associated Salmonellosis in a Young Adult. *Journal of Adolescent Health 17*, 2 (Aug 1995), 120–122.

[199] WELLS, W. F. On Air-Borne Infection. S II. Droplets And Droplet Nuclei. *The American Journal of Hygiene 20* (Jun 1934), 611–618.

[200] WILCKE JR, B. W., MIDURA, T. F., AND ARNON, S. S. Quantitative evidence of intestinal colonization by *Clostridium botulinum* in four cases of infant botulism. *The Journal of Infectious Diseases 141*, 4 (Apr 1980), 419–423.

[201] WILLIAMS, B. M. Environmental considerations in salmonellosis. *The Veterinary Record 96*, 14 (Apr 1975), 318–321.

[202] WILLIAMS, E. Veterinary surgeons as vectors of salmonella dublin. *British Medical Journal 280*, 6217 (Mar 1980), 815–818.

[203] WOLFE, N. D., DUNAVAN, C. P., AND DIAMOND, J. Origins of major human infectious diseases. *Nature 447* (May 2007), 279–283.

[204] WOOLHOUSE, M., AND GAUNT, E. Ecological Origins of Novel Human Pathogens. *Critical Reviews in Microbiology 33*, 4 (Aug 2007), 231–242.

[205] WOOLHOUSE, M. E. J. Personal communication regarding: *Emerging Infectious Diseases 11*, 12 (Apr 2011), 1842–1847.

[206] WOOLHOUSE, M. E. J., AND GOWTAGE-SEQUERIA, S. Host Range and Emerging and Reemerging Pathogens. *Emerging Infectious Diseases 11*, 12 (Dec 2005), 1842–1847.

[207] WOOLHOUSE, M. E. J., TAYLOR, L. H., AND HAYDON, D. T. Population Biology of Multihost Pathogens. *Ecology and Evolution of Infection 292*, 5519 (May 2001), 1109–1112.

[208] WORLD HEALTH ORGANIZATION (WHO). *Zoonoses*, Jul 2008. `http://www.who.int/topics/zoonoses/en/`.

[209] WORLD ORGANIZATION FOR ANIMAL HEALTH, OFFICE INTERNATIONAL DES ÉPIZOOTIES (OIÉ). *Old Classification of Diseases Notifiable to the OIÉ*. `http://www.oie.int/eng/maladies/en_oldclassification.htm`.

[210] WRAY, C., SOJKA, W. J., AND BELL, J. C. Salmonella infections in horses in England and Wales, 1973 to 1979. *The Veterinary Record 109*, 18 (Oct 1981), 398–401.

[211] WUCHTY, S., AND STADLER, P. F. Centers of complex networks. *Journal of Theoretical Biology 223*, 1 (Jul 2003), 45–53.

[212] XIE, X., LI, Y., CHWANG, A. T., HO, P. L., AND SETO, W. H. How far droplets can move in indoor environments - revisiting the Wells evaporation-falling curve. *Indoor air 17*, 3 (Jun 2007), 211–225.

[213] XU, R., AND WUNSCH, D. Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks 16*, 3 (May 2005), 645–678.

[214] ZHAO, Y., AND KARYPIS, G. Data Clustering in Life Sciences. *Molecular Biotechnology 31* (2005), 55–80.

Index