The Impact of Anthropologically Motivated Human Social Networks on the Transmission Dynamics of Infectious Disease

by

Hans P. Nesse

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved July 2018 by the
Graduate Supervisory Committee:

Ana Magdalena Hurtado, Chair
Carlos Castillo-Chavez
Anuj Mubayi

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

Understanding the consequences of changes in social networks is an important anthropological research goal. This dissertation looks at the role of data-driven social networks on infectious disease transmission and evolution. The dissertation has two projects. The first project is an examination of the effects of the superspreading phenomenon, wherein a relatively few individuals are responsible for a disproportionate number of secondary cases, on the patterns of an infectious disease. The second project examines the timing of the initial introduction of tuberculosis (TB) to the human population. The results suggest that TB has a long evolutionary history with hunter-gatherers. Both of these projects demonstrate the consequences of social networks for infectious disease transmission and evolution.

The introductory chapter provides a review of social network-based studies in anthropology and epidemiology. Particular emphasis is paid to the concept and models of superspreading and why to consider it, as this is central to the discussion in chapter 2. The introductory chapter also reviews relevant epidemic mathematical modeling studies.

In chapter 2, social networks are connected with superspreading events, followed by an investigation of how social networks can provide greater understanding of infectious disease transmission through mathematical models. Using the example of SARS, the research shows how heterogeneity in transmission rate impacts superspreading which, in turn, can change epidemiological inference on model parameters for an epidemic.

Chapter 3 uses a different mathematical model to investigate the evolution of TB in hunter-gatherers. The underlying question is the timing of the introduction of TB to the human population. Chapter 3 finds that TB's long latent period is consistent with the evolutionary pressure which would be exerted by transmission on a hunter-

gatherer social network. Evidence of a long coevolution with humans indicates an early introduction of TB to the human population.

Both of the projects in this dissertation are demonstrations of the impact of various characteristics and types of social networks on infectious disease transmission dynamics. The projects together force epidemiologists to think about networks and their context in nontraditional ways.

To my family.

ACKNOWLEDGMENTS

This dissertation would not have been possible without the help of my own social network. First, my unfailingly supportive advisor, Dr. Hurtado, has spent countless hours discussing anthropology and writing with me. She has also been an invaluable advocate. My dissertation has grown and improved under her suggestions and advising, and by the same token, I have grown and (hopefully) improved as an researcher. It is to her that I owe my interest in global health. I have learned a great deal as her student. My committee, Dr. Castillo-Chavez and Dr. Mubayi have helped tremendously in this process, and I owe them both a great deal of thanks. Each has contributed hours of reading, suggestions, and advice which has proven essential in this process. Dr. Mubayi's willingness to talk me through difficult problems has been incredible, and Dr. Castillo-Chavez's suggestions for improving the text have improved this dissertation immeasurably.

My trek through to this point in my academic career was helped along by many others at Arizona State University. Dr. Chowell-Puente first suggested this as an area of research, and guided me on many early problems. I owe a great debt to Dr. Marean, who gave me my first field work experience at ASU. The administrative and support staff in the School of Human Evolution and Social Change, especially Dr. Georgianna Oakley Miller, Chelle Brookes, and Genise Denoire have all done incredible work keeping me on track, listening to my problems, and helping me find a way through them.

My social network of support also must include E. Susanne Daly. She has been a constant source of support throughout my time in graduate school. Her goofy and generous spirit has improved me as a student and as a person. She kept my attitude positive and focused, even through difficult times. More broadly, the other current and former graduate students of SHESC have been a tremendous help in my time at

ASU, and deserve far more thanks than I can put in this limited space. They are my cohort, lab mates, coauthors, and most importantly, friends. My work on this dissertation would not have been possible without their constant support.

During my final years at ASU, I was honored to work at the New School for the Arts and Academics (NSAA). It was a privileged to work with incredible group of educators and administrators. They have been supportive through my time writing this dissertation. And I need to also thank my students at NSAA, from whom I have learned so much.

Finally, although certainly not least, I owe a tremendous debt to my family, particularly Jan Leistikow, my mother. The many Nesses, Petersons, Hedges, and Leistikows have been a constant source of nurturing support. It is to them that I dedicate this dissertation.

TABLE OF CONTENTS

Page

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER

1 INTRODUCTION 1

1.1 Research Questions 2

1.2 Background 3

1.2.1 Connection to Medical Anthropology 3

1.2.2 Social Networks and Disease Transmission 8

1.2.3 Biological Evolution of Humans and Infectious Disease 12

1.2.4 Mathematical Modeling 16

1.2.5 Subcritical Epidemics and R_0 21

1.2.6 Terminology 26

1.2.7 Subcritical Disease Transmission on Social Networks 26

1.2.8 TB in Hunter-Gatherer Populations 27

1.3 Brief Results 27

2 SUBCRITICAL EPIDEMICS AND SUPERSPREADING 29

2.1 Introduction 29

2.1.1 Severe Acute Respiratory Syndrome 31

2.1.2 Networks and Superspreading Events 33

2.2 Hypothesis 34

2.3 Methods 35

2.3.1 Modeling Logic 35

2.3.2 Model Structure 35

2.3.3 Justification of the Negative Binomial 36

CHAPTER Page

2.3.4 Model Parameters 37
2.3.5 Simulations 41
2.4 Results 42
2.5 Discussion 44
2.6 Conclusions 49
3 THE EFFECT OF CONTRACT NETWORKS ON TB TRANSMISSION 51
3.1 Introduction 51
3.2 Background 52
3.2.1 TB Origins and Evolution 52
3.2.2 TB in Hunter-Gatherer Populations 55
3.2.3 Hunter-Gatherer Population Structure 56
3.2.4 The Basic Reproductive Number R_0 57
3.3 Methods 58
3.3.1 Model Construction 59
3.3.2 Parameters 60
3.3.3 Calculation of Active R_0 64
3.4 Results 65
3.4.1 Main Result 65
3.4.2 Active R_0 66
3.4.3 Parameter Sensitivity 66
3.5 Discussion 66
3.6 Conclusion 71
4 CONCLUSION 72
4.1 Anthropological Conclusions 73

CHAPTER Page

4.2 Infectious Disease Conclusions 74
4.3 Discussion 75
4.3.1 Limitations 76
4.4 Conclusion 77
REFERENCES 78

LIST OF TABLES

Table Page

1.1 Definitions of mathematical epidemiology terms used in this paper. 25

2.1 The parameter values used in this model. 39

2.2 Percentage of simulations which were between the specified range of secondary cases. * The percentage rounded to zero, however there were two instances of 1000+ sized simulations for $\phi_\alpha = 5$ and four instances for $\phi_\alpha = 7$. 42

2.3 Total number of cases of SARS by country, data compiled by WHO (2003b). No distinction was drawn between individuals who arrived in each country with SARS, and within-country transmission. China includes only mainland China, excluding its Special Administrative Regions. 46

2.4 Cities with case counts exceeding 50. It is very likely that there are additional cities in China which have not been recorded or reported. Most cities and towns which had reported case counts reported fewer than 50. 47

3.1 The best estimate of the parameter values used in the model. See subsequent discussion for references and estimates of variability in these parameters. 61

3.2 Active R_0 values for different values of p_a using the model parameters found above. These R_0 values indicate very high rates of transmission. . 68

3.3 Testing the sensitivity of the parameters involved estimating the effect at a 10x increase and 10x decrease of each of the constant parameters. . 69

LIST OF FIGURES

Figure Page

2.1 This shows the results of 500,000 simulations of the model at baseline parameter values and with all simulations shown. The overdispersion parameter was set $\phi_\alpha = 5$. 43

2.2 This shows the results of 500,000 simulations of the model at baseline parameters and with $\phi_\alpha = 5$, with only those simulations in which 5 or more secondary infections occurred. 44

2.3 This shows the results of 500,000 simulated epidemics for each $\phi_\alpha = 2$ through $\phi_\alpha = 7$. The vertical axis is the fraction of simulated epidemics which were smaller than 5 secondary infections, including those which resulted in no secondary infections. 45

2.4 This shows the results of 500,000 simulated epidemics for each $\phi_\alpha = 2$ through $\phi_\alpha = 7$. The vertical axis is the fraction of simulated epidemics which greater than 100 secondary infections. 47

3.1 As the percent of active infections increases, the average size of the outbreak at 100 years decreases. Each point on this graph is the average of 300 simulations of the model with the same set of parameters. 67

Chapter 1

INTRODUCTION

This dissertation connects aspects of social networks to the transmission of infectious disease in two projects. The first is an examination of the effects that superspreading can have on subcritical epidemics; superspreading is the phenomenon wherein a small number of infected individuals account for a disproportionate number of secondary cases of an infectious disease. Subcritical epidemics are those which, intuitively, should self-limit and become extinct (see text for more formal definitions of these concepts). Particularly, using Severe Acute Respiratory Syndrome (SARS) as an example, subcritical epidemics are shown to be potentially as large as supercritical epidemics. The second project examines the evolution of tuberculosis (TB) in hunter-gatherer populations. This project demonstrates that the structure of a social network can alter the evolutionary pressures on an infectious disease, potentially providing insight into the timing of the spillover event bringing TB to the human population.

The first project, modeling subcritical epidemics which exhibit superspreading, uses a stochastic jump process model. The resulting simulated epidemics are similar in size, in terms of total case count by city, to the reported case counts in the SARS epidemic. While not arguing that SARS was indeed subcritical, this example of how the network dynamic of superspreading can give rise to potentially large epidemics.

The second project, on the evolution of TB, simulates the spread of TB in a hunter-gatherer social network. The TB persists for long periods of time, and is shown to have an advantage in the specific social network if it has a lengthy latent period, rather than always becoming infectious immediately. Although this has been

suggested qualitatively (Gagneux, 2012), this confirms this qualitative consideration with a realistic model. Since the long latent period of TB would primarily be an advantage in a hunter-gatherer social network, and is observed in modern TB strains, this suggests a lengthy period of coevolution with humans and a temporally distant introduction to the human population.

Both of these projects use mathematical models and tools of mathematical epidemiology to advance anthropological questions. Both demonstrate the importance of social networks to epidemiological inquiries and the importance of anthropological approaches to epidemiology.

1.1 Research Questions

This is a dissertation in two parts, each exploring the effects of contact network structure on the transmission dynamics of infectious pathogens. Chapter 2 is an investigation into the consequences of human variation, of a particular type, on infectious disease transmission using a stochastic process model. The model developed in chapter 2 addresses the question of whether the social network phenomenon of superspreading can give rise to large, yet subcritical epidemics.

Chapter 3 investigates the evolution of *Mycobacterium tuberculosis* infections in a hunter-gatherer community. The question addressed is the origin of TB in humans. Specifically, if TB has a long coevolutionary history, the transmission parameters may be particularly well-suited to transmission in hunter-gatherers. Chapter 3 addresses the question of whether TB's current transmission pattern may be the result of a long coevolution with human hunter-gatherers.

Each chapter examines a different aspect of the impact of contact networks on infectious pathogen transmission. The model developed in chapter 2 does not explicitly model networks, which is a simpler approach, but which attempts to replicate

network-like behavior through the incorporation of heterogeneity in transmission rate. The model developed in chapter 3 explicitly is a network model, using a metapopulation structure. There is a great deal of interest from both anthropological and epidemiological researchers in the impact of networks on infectious pathogen transmission (Newman, 2002; Morris, 1993).

1.2 Background

1.2.1 Connection to Medical Anthropology

The interface between epidemiology and anthropology is well recognized. Investigations into how culture influences human health, how health and disease are defined and characterized as part of a cultural context, and how both are shaped by the environment are all broadly included in the field of medical anthropology (McElroy and Townsend, 1996). Of particular interest is the way in which wellness and disease are impacted by social networks, and how social networks impact the transmission and evolution of pathogens. The interaction of cultural norms and disease can be quite complex; expectations for interactions between individuals, in reference to age, sex, or other factors, can enhance or inhibit disease transmission, and add evolutionary pressures on to the pathogen.

Examples of diseases altering social norms abound in the anthropological literature (Lindenbaum, 1979; Herring and Swedlund, 2010). The alteration of social norms, consequently alters the contacts between individuals which may result in the transmission of infectious disease. Disease stigma and its concurrent alterations of social contacts, can result from long endemic disease (Weiss, 2008) or recent epidemic diseases (Bond and Nyblade, 2006). Changes in behavior around disease are well documented among psychologists and sociologists (Strong, 1990; Schaller and Park,

2011); changes include normally nurturing relationships becoming distant and new social connections may form to combat the disease. This results from both individuals with illness changing their behavior, as well as changes in behavior among those who would normally be involved in the ill individual's care.

These changes to an ill person's social network can have consequences for the transmission of infectious disease. Social stigma related to disease can be a barrier to seeking treatment (Weiss, 2008). Treatment itself is often carried out by professionals, who themselves form a different social network. Hospitals, or other centers of care, can become centers for the spread of infectious disease (Khan *et al.*, 2017; Hsieh *et al.*, 2004). Cultural expectations regarding behavior surrounding illness can greatly impact the social network through which an individual can transmit disease or become infected. The history of understanding how social networks can alter infectious disease transmission are explored in section 1.2.2; nosocomial infections play an important role in the description of the superspreading of SARS in section 2.1.2.

Section 1.2.3 below goes into greater depth on the impact changes in social networks can have for the emergence of infectious disease. Two examples are explored: the emergence of HIV, and the evolution of drug resistant *Staphylococcus aureus* in hospitals. The changes to social networks induced changes to the pressures on these infectious diseases; HIV was able to become a global pandemic due to the changes in the social network associated with colonization of Africa (Sharp and Hahn, 2010), and *S. aureus* has evolved resistance to antibiotics used in treatments in and out of hospitals (Vysakh and Jeya, 2013).

Cultural Epidemiology

The classic investigation of Kuru, Lindenbaum (1979), relied heavily on epidemiological data to understand both the epidemic and the Fore peoples' worldview in which

the epidemic resided. In addition to understanding the meaning of health and disease states, anthropology can identify cultural practices that give rise to epidemics. Another study in Papua New Guinea, Maddocks (2013) illustrates this, showing how cultural practices could give rise, independently, to helminth infections, foot lacerations, and illness associated with a sedentary lifestyle. Studies such as these bring up what Trostle and Sommerfeld (1996) called "cultural epidemiology."

Cultural epidemiology, or the broadly similar proposed subfield within epidemiology of "social epidemiology" has been controversial. Perhaps most notably, Zielhuis and Kiemeney (2001) argued social epidemiology (considered as a subfield of epidemiology concerned with social determinants of health and disease), was not epidemiology at all since it considered factors other than human biology in explaining health and disease. In the same issue of the International Journal of Epidemiology, five responses to this claim were also published with opinions ranging from social epidemiology being necessary for a fuller understanding of health, to various attempts at refinement and compromise. Much of this work has been done under the rubric of "social determinants of health," which examines the "upstream" causes of health, disease, and disparity in health outcomes (Braveman *et al.*, 2011).

Meanwhile, other authors have developed anthropological methodologies aimed at uncovering cultural meanings and definitions of health and disease (Weiss, 1997, 2001). Such work has been applied to both infectious and noninfectious disease. In one example, Sundaram *et al.* (2014) examines how viral influenza is understood in Pune, Maharashtra, India, and how that understanding shapes the epidemiology of the infection. Similarly, Weiss *et al.* (2008) examined tuberculosis in several locations, characterizing the perceptions of the infection and their consequences. Although not universal, a sizable portion of the literature investigating cultural epidemiology has focused on disease stigma (Weiss, 2008).

Many authors have pointed out the natural overlap between questions in medical anthropology and epidemiology (Trostle and Sommerfeld, 1996; DiGiacomo, 1999; Inhorn, 1995; Ashan, 2016). However most authors since Trostle and Sommerfeld (1996) have described this overlap as under-utilized. Trostle and Sommerfeld (1996) identifies several areas which may be fruitful for anthropological and epidemiological collaboration: disease and stress, stratification and inequity, identifying culturally relevant social variables, understanding the context of risk, classifying illness, improving epidemiological methodologies, using disease as a metaphor, and differentiating illness at the population versus individual levels.

These areas have received some attention in both anthropology and epidemiology. Din-Dzietham *et al.* (2004) exemplifies the importance of understanding the perception of stress on health with a study of African-Americans in Atlanta, Georgia, United States. The study found the experience of racial discrimination at work among African-Americans, when the discriminatory actions were undertaken by non-African-Americans, resulted in significantly higher rates of hypertension. However, when compared to discrimination perpetrated by other African-Americans, no such increase in hypertension was found. The study built on earlier anthropological work on stress and hypertension (eg Dressler (1984)), and touches on two areas of collaboration noted in Trostle and Sommerfeld (1996).

Epidemiologists often classify factors related to the incidence of a disease by the risk that the presence of such a factor brings for the onset of the disease. These risk factors can be biological, however many are cultural (Braveman *et al.*, 2011) and thus need to be contextualized in a particular culture. Factors such as social status, wealth, ethnic group, or family connectedness are intrinsically cultural variables and thus it is fruitful to understand these factors in the context of the culture in which they reside.

A particularly clear example of a cultural category which has become a unique descriptor of health comes from the United States. In the biomedical literature, beginning in the 1950s, two personality categories were distinguished: type A and type B, describing a risk factor for heart disease due to personality. Although it was not known publicly at the time, research into this distinction was funded by tobacco firms, possibly in an attempt to shift blame for coronary artery disease away from their products (Petticrew *et al.*, 2012). So these uniquely American cultural descriptors, type A and B personalities, became commonly identified and described as risk factors related to coronary artery disease (Riska, 2000). Similar examples are likely to be found in nearly every culture.

Finally, Trostle and Sommerfeld (1996) identifies ways in which anthropology can improve epidemiological practice. Data collection in epidemiology often relies on informants and their description of health, disease, or symptoms. These descriptions are rooted in a taxonomy of disease and an understanding of symptoms which may be unfamiliar; using anthropologists to assist in the unpacking of these descriptions can be a fruitful collaboration. Anthropologists may even be able to identify interview questions that yield poor results. Ashan (2016) gives the example of an epidemiological investigation of HIV among intravenous drug users. Traditional epidemiological survey instruments identified high rates of bleach being used to rinse injectors between use; anthropological investigation, however, found that bleach being supplied for this purpose was instead being used to launder clothing. The anthropological approach of participant observation identified an epidemiologically relevant fact, which traditional epidemiological survey instruments did not uncover.

1.2.2 Social Networks and Disease Transmission

Humans are fundamentally social creatures, forming connections with other people for a variety of purposes, including kin and non-kin (Bott, 1971). The broad history of the concept of a social network is reviewed in Mitchell (1974), which could be assumed theoretically either in a static or dynamic way. The importance of these connections to the transmission of infectious disease has not gone unnoticed in sociology and anthropology (Morris, 1993). Broadly there are two areas where anthropological investigations into social networks have intersected the study of infectious disease transmission. The first is looking at infectious disease transmission on a social network as a metaphor for the transmission of ideas (Blackmore, 2008; Inhorn and Brown, 1990; Cavalli-Sforza and Feldman, 1981). The second is looking at the consequences of changes to some aspect of social network structure on the transmission of a specific infectious disease (Klovdahl *et al.*, 1994).

It is important to recognize from the outset that in reference to infectious disease transmission, there are potentially as many or more social networks as there are infectious diseases; the unique characteristics of the transmission of specific infectious pathogens characterizes what constitutes contact between individuals (Jacobsen, 2008). This dissertation uses the terms "social network" and "contact network" to mean the collection of probabilities of transmitting a specific infectious disease between individuals in a population. In all cases discussed herein, the size of the population is so large that these transmission probabilities are described statistically. These transmission probabilities are a function of how the strength of contact between individuals in the network change over time.

Epidemics as a Metaphor

Anthropology, broadly, is sometimes described as the study of culture (Kroeber, 1988), and it may be the transmission of culture which most clearly differentiates humans from other species (Hill *et al.*, 2009). Although the evolution of humans has been discussed since Darwin, much of anthropological work in the 20th century centered on treating humans as a quintessentially unique species (Durkheim, 2000). Although there were many attempts to put anthropology onto a biological footing (McGee and Warms, 2000), arguably the most successful was the introduction of sociobiology by E. O. Wilson in 1975 (Wilson, 2000).

Sociobiology as applied to humans broadly split into three areas of study: human behavioral ecology, evolutionary psychology, and dual inheritance theory (Smith, 1999). The third of these, dual inheritance theory, describes two routes for the transmission of human behavior: cultural transmission and biological inheritance (Cronk *et al.*, 2000; Boyd and Richerson, 1988). The descriptions for the transmission of culture, both vertically from one generation to the next, and horizontally to those in the same generation, borrowed extensively from the language and approaches found in mathematical epidemiology (Cavalli-Sforza and Feldman, 1981).

Perhaps nowhere is this borrowing clearer than in memetics (Blackmore, 2001). A meme, as defined by Richard Dawkins, analogously to a gene being a unit of biological inheritance, a meme as "a unit of cultural inheritance, or imitation" (Dawkins, 1989). Important in this understanding of cultural transmission is the meme as a self-replicator, using humans as its host, for better or worse in terms of the humans' goals. For this reason, memes are sometimes analogized to infectious agents moving between humans. For an overview of this view and its critics, see Aunger (2000).

Social Networks' Impact on Transmission Dynamics

This dissertation does not take infectious disease epidemics as a metaphor for social network transmission. Instead, it examines the impact that human social networks of the type found in anthropological data, or specific aspects of them, have on infectious disease transmission dynamics. This type of analysis is not new in anthropology. The importance of social networks to disease transmission was first identified in the epidemiological literature, related to sexually transmitted infections (Hethcote *et al.*, 1982; Yorke *et al.*, 1978), and drew attention from anthropologists in the 1980s (Potterat *et al.*, 1985; Klovdahl, 1985).

Sexually transmitted infections remain a rich area in the overlap between infectious disease dynamics and social structure. HIV transmission has been examined in terms of social network factors leading to the increase or decrease in prevalence (Klovdahl, 1985; Klovdahl *et al.*, 1994; Rothenberg *et al.*, 1998), sexual partner patterns and the influence of migration (Khanna *et al.*, 2014), and in terms of population heterogeneity (Williams and Dye, 2018). Attempts to measure the spread of sexually transmitted infections have explicitly measured contact networks (Bearman *et al.*, 2004), as well as characterizing the network in terms of a statistical abstraction of networks (Eames and Keeling, 2002).

The examination of the impact of changing aspects of social networks on disease transmission dynamics has extended beyond sexually transmitted infections, however. Klovdahl *et al.* (2001) analyze the contact network on which tuberculosis spread, after an outbreak, in a manner similar to contact tracing from epidemiological research (Shaban *et al.*, 2008; Armbruster and Brandeau, 2007), with the goal of identifying clusters on the network of particular risk. High resolution spatial data has been used by Salathe *et al.* (2010) to measure close proximity interactions which may

be conducive to the spread of diseases which are spread through aerosolized droplets. While the data were collected absent an epidemic, the measurement of close proximity interactions allowed them to examine the effects of network structure on the spread of diseases through simulation. Zelner *et al.* (2012) was one of only a few examples where the protective role of social networks was explicitly examined in relation to infectious disease, considering how community cooperation and interdependence can lower, rather than raise, rates of infectious disease.

A very rich area of interest is the description of mixing heterogeneity in human populations as a factor in altering infectious disease transmission dynamics (Eames and Keeling, 2002). Superspreading is the phenomenon of a small number of individuals being responsible for disproportionately many secondary cases (Stein, 2011). Like many terms, it is the term "superspreader" or "superspreading event" describes a gradient of values, not a difference in kind between superspreading and non-superspreading individuals or events. The two most commonly cited cut-off values are the top 20% of individuals, in terms of the number of secondary cases caused (Woolhouse *et al.*, 1997), or the top 1% (Lloyd-Smith *et al.*, 2005). Both of these values, however, are largely arbitrary.

Superspreading on contact networks has been suggested or described in the spread of ebola (Lau *et al.*, 2017), tuberculosis (Ypma *et al.*, 2013), SARS (Chen and Leo, 2006), Middle East Respiratory Syndrome (MERS) (Kucharski and Althaus, 2015; Oh *et al.*, 2015), and others (Galvani and May, 2005). Superspreading has also been investigated in non-human animal diseases, including West Nile Virus in birds and brucellosis in water buffalo (Stein, 2011), however given humans' social nature, much of the work has focused on human to human transmitted diseases.

Some work has attempted to identify common features of either individuals (superspreaders) or time-and-place factors (superspreading events) which lead to this

sort of disproportionate transmission. Broadly, the superspreading phenomenon can result from individual variation in the number of social contacts an infected individual in a network has, or from individual variation in the rate of transmission, due to physiological, immunological, or behavioral factors, an infected individual in a network has (Lloyd-Smith *et al.*, 2005). However attempts to identify general patterns which result in a priori identification of superspreaders or circumstances which lead to superspreading events is in its formative phase. Immunological factors have been identified as relevant in some diseases (Gopinath *et al.*, 2014), properties of the disease itself has also been considered (Galvani and May, 2005). Behavioral risk factors leading to superspreading events have also been identified (Yu *et al.*, 2007). However no general *a prior* method for identifying superspreading yet exists (Stein, 2011).

Consideration of the effects of superspreading, as a social network phenomenon affecting infectious disease transmission, has largely focused on large epidemics, either retrospectively examining known epidemics (Yu *et al.*, 2007) or simulating prospectively epidemics which might occur using estimates of superspreading (Fujie and Odagaki, 2007). Missing from the literature is an examination of the effect that superspreading could have on epidemics which, absent superspreading, would be small and self-limiting. This project addresses that gap.

1.2.3 Biological Evolution of Humans and Infectious Disease

The biological evolution of any species is ultimately driven by factors affecting reproductive success, often mediated through simple survival (Barton, 2007). Examination of the Ache, gathering data from informants about pre-contact mortality in this hunter-gather group, finds that disease was a common, though by no means the largest, cause of death (Hill and Hurtado, 1996). This is suggestive, though certainly not definitive, that disease was a significant factor in human mortality in the past,

and so shaped human evolution. Moreover, occasional epidemics, such as respiratory infections in the post-contact Ache, and which have been described in many other groups (Milner, 1980), may substantially increase the effect that disease had on human mortality in the past.

The human immune system co-evolved with parasites and pathogens, each shaping and altering the evolutionary trajectory of the other (Nesse and Williams, 1996). Pathogens nearly always have a much shorter generation time than humans, and so biologically can evolutionarily out-pace the human immune system. However pathogens do not necessarily evolve towards increasing harm of their host; indeed, in many cases, this is counter-productive from the perspective of the pathogen. Pathogens which kill their hosts are usually selected against. As a result, humans have evolved in constant contact with a wide range of pathogens, in many cases evolving toward commensal tolerance, and in a few cases even coming to benefit from inhabitation by these organisms (Gluckman *et al.*, 2009).

The unprecedented industrialization and explosion in human population over the past 200 years has radically transformed the landscape of human-pathogen coevolution. Among these changes has been alteration to human social networks, to include rapid, long-distance migration and changes to domestic animal practices, both of which have been implicated in the global emergence of novel infectious diseases (Hassell *et al.*, 2017). This has been referred to as the "third epidemiological transition" (Herring and Swedlund, 2010).

This phenomenon can be illustrated with an account of emergence of HIV. There are two modern circulating strains of HIV, denoted HIV 1 and HIV2, both of which entered the human population through contact with animals, most probably blood-blood contact while hunting or butchering (Sharp and Hahn, 2010). Both strains had separate origins from different primate species, and both have become worldwide,

although HIV-2 remains at much lower prevalence outside of Africa (Adjorlolo *et al.*, 1993). The emergence and global spread of two strains of a virus in a few decades, due to a spillover event–hunting–which humans have participated in for the entirety of human existence, is suggestive of a change in the human social network which facilitated the spread of the pathogen. In all probability, such spillover events occurred regularly throughout human history, however only recently, due to changes in the human social network, became capable of spreading globally.

HIV typifies the rapidity of the changes to the human social network on which infectious pathogens may transmit. Infectious diseases which previously were isolated and went extinct in humans may now, due to changes in the human social network, population density, or behaviors associated with industrialization, spillover into human populations and spread (Taylor *et al.*, 2001; Daszak *et al.*, 2000).

Another example of such evolution in response to a changing social network is the evolution of antibiotic resistance in *S. aureus*. Antibiotic resistant *S. aureus* is most commonly methicillin-resistant *Staphylococcus aureus* (MRSA), which is generally resistant to both the penicillin family of antibiotics, as well as methicillin and its derivatives (Otto, 2013). Resistance to these antibiotics emerged shortly after their introduction, initially in hospitals (Chambers, 2001). MRSA has since spread globally in and out of hospitals.

Didelot *et al.* (2016) has demonstrated that the evolution of drug resistance in bacterial species can originate within a single host. Antibiotics strongly select for antibiotic resistance through the differential removal of susceptible pathogens. High rates of antibiotic use, therefore, can be expected to result in higher rates of antibiotic resistance. Further, interactions between MRSA and antibiotic-susceptible *S. aureus*, as might be expected to occur at a higher rate in a health-care setting, can result

in horizontal gene transfer of the resistance-related genes to the susceptible strain (Lindsay, 2014).

Hospitals are both frequent users of antibiotic therapies and provide a network on which disease can transmit. Individuals who are ill, potentially with weakened immune systems, are disproportionately found on the contact network of a hospital (Otto, 2013). Further, with the prospect of horizontal gene transfer is heightened by individuals with antibiotic-susceptible *S. aureus* infections potentially interacting with individuals with MRSA. Even within MRSA, strains vary in their antibiotic susceptibility. Hospitals may become centers for the creation of more strongly resistant MRSA strains through horizontal gene transfer. These factors together result in higher rates of MRSA, and more strongly drug resistant MRSA, in hospitals than outside (Vysakh and Jeya, 2013). As individuals seek treatment, their changing social network associated with treatment exposes them to changing risks of infectious disease and alters the evolutionary trajectory of the pathogens.

Superspreading and Human-Pathogen Coevolution

Very little has been written regarding human-pathogen coevolution resulting in superspreading as a mechanism for pathogen emergence. As noted previously, there are broadly two causes of superspreading as a phenomenon on disease transmission networks: disproportionately many contacts for a particular infected individual or small collection of individuals, or a change in the immune response, pathogen, or behavior which allows higher transmission rates than in the rest of the population.

Given that, one evolutionary strategy may be for a pathogen to evolve to be asymptomatic, at least in some individuals. Asymptomatic carriers have been frequently identified as superspreaders in typhoid (Yang *et al.*, 2018) and polio (Mehndiratta *et al.*, 2014). It is reasonable to speculate that such asymptomatic carriers are

the result of human-pathogen coevolution to create human superspreaders.

This project, in chapter 3, demonstrates the fact that disease transmission dynamics which might favor the pathogen in one human social network could be disadvantageous in another human social network. This fact can use used to identify the likely social network on which a pathogen evolved, and thus estimate the time it entered the human population.

1.2.4 Mathematical Modeling

The use of mathematical models in the biological sciences has a long history (Kingsland, 1995). Quantitative models for outbreaks of infectious disease have taken a wide variety of forms, from the very data-intensive agent based models to very simple models (Brauer *et al.*, 2008). First order autonomous ordinary differential equations (ODEs) has an important place in mathematical biology. They are widely used in biology (Murray, 2002), especially ecology (Kot, 2001) and infectious disease epidemiology (Anderson and May, 2010). Deterministic solutions provide a useful starting point, however much of biology involves quantifying behavior that has inherent and natural stochasticity. Phenomena such as transmission of infection or birth/death processes are fundamentally probabilistic in nature.

The need for probabilistic models can be seen in many fields, but particularly in epidemiology. Ordinary differential equations (ODEs)-based models of infectious disease capture deterministic and homogeneous mixing patterns, which may have limited biological implications. There are many ways in which this limitation can be overcome, however each has a shortcoming. The approach that is proposed here likewise has limitations.

Modeling diseases on different contact networks can, therefore, be challenging in two ways. The first is the development of a mathematical model which reflects and

can, at least potentially, capture the disease dynamics induced by the structure of the contact network. The second is that information about the contact network may not always be available or adequately fine-resolution to make meaningful predictions, or even explanations after the fact. Collection of high resolution data on contact networks can be both expensive and invasive.

This project develops a mathematical model which can capture disease dynamics caused by heterogeneity in the contacts, while simultaneously not requiring detailed information about the network. The model developed in chapter 2 is not explicitly a network model, as the network structure for most infectious diseases is difficult to determine. Instead, the approach uses a stochastic approach to simulate the super-spreading phenomenon without explicitly modeling contacts and individuals.

Incorporation of Heterogeneity in Models

There are, in principle, many ways to incorporate heterogeneity of the population in epidemiological models. Several common approaches include partial differential equations (Anderson and May, 2010), adding demographical and epidemiological structure to compartmental models, branching process models, and individual-based models (Getz *et al.*, 2006). Gustafsson (2000) proposed another method, a Poisson-process based model, which forms the basis for this project. The Poisson method is described below, after a survey of the other existing methods.

Compartmental models have played an important role in the history of epidemiological modeling (Brauer, 2017). Naturally, early attempts to incorporate heterogeneity in models utilized these models. Making compartments for an epidemiological state such as susceptible, but which differed by age, sex, or other characteristic, allowed the modeling of these compartments as different from each other. That is, splitting the susceptible compartment of an SIR model into male and female suscep-

tibles allowed males and females to be modeled as having different susceptibilities.

This additional-compartment approach to modeling heterogeneity has been used to capture superspreading events or superspreading individuals (Hethcote *et al.*, 1982; Mummert, 2011), although it is primarily used to model variation in pathogen transmission dynamics which vary according to an identifiable characteristic such as age or sex.

Some such characteristics, such as age, are continuous in their variation. This suggests the use of a partial differential equation (for a pedagogical development of this ideas, see Kot (2001)). A partial differential equation can model all possible ages and use a function to describe how the epidemiological parameters vary as a function of age, for example. Partial differential equations are a standard tool in mathematical modeling of epidemics (Anderson and May, 2010). A curious historical note is that McKendrick (1925) actually developed the essentials of this approach prior to the better-known ordinary differential equation approach (Kermack and McKendrick, 1927), however the approach was primarily popularized by von Foerester (Keyfitz and Keyfitz, 1997).

A third approach to incorporating heterogeneity in epidemiological models is with branching process based models (Jacob, 2010). The approach of a branching process model, which is a realization of a random graph model, is distinct from compartmental models. Branching processes, most frequently Galton-Watson branching processes, are modeled as a series of events, each of which gives rise to a random number of secondary events. For instance, an infected individual gives rise to n_1 secondary infections, following some distribution, and each of those give rise to additional infections, also following the same distribution, and so forth (Harris, 1963). Branching process models are typically what are used to model subcritical epidemics

(see below) (Becker, 1977; Farrington and Grant, 1999; Blumberg and Lloyd-Smith, 2013a).

A fourth approach to incorporating heterogeneity in epidemiological models is individually-based models. There is no single definition for what constitutes an individually-based model (or agent based model), however the premise is that each individual may potentially be modeled to have unique characteristics related to the transmission of disease. These models can be large and difficult to parameterize, but offer a tremendous flexibility in the ability of capture heterogeneity (Willem *et al.*, 2017).

The final approach to incorporating heterogeneity is less often used in mathematical epidemiology. First proposed by Gustafsson (2000), the Poisson method is a stochastic approach using a Poisson process describing a jump process, the mean value of which is given by the corresponding ODE. Gustafsson and Sternad (2007) has proposed the Poisson method as a general method for models of an intermediate complexity, at once providing the flexibility of compartmental modeling while using stochasticity as a means of capturing different solution trajectories due to heterogeneity in the population. The method itself is best illustrated in one dimension, although multidimensional versions are important. Given the differential equation describing state variable x at time t,

$$\frac{dx}{dt} = f(x,t), \tag{1.1}$$

the Euler method for simulating this involves discretizing the model into time intervals Δt. The value of the state space variables at time t, denoted x_t can therefore be approximated by equation 1.2.

$$x_{t+1} = x_t + f(x_t,t)\Delta t \tag{1.2}$$

The Poisson method developed in Gustafsson (2000) takes the change in the state space variable, $f(x_t, t)\Delta t$, and simulates it as a Poisson random variable, with rate (and therefore mean) set to be $f(x_t, t)\Delta t$. The project in chapter 2 extends this approach by using, instead of a Poisson, a negative binomial random variable. The mean of this random variable is again $f(x_t, t)\Delta t$, and the overdispersion is set to be a new parameter ϕ.

Uses of the Negative Binomial

A central mathematical method in chapter 2 is the replacement of the Poisson distribution in an existing method Gustafsson (2000) with the negative binomial distribution in order to capture superspreading events. The negative binomial is a widely-used statistical distribution for non-negative discrete random variables (Casella and Berger, 2002). It has a lengthy history into the earliest days of probability; early descriptions of the negative binomial date to the 17th Century, due to Blaise Pascal and Pierre de Fermat, however the precise formulation was not given until 1714 (Bartko, 1962). Nevertheless, this early history gives rise to a second name for the negative binomial: the Pascal distribution. The early applications involved the formulation of the negative binomial as the length of a sequence of independent, identically distributed Bernoulli trials before a specified number of failures has been reached. A Bernoulli trial is one in which the outcome is either 0 or 1 (Casella and Berger, 2002).

The first connection between the Poisson and the negative binomial distributions is usually attributed to Student (1907), although later works such as Greenwood and Yule (1920) made explicit the idea that the negative binomial is a "generalized Poisson." Feller (1943) included the negative binomial in a class of "contagious" distributions, so named because of their ability to capture distributions resulting from processes in which "each 'favorable' event increases (or decreases) the probability of

favorable future events." He would go on in that article to describe how heterogeneity in rates in a process could give rise to the same distributions. The terminological connection to infectious disease was not an accident; the negative binomial was a widely-used distribution for biological phenomena resulting from either heterogeneity in rate, or the previously described autocatalysis (Gurland, 1959).

In this vein of "contagious" distributions, the negative binomial has been used to model phenomena as diverse as counting bacteria in a microscope field (Jones *et al.*, 1948), dental carries in children (Gurland, 1959), or parasite loads in people (Grenfell *et al.*, 1990). The negative binomial has also been widely used in generalized linear models, including in medicine and epidemiology, to model variables which arise from Poisson processes, but which have higher-than-expected variance (Dobson, 2002).

This project is situated in the same meaning of the negative binomial as an "overdispersed Poisson." That is, incidence is expected to be an approximately-Poisson process, however with heterogeneity in the transmission rate or connectedness of individuals. Consistent with these previous uses, a negative binomial is used to model incidence where transmission probabilities.

1.2.5 Subcritical Epidemics and R_0

History of R_0

The basic reproductive number R_0 is one of the most recognized parameters in mathematical epidemiology. The intuition behind R_0 is it is the average number of secondary cases caused by an infected individual during the infectious period in a totally susceptible population (Anderson and May, 2010). The demographic parallel, which predated the development of the concept in epidemiology, was the growth rate of the population (Cushing and Diekmann, 2016). However this connection was slow to be

drawn (Heesterbeek, 2002).

Just as demography and population biology models often centrally place the idea of population growth rate, infectious disease epidemiology models often use R_0 as a centrally important descriptive parameter (or combination of parameters) to describe a first order approximation of the population trajectory (Anderson and May, 2010). If $R_0 > 1$, the population of infected individuals will grow (at least initially), while if $R_0 < 1$, the population will shrink to zero. For this reason, estimating R_0 has become a centrally important task in the prediction of infectious disease outbreaks.

There are simple methods for estimating R_0 from data of differing types. If transmission probability (probability of transmission of an infection given a contact between an infectious and a susceptible individual) β, contact rates (average number of contacts between an infected individual and susceptible individuals) n, and the duration of infectious period τ are known experimentally for a disease, then

$$R_0 = n\beta\tau. \tag{1.3}$$

These parameters, however, are practically not normally known. As described elsewhere in this project, the precise definition of what constitutes "contact" is critical to this formulation, but is challenging to estimate. Nevertheless, it is useful as an intuition into the behavior in more complex estimation.

Branching process models also have been used to estimate R_0, described below. However a very common approach is to develop a simple compartmental model for the disease, fit the model, and use the resulting fit parameters to estimate R_0 (Chowell and Brauer, 2009). The precise formula for R_0 is dependent on the model chosen, however many models have well-described formulations.

There are two critically important considerations in the estimation of R_0 which are sometimes overlooked, although both have been recognized. The first is that

R_0 is dependent on both the biology of the disease and the properties of the host. Changes to human contact networks can wildly alter the R_0 for a particular disease; despite the attraction and simplicity, tables of "diseases and their R_0 values" leave off the important contribution of the human social network. Second is the critical importance of the model choice; relatively subtle changes to a model can yield very different values of R_0.

Subcritical Epidemics

This project uses the term "subcritical epidemics" to describe infectious disease outbreaks for which the basic reproductive number R_0 is less than 1. These outbreaks have been previously described as "stuttering outbreaks" or simply "mortal" and have generally been considered using branching process models (Bailey, 1957; Becker, 1974; Kimmel and Axelrod, 2015; Farrington and Grant, 1999). However modeled, $R_0 < 1$ describes an outbreak which generally dies out.

Branching process models are within the scope of random graph models and most commonly are Galton-Watson branching process models. They describe a population X_t at a sequence of discrete times t. The size of the population at time $t + 1$, conditioned on X_t being known, is the sum of X_t independent, identically distributed random variables Z_t. The expected value of Z_t, in this formulation, is R_0. Although $R_0 < 1$ guarantees termination (also called extinction) of the branching process, there is a positive probability that branching processes with $R_0 > 1$ also terminate. Thus a goal in the analysis of such models is the determination of R_0 and in particular, estimating the probability that $R_0 < 1$ (Guttorp and Perlman, 2015).

In epidemiology, there is a classic result (Becker, 1974) that is seen as a first order estimate for the size of subcritical epidemic outbreaks (De Serres *et al.*, 2000). Under the assumption of a geometric distribution for the number of connections at a node,

the average size of the outbreak N when $R_0 < 1$ is given in equation 1.4

$$N = \frac{1}{1 - R_0} \tag{1.4}$$

This suggests, at least if the distribution is geometric, that large subcritical epidemics only occur if R_0 is very close to 1 or exceedingly rarely.

Other distributions besides the geometric are possible, however. In the context of epidemiology, Farrington and Grant (1999) examined in branching processes include the Poisson and Bernoulli, which are analytically more difficult to describe. Nevertheless these can also be fit to outbreaks which terminate. Superspreading has been incorporated into branching process models through the use of a negative binomial distribution for the number of offspring per individual in each generation (Garske and Rhodes, 2008). This is similar in principle to the present project, although the present project does not use a branching process model. Much of the justification for the model choice, however, is parallel to Garske and Rhodes (2008). The effect of a negative binomial branching process model in an epidemic was surprisingly large outbreaks.

Very little work on subcritical epidemics has been done outside of branching process models. Branching process models are particularly applicable to subcritical epidemics since they assume, as is assumed here, an infinite population and easily extend from an individual process to a population. However they are limited in their extensions and do not have the history and development found in ODE models. As such, a method for easily incorporating superspreading into ODE models is the need met by this project.

Term	Definition
R_0	The basic reproductive number describes the number of secondary infections caused by an infected individual, on average, in a totally susceptible population. The value of R_0 is always non-negative.
Supercritical	Supercritical describes an epidemic of infectious disease in which each infected individual causes more than one secondary infections. It can be described as having $R_0 > 1$.
Subcritical	Subcritical describes an epidemic of infectious disease in which each infected individual causes fewer than one secondary infections, on average. It can be described as having $R_0 < 1$. These are sometimes referred to as "stuttering" epidemics.
Spillover event	An animal to human transmission event for an infectious disease.
Superspreading event	An instance where an infected individual causes a large number of secondary cases.
Parameterize	Determine the key rates and constants applicable to a particular model.

Table 1.1: Definitions of mathematical epidemiology terms used in this paper.

1.2.6 Terminology

1.2.7 Subcritical Disease Transmission on Social Networks

The method used in chapter 2 to capture the network phenomenon of superspreading is epidemiological modeling (Anderson and May, 2010). Differential equations provide a useful framework for describing how prevalence changes through time by describing the reasons why prevalence changes, together with an initial condition. The central goal of chapter 2 is the impact which a specific type of social network heterogeneity has on infectious disease transmission. This goal fits into the framework of anthropological epidemiology identified in Trostle and Sommerfeld (1996). While other anthropologists have described other types of social networks on disease transmission (see review in section 1.2.2), this is the first examination of superspreading in an anthropological context on subcritical epidemics.

Subcritical epidemics, sometimes also described as "sputtering outbreaks," are distinguished from epidemics by the basic reproductive number R_0. The basic reproductive number is the number of new infections caused by an infected individual in a completely susceptible population (Diekmann *et al.*, 1990). If $R_0 > 1$, the outbreak will become an supercritical epidemic, generally infecting a substantial portion of the population; the intuition is straight-forward, if each infection causes more than one additional infection, then the infection will spread. However if $R_0 < 1$, the intuition is the disease will die out, but might persist for a few additional cases before doing so.

Networks can provide a contrast to this intuition about subcritical epidemics. A network may have structure which permits local transmission to a significant number of individuals, but considering the larger network, still have an $R_0 < 1$. The specific method was chosen to incorporate a lack of extensive knowledge of the likely contact

network. The research question being investigated is whether we can distinguish subcritical epidemics from supercritical epidemics. While this has been investigated before (see review in section 1.2.5), this is a novel description of how human variation impacts this assessment.

1.2.8 TB in Hunter-Gatherer Populations

There is an open question of the origin of Tuberculosis (TB) in humans. The behavior of a TB infection is also somewhat surprising: most new TB infections are not immediately infectious. Instead, they only become infectious years or decades later, if ever. This delay might be explained by the network structure of hunter-gatherers. The question of chapter 3 is whether the network structure of hunter-gatherers can explain the seemingly paradoxical behavior of TB infections.

To address this question, the project develops a model of TB infecting a simulated group of hunter-gatherers. The infection is introduced from a single active case, and run for 100 years. The final outbreak size is measured as the number of individuals infected at the end of the 100 year period. The network is the critical part of the model: the simulation is on a network of discrete 25 person bands with limited contact between bands. Birth and death processes are also modeled.

These simulations use realistic values of life history and TB parameters drawn from the literature, varying the percent of individuals who become active immediately after being infected. In TB, this value is low. The final outbreak size and the R_0 for different values of the percent active are calculated.

1.3 Brief Results

The two projects in this dissertation show several important results on the research questions outlined in section 1.1. On the question of subcritical epidemics

with superspreading, chapter 2 finds that as the degree of superspreading increases, so too does the size of the rare, major epidemics. While these remain rare, and most subcritical epidemics are immediately extinguished, the large epidemics which do occur are major. The question of the origin of TB in the human population is addressed in chapter 3. The model demonstrates that over a wide range of parameter values, TB can persist in the hunter-gatherer essentially indefinitely due in part to its long latent period. This long latent period would be a disadvantage in a homogeneously mixed population, suggesting that it evolved on a hunter-gatherer social network.

The idea of subcritical epidemics is certainly not new, however the project presented in chapter 2 is a demonstration of the complexity of the problem of distinguishing between supercritical and subcritical epidemics based on case count data. Since interventions are often aimed at reducing a supercritical epidemic into a subcritical epidemic Anderson and May (2010), this is not a purely academic problem. The results in this chapter demonstrate the importance of the anthropological consideration of network structure has to epidemiological practice.

Epidemiological methods and models can also be fruitfully applied to anthropological investigations, as is demonstrated in chapter 3. The important demonstration of the *Mycobacterium tuberculosis* pathogen's long-term coevolution with humans illustrates how epidemiological characteristics can provide insight into the anthropological question of the origins of TB in humans.

Both projects are demonstrations of the importance and interconnectedness of epidemiological and anthropological research. They can be considered as contributions to the literature at this interface.

Chapter 2

SUBCRITICAL EPIDEMICS AND SUPERSPREADING

2.1 Introduction

Human to human infectious disease transmission has been a topic of anthropological interest in a variety of contexts (Inhorn and Brown, 1990). Of particular interest is the impact that contact networks (sometimes referred to as social networks) have on the spread of disease (Richardson and Gorochowski, 2015; Nunn *et al.*, 2015; Khanna *et al.*, 2014). The spread of disease has been used as an analogy to phenomena such as the transmission of ideas or culture (Blackmore, 2008; Inhorn and Brown, 1990). Approaches to modeling the transmission of culture has borrowed from both the language and model structure of epidemiology (Cavalli-Sforza and Feldman, 1981). However this is not a project which takes a metaphorical approach.

In addition to the metaphorical approach, anthropologists have provided insight into disease transmission dynamics by showing the effect of cultural factors can have on these dynamics (Sattenspiel, 1990; McGrath, 1988). As with any eusocial organism, human social networks play an important role in determining pathogen transmission (Stroeymeyt *et al.*, 2018; Cremer *et al.*, 2007). Anthropologists and epidemiologists have looked at the influence of group size on pathogen transmission, in both non-humans (Davies *et al.*, 1991) and humans (Nunn *et al.*, 2015). Human sexual networks have been widely implicated as a determining factor for sexually transmitted disease prevalence (Khanna *et al.*, 2014; Yorke *et al.*, 1978; Bearman *et al.*, 2004; Day, 1994). This project builds on these by examining heterogeneity in contact networks, specifically the presence of superspreading events, as a factor which can increase the

size of certain epidemics.

Superspreading is a characterization of a feature of some infectious disease transmission networks in which one or a few individuals are responsible for a disproportionate number of secondary infections (Lloyd-Smith *et al.*, 2005). Descriptions of superspreading include both superspreading individuals and superspreading events, the latter emphasizing the importance of the confluence of both individual factors and their circumstances or connections with others. The causes and consequences of this superspreading network feature have been investigated in many infectious diseases (Lau *et al.*, 2017; Brown and Kelly, 2014; Salathe *et al.*, 2010). The present project examines changes in disease transmission dynamics created by superspreading events, using the example of SARS. Specifically, this project asks the question whether the network phenomenon of superspreading events can give rise to SARS-sized epidemics of disease, if the epidemics would, absent superspreading, be relatively small.

A classic epidemiological approach to predicting the future of an epidemic is the estimation of the basic reproductive number (Anderson and May, 2010; Cushing and Diekmann, 2016). The basic reproductive number, R_0, has the intuition of being the number of secondary infections caused by an infected individual, on average, in a completely susceptible population. Supercritical epidemics, those for which $R_0 > 1$, are generally thought to be large, infecting a significant portion of the population, while subcritical epidemics, those for which $R_0 < 1$, are usually thought of as small and limited (Anderson and May, 2010).

The 2002-2003 epidemic of Severe Acute Respiratory Syndrome (SARS) had a global reach, with more than 8000 people infected (WHO, 2003a). On the basis of the large size alone, it is frequently identified as likely supercritical(WHO, 2003a). Some estimates of R_0 for SARS supported that (WHO, 2003a). However some of the early modeling (Chowell *et al.*, 2004) found a best fit to the epidemic to be subcritical,

that is $R_0 < 1$, although with error bars in the estimate which included $R_0 > 1$. The possibility of SARS being subcritical has largely been ignored, as the total case count was far larger than what a subcritical epidemic process seems to which could give rise. This project investigates whether superspreading events, which are well documented in the SARS epidemic (Lloyd-Smith *et al.*, 2005), could result in subcritical epidemics which are of a similar or larger size to the SARS epidemic in cities.

Contact networks can dramatically change the transmission dynamics of infectious diseases (Keeling and Eames, 2005). Superspreading is an aspect of this phenomenon, characterized by a highly variable number of transmission events per infected individual. Conceptually, a superspreading event is a confluence of circumstances which allow a single infected individual to infect a large number of other individuals, while under most circumstances secondary cases are much rarer.

This project develops a simple model of subcritical epidemics which can incorporate the network-level effect of superspreading. Using this model and the transmission parameters from the SARS epidemic, it is shown that subcritical epidemics similar in size to the SARS epidemic are possible. While this is not a demonstration that SARS itself was a subcritical epidemic, it demonstrates that subcritical epidemics can be large, even global in reach due to the network effect of superspreading.

2.1.1 Severe Acute Respiratory Syndrome

SARS is a potentially fatal viral infection which is spread through human to human transmission, killing approximately 11% of those infected (WHO, 2003a). The only large outbreak of SARS started in 2002, ultimately infecting more than 8000 people before the outbreak ended in 2003. SARS has not re-emerged since, there is no known-effective treatment and no vaccine (Stockman *et al.*, 2006). The disease is genetically most similar to strains found in bats, leading most to believe SARS emerged as a

zoonosis from bats in Southern China (Li *et al.*, 2005). It exemplifies a class of pathogens which pose a significant threat: spillover events of previously undescribed infectious diseases, capable of rapid transmission (Daszak *et al.*, 2000).

The outbreak ultimately spread to 26 countries, causing 916 deaths (WHO, 2003a). The rapid spread of the disease was facilitated by the airborne droplet and indirect contact transmission. Air travel (Breugelmans *et al.*, 2004) and the movement of patients and infected staff during medical treatment spread the disease between and within communities. Other disease transmission routes have been investigated, including vector borne transmission and fecal-oral route, but little evidence supports these (Ng, 2003). Symptoms of SARS included viral pneumonia and symptoms of a respiratory tract infection, and diarrhea is also a feature in some infections (Wong and Yuen, 2005).

Responses to SARS included hospitalization and quarantine, travel restrictions, and attempts at treatment (WHO, 2003a; Hsieh *et al.*, 2007). There is a good deal of modeling evidence that indicates quarantine measures were effective at limiting the spread of SARS and ultimately eliminating the infection (Hsieh *et al.*, 2007). Treatment measures were generally ineffective (Stockman *et al.*, 2006) and some medical procedures may have increased the risk of transmission, especially to hospital staff (Stein, 2011).

The spillover event has never been conclusively identified, however initial cases were disproportionately employed in food preparation or production (WHO, 2003a). While the strain of SARS which infected humans most closely resembles strains found in bats, related viruses have been located in a wide range of other species (Li *et al.*, 2005).

SARS was chosen for this project given its propensity to spread through so-called "superspreader events," that is a single individual who is responsible for a large

number of secondary cases. This project is examining the hypothesis that subcritical epidemics, when superspreading is possible, can appear to be supercritical epidemics.

2.1.2 Networks and Superspreading Events

The structure of a contact network (variously called social network or simply network) can have a significant impact on the transmission of an infectious disease (Newman, 2002). Network structures are oftentimes assumed to be simple or homogeneous, with each individual in the network having the same probability of coming into contact with another individual (Newman, 2002; Anderson and May, 2010). For the most part, network structures are very complex and heterogeneous. The structure of the network on which a disease transmits can have important implications for the size, frequency, and other epidemiological characteristics of an outbreak (Castellano and Pastor-Satorras, 2010; Martin and Boland, 2018), as well as control measures such as vaccination (Kucharski *et al.*, 2016). One feature of networks which can give rise to unusual behavior is the superspreader phenomenon (Fujie and Odagaki, 2007).

Superspreading events, or superspreading individuals, are individuals who are responsible for a disproportionate number of secondary infections (Stein, 2011). This can be the result of individuals having more contacts in a network, or other factors such as increased transmissible (Lloyd-Smith *et al.*, 2005; Gopinath *et al.*, 2014). There is no specific cut-off which defines an individual as a superspreader or an event as a superspreading event. Woolhouse *et al.* (1997) argued that 80% of new infections were generally caused by 20% of individuals, while Chun (2016) arbitrarily defines the 99th percentile of individuals in terms of their individual R_0 (Lloyd-Smith *et al.*, 2005).

Although there is no clear and definitive cut-off for which individuals are superspreaders, there is a great deal of interest in identifying and limiting the impact of

those individuals (Galvani and May, 2005; Kitsak *et al.*, 2010). This task is computationally complex, if even possible (Gu *et al.*, 2017), but more than that it is highly contextual. Transmission depends on both a changing physical and social environment, as well as being dependent on the specifics of the pathogen involved. Effectively, every pathogen can potentially have a different contact network.

Anthropologists have long recognized the importance of social networks to understanding culture and even its relevance to disease transmission (Morris, 1993; Salathe *et al.*, 2010; Nunn *et al.*, 2015). The structure of a social network, in the context of infectious disease modeling, means the potentially changing patterns of contact through which disease may be transmitted. In this way, disease contact networks are often unique to the specific disease, and not universal to a particular society. A great deal of effort has been put into identifying the structure of social networks, but the task is monumental. Contact networks are disease-specific, since different diseases can have different properties and routes of transmission, and they can be ephemeral.

2.2 Hypothesis

This model does not have the goal of demonstrating SARS was certainly a subcritical epidemic, that is an outbreak where $R_0 < 1$. This project investigates the question whether the influence of average contact network, specifically the existence of superspreading events, can result a similar total case count as seen in SARS without requiring $R_0 > 1$. This provides information on the question of whether separating subcritical epidemic from supercritical epidemics is a feasible goal.

2.3 Methods

2.3.1 Modeling Logic

The question here is whether a SARS-like epidemic could be reasonably modeled as a subcritical epidemic. To address this question, a simple model of subcritical epidemics is described, and realistic parameters are drawn from the literature. Previous work on outbreaks for which $R_0 < 1$ has argued them to be generally small, sometimes called "stuttering" epidemics (Blumberg and Lloyd-Smith, 2013b). Thus the metric used to compare model output to the actual SARS data is total case count.

The model output matching the SARS data demonstrates the possibility that SARS-sized outbreaks can result from subcritical epidemics, due to the presence of superspreading events.

2.3.2 Model Structure

This model starts with a simple subcritical epidemic ordinary differential equation, equation 2.1 described here as the EI model. There are two compartments in the EI model, E which describes the number of exposed individuals, and I which describes the number of infectious individuals. As a deterministic system, however, the EI model does not have interesting dynamics.

$$\frac{dE}{dt} = \alpha I - \beta E \tag{2.1}$$

$$\frac{dI}{dt} = \beta E - \delta I \tag{2.2}$$

Stochasticity is added to this model using the Poisson method, described in Gustafsson (2000). Transitions between the E and I compartments it taken to be a Poisson random variable, as is the transition out of the I compartment. The entrance

into the E compartment is modeled to have a rate of αI. However to capture the effect of superspreading events, this is taken to have a negative binomial rate, with an overdispersion parameter of ϕ_α. The justification for using the negative binomial is given in section 2.3.3 and section 1.2.4.

It is important to note that this model only has zero as a stable equilibrium R_0 is less than 1. If $(E, I) = (0, 0)$ becomes an unstable equilibrium, the number of infected individuals can potentially go to infinity, which is considered unbiological. As such, this model is only applicable to subcritical epidemics.

2.3.3 Justification of the Negative Binomial

The negative binomial can be realized as the mixture of a gamma random variable as the argument of a Poisson (Casella and Berger, 2002). That is a random draw from $X \sim$ Negative Binomial(μ, ϕ_α) is equivalent to a random draw γ from Gamma(μ, ϕ_α) random variable, and then another random draw from a Poisson(γ) random variable.

The rationale behind this procedure is to add additional variance to the Poisson random variable. It allows for some departures from a Poisson process, making it useful as a distribution which fits data with heterogeneous Poisson rates. There is a very reasonable question as to whether that heterogeneity is captured by a Gamma distribution, however that is an empirical question. There are two ways to justify the choice, therefore. First, comparison to existing work, for example Lloyd-Smith *et al.* (2005) shows a good the negative binomial is a good empirical fit to the individual reproductive number. This can be generalized using the argument below to justify its use for incidence. Second, the Poisson is the limiting case of the Negative Binomial (as $\phi_\alpha \to 1$), so in a sense, the Negative Binomial can be thought of as an almost-Poisson process with more flexibility to account for at least some heterogeneity in rates, even if it can not capture it all or perfectly.

The justification for moving from an individual reproductive number in the sense of Lloyd-Smith *et al.* (2005) to incidence requires additional work. Since the sum of negative binomials is also negative binomially distributed, the distribution for incidence is also negatively binomially distributed.

2.3.4 Model Parameters

There have been multiple models of SARS which use similar parameters to the three disease parameters in this model. One of the first attempts at estimation of the key parameters is Donnelly *et al.* (2003), which used data from the outbreak in Hong Kong. They estimated an incubation period (time until symptoms appear) of 6.37 days, with a variance of 16.69 days. The estimated hospital admission time had a mean of 23.5 days and a variance of 62.1 days, however if the patient died in hospital, the admission to death time was longer, with a mean of 35.9 days and a variance of 572.9 days. The period of highest risk to the public–the period after the onset of symptoms but before admission to the hospital–varied as the outbreak progressed. Early in the outbreak, the mean time to admission after the onset of symptoms was 4.85 [1] and a variance of 12.19 days, mid-outbreak the time to admission dropped to 3.83 days (variance 5.99, and nearing the end of the outbreak, dropped to 3.67 days (variance 10.71 days).

These data were used by Chowell *et al.* (2004) to estimate the distribution of the basic reproductive number R_0 using a statistical model and sensitivity analysis. They also included an estimate of the transmission rate of 0.25 new cases per infected person per day. Fitting the Hong Kong outbreak to their model, Riley *et al.* (2003) estimated the transmission rate to be 0.062 new infections per infected person per day, during the initial phase of the infection, although the parameter in their model

[1] Donnelly et. al. contains a typo, listing this number as 48.5

was allowed to vary at later points in the outbreak.

The WHO consensus report (WHO, 2003a) summarized multiple sources of data from a variety of studies of SARS in different regions, finding the men incubation period ranged from 4 - 7.2 days. They also report real-time polymerase chain reaction studies of viral shedding in respiratory and fecal samples. Fecal samples show a peak around 10 days after the appearance of symptoms, with respiratory samples showing a peak 12-14 days into the symptomatic period.

As challenging as it is to estimate the rate of new infections per infected individual, it is the variance in that estimate that is of great interest to capture the superspreader phenomenon. Lloyd-Smith *et al.* (2005) investigated the distribution of the number of new infections caused by an infected individual in SARS and seven other infectious diseases. They describe this "individual reproductive number," finding that the negative binomial was a good fit to the data from the Singapore and Beijing. They fit a dispersion parameter ν to the Singapore data, finding a maximum likelihood fit of 0.16.

Using the preceding, the parameter estimates used in this project are shown in table 2.1. The rationale for each is discussed below.

This model is only applicable for subcritical epidemics, since any supercritical epidemic would probably result in the number of infections going to infinity. As such, only those values of the parameters which result in $R_0 < 1$ are considered.

The Rate of New Infections α

Perhaps the most challenging parameter to set, α represents the number of new infections caused by an infected individual per day of their infection. This can be thought of as the combination of two factors: the number of contacts between infected individuals and uninfected times the probability that such a contact leads to a new

Variable	Value	Range	Description
α	0.15	$0.05 - 0.25$	The infection rate per infected individual per day.
ϕ_α	5	$1.94 - 7.25$	The overdispersion parameter of α
β	$\frac{1}{5}$	$\frac{1}{7.2} - \frac{1}{4}$	The reciprocal of the latent period.
δ	$\frac{1}{5}$	$\frac{1}{19} - \frac{1}{3}$	The reciprocal of the average time between the initial infection and the ability to infect others.

Table 2.1: The parameter values used in this model.

infection. The EI model is constructed to simulate situations where $R_0 < 1$. Since $R_0 = \frac{\alpha}{\delta}$, and δ is reasonably estimated to be less than $\frac{1}{3}$, the Chowell *et al.* (2004) estimate of $\alpha = 0.25$ provides a reasonable upper bound. Riley *et al.* (2003) estimated the transmission rate to be 0.062. This was rounded down to 0.05 as the lower bound of the range.

In addition to the parameter α itself, it is also necessary to estimate the overdispersion ϕ_α. Using Lloyd-Smith *et al.* (2005) estimate of the dispersion parameter $\nu = 0.16$, it is possible to estimate a range of ϕ_α. The dispersion parameter ν is related to this model's overdispersion parameter ϕ_α according to the formula $\phi_\alpha = 1 + \frac{R_0}{\nu}$. Since $R_0 = \frac{\alpha}{\delta} < 1$ in this model, and $\alpha \approx 0.05 - 0.25$ while $\delta \approx \frac{1}{16} - \frac{1}{3}$ (see below), R_0 ranges from 0.15 to 1. Thus the overdispersion parameter ranges from $1.94 - 7.25$.

Since the overdispersion is how this model captures the effects of superspreading events. As such, it is the only overdispersion parameter which will include a range of possible values. The larger values of ϕ_α describe more connected individuals, however necessarily fewer of them.

The Reciprocal Latent Period β

The latent period is defined as the starting with the infection of an individual and ending when that individual becomes infectious. Infectiousness is not always apparent, so a related quantity, the incubation period has been better reported. The incubation period is ends with the first appearance of symptoms. Since SARS is spread by airborne droplet, it is probable that the appearance of symptoms and ability to infect others are close in time. Nevertheless, that is an assumption.

The WHO consensus report (WHO, 2003a) estimated the incubation period to be between 4 and 7.2 days; Donnelly *et al.* (2003) made a consistent estimate of 6.37 days. As such, the incubation period is taken to be approximately 5 days, setting $\beta = \frac{1}{5}$, with a range $\frac{1}{7.2} - \frac{1}{4}$.

To determine the appropriate distribution, note Donnelly's estimate of a variance of 16.69, which is considerably larger than the estimate of the incubation period of 5 days. Since overdispersion of the β parameter would be overdispersion of the reciprocal of the incubation period, this variance of the incubation period does not immediately lead to an estimate of the variance of the reciprocal. Without specifying a distribution, it is not possible to estimate the variance of the reciprocal. Choosing the negative binomial is not possible, since the presence of zeros means the reciprocal is not defined. A truncated negative binomial, with the zeros removed, does have a reciprocal, and the results are underdispersed. The normal distribution contains negative values and values close to zero, which wildly inflates the variance of the reciprocal. As no distribution is obvious, a Poisson distribution was chosen as a default.

The Reciprocal Infectious Period δ

Quarantine, if completely effective, can be thought of as shortening the infectious period. Quarantine is thought to be a critical factor in ultimately controlling the spread of SARS. However hospitalization was not always effective in limiting the spread of SARS. Thus this model takes the lower bound of the range of $\frac{1}{\delta}$ to be about 4 days, which is approximately the mean time to hospitalization. Although hospitalization continued to 23 days, the viral shedding in the respiratory tract peaked around day 12-14 (WHO, 2003a). The infectious period was extended to a mean of 19 days, for a situation in which the hospitalization was ineffective at controlling the spread of SARS. Under the assumption that the hospitalization is effective at controlling the spread of the disease, the mean infectious period is chosen to be 5 days, making $\delta = \frac{1}{5}$.

For similar reasons to the choice of distribution for β, the distribution of δ is assumed to be Poisson.

2.3.5 Simulations

The model described above was implemented in Python 2.7.12 and simulated using a time step of 0.01 days. For most parameters, the model was simulated 500,000 times at each of the six integer values of ϕ_α. Each of the 500,000 simulations represents a new spillover event or a new infected individual arriving at a totally susceptible city. The total case count was recorded for each of the model runs. Larger values of ϕ_α correspond to more superspreading events, however to keep the mean fixed, necessarily there will also be more outbreaks which stop immediately.

Each simulation was begun with a single individual in the I compartment, simulating an infected individual arriving in the city. The maximum time of any single

	Number of secondary cases				
	0-4	5-49	50 - 99	100-999	1000+
Baseline, $\phi_\alpha = 5$	90.4	8.3	0.8	0.5	0*
Baseline, $\phi_\alpha = 7$	91.5	7.1	0.8	0.6	0*

Table 2.2: Percentage of simulations which were between the specified range of secondary cases. * The percentage rounded to zero, however there were two instances of 1000+ sized simulations for $\phi_\alpha = 5$ and four instances for $\phi_\alpha = 7$.

simulation was 5000 days, although no simulations reached this value. Simulated outbreaks which ended (no exposed or infectious individuals remaining) were stopped.

In each simulation, the total number of secondary infections was tracked. The result was a list of case counts for 500,000 attempts at the introduction of SARS to a city, for each of the six values of ϕ_α in its range.

2.4 Results

The simulation of 500,000 spillover/new infection events with parameters at their baseline values and $\phi_\alpha = 5$ demonstrated an overwhelming number of simulations (78.6%) result in no secondary infections (see figure 2.1). Fully 90.4% of simulations showed 4 or fewer secondary cases.

Since there is a possibility that cities with few cases might not have detected the cases, only simulations which had 5 or more secondary cases are shown in figure 2.2

The percentage of secondary cases at the baseline parameters is shown in table 2.2.

The effect of increasing ϕ_α over its range (1.94 - 7) is to increase the size of rare, large epidemics while also increasing the number of epidemics which end immediately with no secondary cases. Figure 2.3 shows that increasing ϕ_α, the overdispersion parameter has the effect of increasing the percent of epidemics which are fewer than five

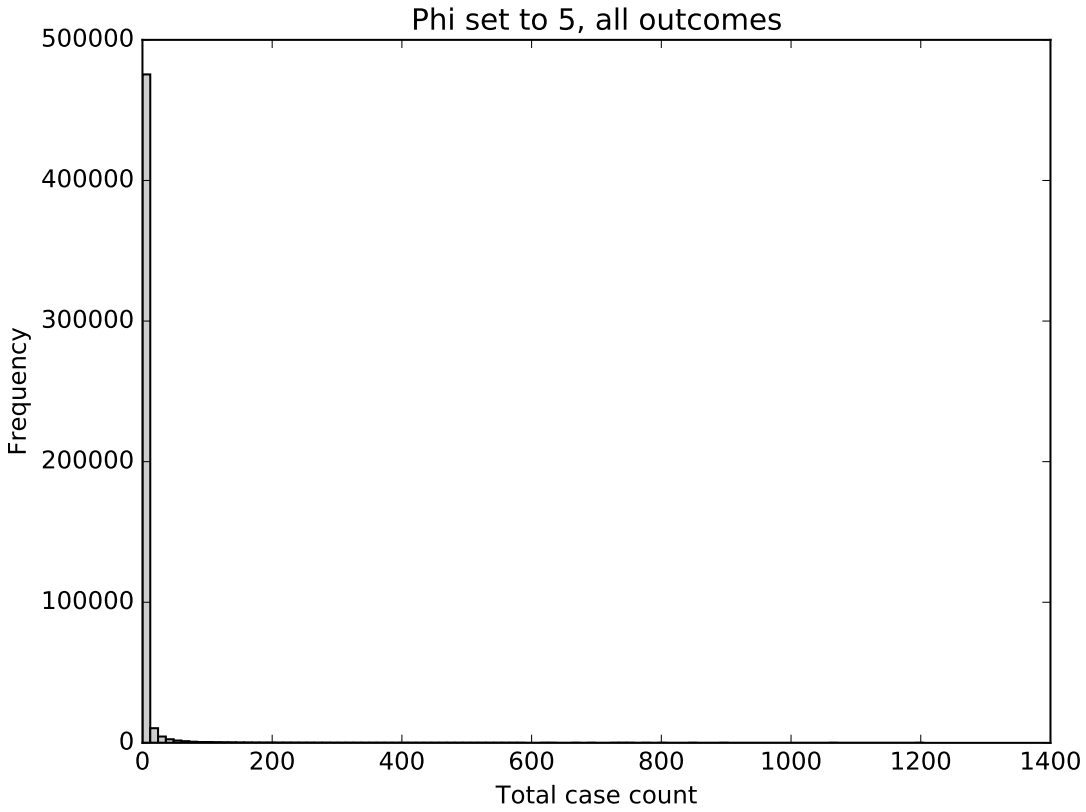

Figure 2.1: This shows the results of 500,000 simulations of the model at baseline parameter values and with all simulations shown. The overdispersion parameter was set $\phi_\alpha = 5$.

individuals. Figure 2.4 shows the percent of epidemics which exceeded 50 individuals for larger values of ϕ_α. The pattern is less-clear in those results.

The pattern shown in figure 2.4 appears to be increasing. Thus the patten of increasing the overdispersion parameter ϕ_α appears to be more small epidemics and larger rare epidemics. The number of introductions which result in small, potentially undetectable epidemics goes up with increasing ϕ_α, but so too does the number of large epidemics, shown here as epidemics which exceed 100 individuals. There is some leveling off of both graphs, which is likely a binning effect.

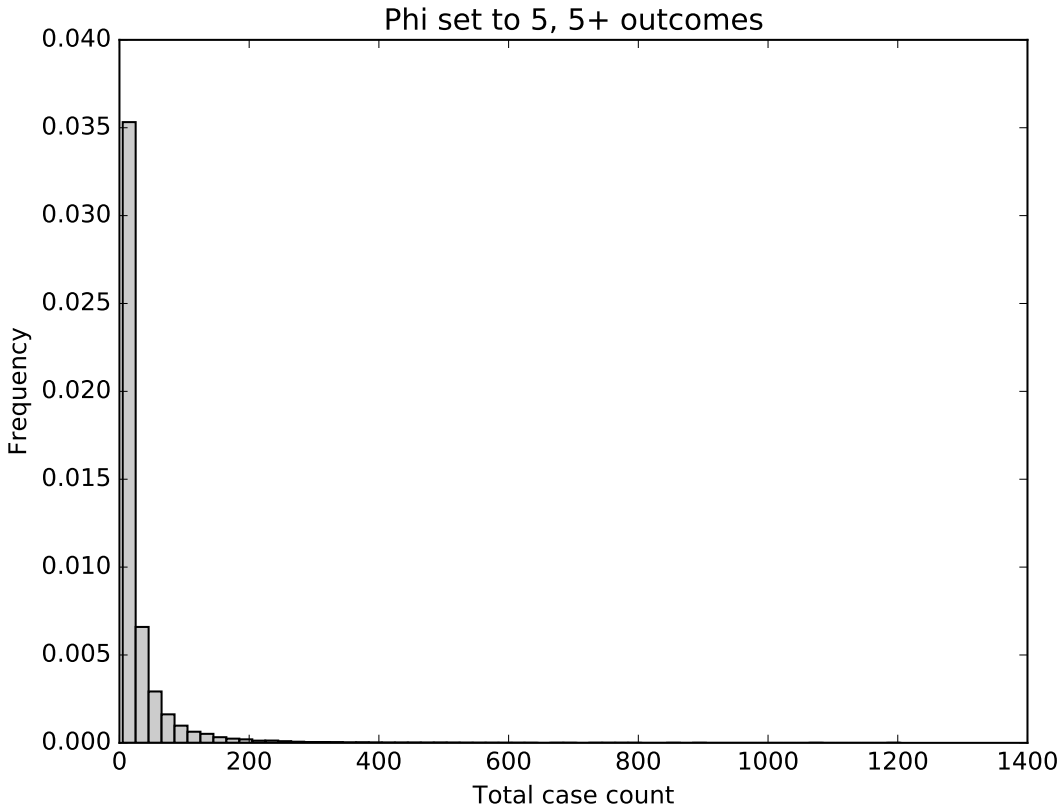

Figure 2.2: This shows the results of 500,000 simulations of the model at baseline parameters and with $\phi_\alpha = 5$, with only those simulations in which 5 or more secondary infections occurred.

2.5 Discussion

The simulations reveal that increasing the value of ϕ_α, the overdispersion parameter, increases the number of outbreaks in the tens of individuals. This is similar, in broad strokes, to the distribution of SARS case counts by city, see tables 2.3 and 2.4. While there is not a perfect match between the distribution in the sizes of the outbreaks, the data are also limited. Only a few cities reported large numbers of cases, and in some cases, individuals who had no or few secondary transmission events may have gone unreported.

The results demonstrate that SARS, which is widely recognized as an epidemic,

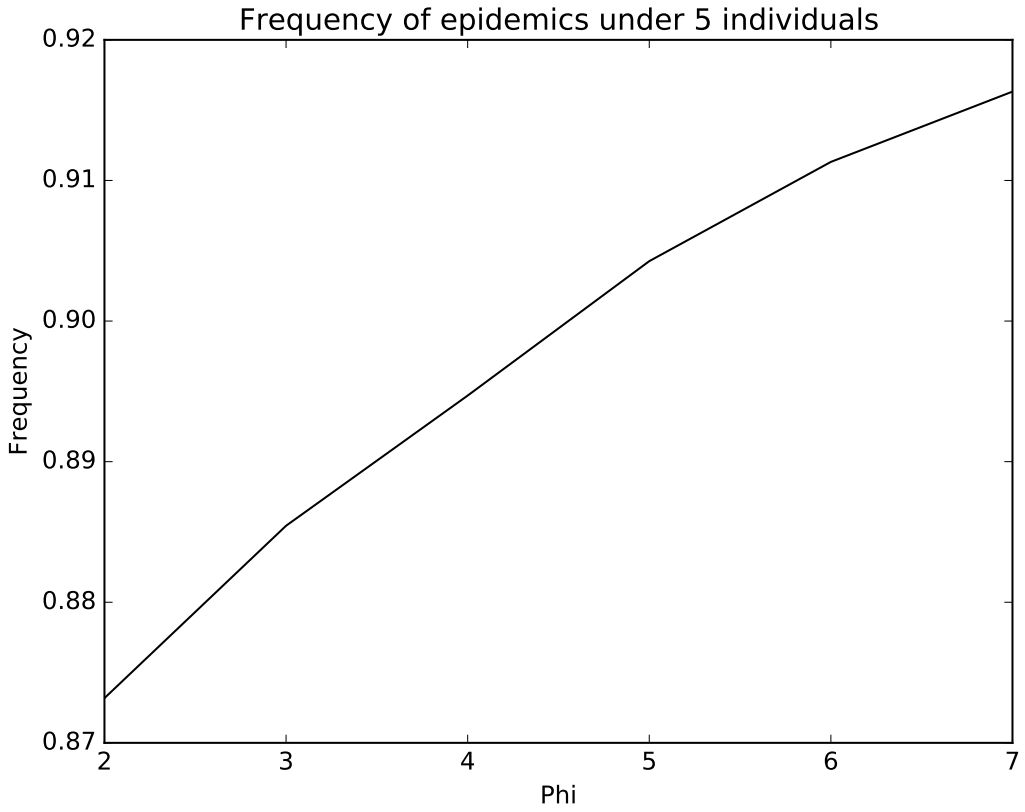

Figure 2.3: This shows the results of 500,000 simulated epidemics for each $\phi_\alpha = 2$ through $\phi_\alpha = 7$. The vertical axis is the fraction of simulated epidemics which were smaller than 5 secondary infections, including those which resulted in no secondary infections.

has similarities in the size of the outbreak to simulations of a subcritical epidemic.

The history of the 2003 SARS outbreak indicates a chain of limited transmission in Foshan and Guangzhou provinces prior to the rapid increase in the disease in early 2003 (WHO, 2003a). This may be characteristic of subcritical epidemic dynamics, sustained as a stuttering chain. However the modeling results demonstrate that even subcritical epidemic, which modeled with superspreading events, a substantial increase in the incidence of the disease. Although there is no evidence that SARS was indeed a subcritical epidemic, the results demonstrate the challenges in distinguishing between supercritical epidemics ($R_0 > 1$) from subcritical epidemics ($R_0 < 1$).

Country	Cases	Country	Cases
Australia	6	Mongolia	9
Brazil	1	New Zealand	1
Canada	251	Philippines	14
China	5327	Republic of Ireland	1
Hong Kong SAR	1755	Republic of Korea	3
Macao SAR	1	Romania	1
Taiwan	665	Russian Federation	1
Colombia	1	Singapore	238
Finland	1	South Africa	1
France	7	Spain	1
Germany	9	Sweden	3
India	3	Switzerland	1
Indonesia	2	Thailand	9
Italy	4	United Kingdom	4
Kuwait	1	United States	33
Malaysia	5	Vietnam	63

Table 2.3: Total number of cases of SARS by country, data compiled by WHO (2003b). No distinction was drawn between individuals who arrived in each country with SARS, and within-country transmission. China includes only mainland China, excluding its Special Administrative Regions.

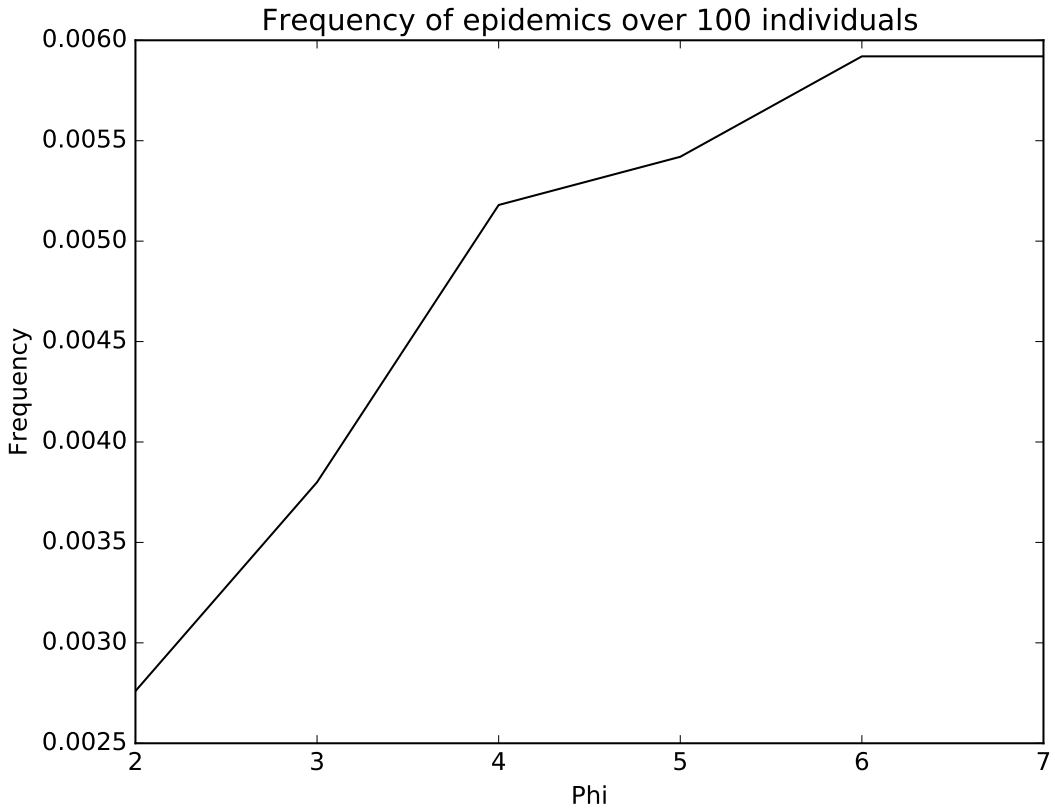

Figure 2.4: This shows the results of 500,000 simulated epidemics for each $\phi_\alpha = 2$ through $\phi_\alpha = 7$. The vertical axis is the fraction of simulated epidemics which greater than 100 secondary infections.

City	Country	Cases	Reference
Beijing	China	2406	(Liang *et al.*, 2004)
Hong Kong	Hong Kong SAR, China	1755	(WHO, 2003b)
Guangzhou	China	1246	(Xu *et al.*, 2004)
Taipei	Taiwan	341	(CDC, 2003)
Singapore	Singapore	238	(WHO, 2003b)
Toronto	Canada	140	(Low, 2004)
Hanoi	Vietnam	63	(Nishiyama *et al.*, 2008)

Table 2.4: Cities with case counts exceeding 50. It is very likely that there are additional cities in China which have not been recorded or reported. Most cities and towns which had reported case counts reported fewer than 50.

Furthermore, since cities with few or no secondary transmission may be missed, detection of subcritical epidemic transmission may be hampered by the appearance of only large or moderate transmission chains.

The results, however, are not conclusive. The simulation of the model at baseline parameters demonstrated that epidemics of a size over 1000 could occur, but were rare. The larger the size of the epidemic, the more rarely it should occur. Given that two cities had epidemics which exceeded 1000 people, this would suggest that SARS does have a higher rate of transmission than the subcritical model can account, although it is impossible to demonstrate this definitively given the uncertainty of the data and the rarity of the events. In the simulation of 500,000 epidemics with $\phi_\alpha = 5$, only twice did the epidemic exceed 1000 secondary cases. When $\phi_\alpha = 7$, four of the 500,000 simulations resulted in an epidemic over 1000. These are possible but rare events. Since the city by city breakdown of the SARS data is not available for China, it is difficult to determine if the distribution of city outbreak sizes is statistically similar to the estimated distribution. However qualitatively, outbreaks of very large size are possible in SARS and qualitatively have a similar distribution.

Differentiation between subcritical epidemic and supercritical epidemics has been a topic of interest (Blumberg and Lloyd-Smith, 2013a; Guttorp and Perlman, 2015), however much of this modeling has been done using branching process models. These have an extremely limited ability to capture disease dynamics which have elsewhere been captured. The negative binomial extension of the Poisson method provides a tool to incorporate superspreading into any ordinary differential equation model, a wealth of which have been developed for a wide range of diseases (Anderson and May, 2010). As such, it is a step in the direction of capturing superspreading events in models, without the simplifying assumptions of branching process models which would limit the utility to all but the simplest cases for the spread of disease.

The model developed here specifically demonstrates surprisingly large outbreaks, similar in size to those expected from epidemics, are possible due to superspreading events. It is not clear from the modeling effort herein described that it is possible or likely that distinguishing between subcritical epidemic and supercritical epidemics is possible from the case count data alone. However any such attempts will undoubtedly need to be founded in models which reflect the transmission dynamics of the diseases, which is challenging with branching process models.

2.6 Conclusions

The model presented here is not a demonstration that SARS was a subcritical epidemic. However the size of subcritical epidemic generated with SARS-like parameters are sometimes similar in size to the SARS outbreaks. This demonstrates the challenge in identifying subcritical epidemic; even if R_0 is small, the potential for superspreading events leads to potentially large outbreaks. This indicates that due to contact network structure, the prospects epidemiological separation of supercritical and subcritical epidemics based on case count is limited.

One of the characteristics apparent in the subcritical epidemics with superspreading is the frequency with which introduction of an infectious disease fails to transmit in humans. Repeated failures of a disease to pass to other humans may be taken by some as evidence that the disease has a low potential to become an epidemic in humans, beyond a few cases. However this research demonstrates that such behavior is not inconsistent with a disease which could become a global pandemic in multiple cities, if the disease and social network are conducive to superspreading. This indicates a clear need for discussion of social networks in the assessment of potential pandemic pathogens.

From a public health standpoint, in SARS and other diseases, a frequent goal

of control efforts is to drive R_0 less than one through interventions (WHO, 2003a). If, as this work shows, it is possible for there to be large epidemics of infectious disease arising from subcritical R_0 values, this may indicate that R_0 is a poor target for epidemiological intervention. It may also have limited utility in predicting the epidemic potential of novel infectious diseases. Instead, research into social network structure may provide more fruitful results. More broadly, diseases which repeatedly spillover to human populations and fail to spread might be considered by some to be a low health risk. However it is possible that history of failure in spreading in humans may not be a good predictor of future epidemics. Superspreading events, as demonstrated here, can result in large, global pandemics. This has clear indications of a need for social network considerations in infectious disease management and planning.

Chapter 3

THE EFFECT OF CONTRACT NETWORKS ON TB TRANSMISSION

3.1 Introduction

Tuberculosis (TB) is a bacterial infection found in many species, including humans, caused by the *Mycobacterium tuberculosis* complex of bacterial species. The origins of the infection in humans is widely debated (Bos *et al.*, 2014; Comas *et al.*, 2013; Brosch *et al.*, 2002; Wilbur *et al.*, 2009), and draws on evidence from a wide range of sources. In this project, a model of TB in a hunter-gatherer community is developed to investigate the effects of the network structure on disease transmission. The goal of this modeling effort is to elucidate the origins of TB in the human population.

TB has been highly successful in infecting humans. TB is one of the most common infectious diseases, approximately 1.7 billion people infected worldwide (World Health Organization, 2017). The simulation of TB on the unique network pattern of hunter-gatherer allows for the investigation of the unusual transmission dynamics. Most of the TB cases are latent, wherein the immune system keeps the TB pathogen in check and reduces the transmissible to effectively zero. However not all cases are latent. Active TB describes an infection which is readily spread from person to person via airborne infection. A person with active TB has both a shortened lifespan, if untreated, and may infect others in close proximity (World Health Organization, 2017).

One explanation for the transmission pattern of TB is the success of the human immune system at fighting off the pathogen. Certainly this is the proximate explana-

tion for the pattern of TB transmission, however this project investigates a different, evolutionary explanation for this pattern. The model developed below is used to assess whether TB is uniquely evolved to transmit in hunter-gatherer contact networks. This may be an additional line of evidence on the ultimate origin of the TB pathogen in the human population.

3.2 Background

3.2.1 TB Origins and Evolution

The introduction of TB to humans is a wide-ranging debate with many sources of data and no clear consensus on the timing or circumstances. This debate has its origins in the mid-20th century, and draws upon data including paleopathological, genetic, and microbial. This project takes a novel look at the debate by considering the effect the social network of hunter gatherers would have had on the evolution of the tuberculosis pathogen.

Investigations in the mid-20th century were largely based on historical records and the paucity of reports of tuberculosis-like symptoms (Clark *et al.*, 1987; Roberts and Buikstra, 2003). Of note was the lack of resistance to tuberculosis found in many indigenous groups in the Americas, which suggested no prior exposure. Investigations of archaeological specimens for signs of skeletal damage consistent with tuberculosis found evidence of lesions in specimens which predated European arrival in the Americas (Morse, 1961). The question of pre-Columbian tuberculosis was mostly settled in 1973 with the identification of probable *Mycobacterium* bacteria in a prehistoric Peruvian mummy (Allison *et al.*, 1973; Clark *et al.*, 1987). The mummy had evidence of skeletal damage consistent with TB, as well as acid-fast bacteria which are most commonly *Mycobacterium*. However being a soil-dwelling genus, contamination with

another *Mycobacterium* species could not be ruled out at that time, and is a problem which persists to the present (Muller *et al.*, 2016). Salo *et al.* (1994) further tightened this identification with the demonstration of *M. tuberculosis* DNA from a Peruvian mummy in 1994.

However, the abundance of skeletal and DNA evidence of pre-European tuberculosis raised the question of origin and spread of tuberculosis in human broadly. Stead *et al.* (1995) hypothesized that TB originated in humans as a spillover event from the strain of TB found in cattle, *M. bovis*, likely around the time of the domestication of cattle or after. This history, they note, is inconsistent with the appearance of tuberculosis in the new world prior to European contact, and as such they dismiss the analysis of Salo *et al.* (1994).

The general contention, however, is that if TB was present in the new world prior to contact with Europeans, it likely had a much more ancient origin in the human species. Genetic evidence has expanded the analytical methods available in the search for the origins of tuberculosis (Tibayrenc, 2004). Broadly these fall into three groups: comparison of closely related species, comparisons of modern strains, and studies on ancient DNA. The oldest date from these studies is the divergence of TB from closely related species. Gutierrez *et al.* (2005) investigated this question, finding a genetically complex story and an origin of approximately 3 million years before present.

The advent of genome sequencing brought a new tool for the analysis of the origin of tuberculosis in humans (Cole *et al.*, 1998). Comas *et al.* (2013) used whole genome sequences of 259 modern strains of TB to develop a phylogeny of modern tuberculosis. Their analysis, along with some assumptions of the mutation rate for TB known as the genetic clock, identify an ancient origin to TB. They characterize pathogens co-evolving with, and migrating with humans as they migrate out of Africa approximately 70,000 years before present. This pattern of TB evolution following

human migration was supported by Wirth *et al.* (2008), who estimated a younger, though still ancient, origin of TB at 40,000 years BP.

Their finding is consistent with a similar investigation conducted by Luo *et al.* (2015), which estimated the age of the Asian strains of TB to be approximately 30,000 years old. Like all studies using a genetic clock, phylogenetic studies have been criticized for lack of firm data on the mutation rate of the disease (Bos, 2018).

The ancient origin of tuberculosis in humans was disputed by Bos *et al.* (2014), in a study of ancient DNA from TB samples found in three pre-contact Peruvian mummies. The strain of TB identified in the study was most closely related to *M. pinnipedii*, a strain primarily found in seals and sea lions and of very recent—approximately 6000 years bp–divergence from their last common ancestor. This indicates a relatively recent origin of TB in humans. Not all studies of ancient DNA have reached a similar conclusion, however. Spigelman *et al.* (2015) examined human remains from a 9000 year old archaeological site, finding evidence of a tuberculosis infection in an infant. Similarly Posa *et al.* (2015) identified tuberculosis from a grave site in Hungary, dating to approximately 7000 years BP, seeming to suggest the introduction of TB detected by Bos *et al.* (2014) was not the origin of TB in humans.

Further supporting the ancient origin of tuberculosis is paleoanthropological data. Bone damage has been identified in the extinct human ancestor species *Homo erectus* which may have been caused by TB (Kappelman *et al.*, 2007), although this is controversial (Roberts *et al.*, 2009). There is broad consensus that TB was present in the New World prior to European contact (Bos *et al.*, 2014), although how much earlier and the pattern of introduction remains unknown.

A final line of evidence has been suggested, although not throughly investigated. Gagneux (2012) has suggested that the pattern of transmission of tuberculosis is indicative of long-term coevolution with humans. Qualitatively, TB may have evolved

traits which make it particularly well-suited to transmission in human populations, although absent a quantitative approach, the precise nature of these traits is not clear (Gagneux, 2012, 2018). Perhaps the most prominent feature of TB which has been suggested is that TB has "sympatrically coevolved" with human hosts (Gagneux, 2012). Transmission is more probable when TB strains are those found local to the people being infected, suggesting local TB adaptation (Fenner *et al.*, 2013; Hirsh *et al.*, 2004). However the meaning of these data on the timescales of human history are unclear. However Gagneux (2012) identified the possibility that other features suggesting coevolution could be found, and this was an identified need in the literature.

The present project seeks to investigate this question by examining whether TB has traits which would be advantageous in a hunter-gatherer population, but disadvantageous in a highly connected population.

3.2.2 TB in Hunter-Gatherer Populations

Investigations into the transmission dynamics of tuberculosis has found considerable variation between different urban populations (Hirsh *et al.*, 2004) and between urban and hunter-gatherer (Hurtado *et al.*, 2003). Prior to European contact, tuberculosis was not known in the Ache, a population of hunter-gatherers in the highlands of Paraguay. As such, they were an immunologically naive population upon the introduction of tuberculosis in the late 1970s or early 1980s. The epidemic differed in several ways from epidemics seen in urban populations. (Hurtado *et al.*, 2003)

The Ache tuberculosis epidemic affected a large percentage of the Ache population; Hurtado *et al.* (2003) reported that of 427 tuberculin purified protein derivative (PPD) tests given, 30.4% tested positive. This number is likely an underestimate, as they also report active cases testing negative on the PPD test, and BCG vaccination status

of individuals was negatively correlated with their PPD test results. This is unusual in comparison to urban populations, where BCG vaccination generally causes a positive PPD test, and the PPD test has generally been found to be reliable (Moran-Mendoza *et al.*, 2013). These are suggestive of immunological complexity in the infection.

Of particular relevance to the current project, Hurtado *et al.* (2003) found active tuberculosis in the Ache had a case fatality rate of 7.7%, and an incidence rate of 3.7 cases per 100 people per year. Given the challenges of assessing latent infections, the rate at which individuals transition from latent to active is difficult to assess, however it is higher than in urban populations. There is also a significant age-effect, with children under 5 years of age having a lower rate of active tuberculosis than other age ranges.

The differences in the transmission dynamics of tuberculosis in the Ache may be the result from immunological or pathogen variation, as has been suggested by Hirsh *et al.* (2004) for urban populations. There may also be higher susceptibility due to social network structure, as has been suggested for other small scale societies (McGrath, 1988).

3.2.3 Hunter-Gatherer Population Structure

Human social life is a key distinguishing feature of human behavior (Hill *et al.*, 2009). For most of the time that anatomically modern humans have been on the planet, they have subsisted as hunter-gatherers (Richerson *et al.*, 2001). Although human social networks vary widely, hunter-gatherer social networks share some similarities (Hill *et al.*, 2011). Hunter-gatherer populations are frequently divided into bands, which have high levels of within-band cooperation and contact. Between-band contact is present, but much less common. The rationale or social norms which give rise to this common structure varies, but Hill *et al.* (2011) found an average band size

of 28 individuals in a survey of hunter-gatherer groups.

Transmission of disease occurs on close contact, defining a unique "contact network" or disease-specific "social network." Here these terms are used interchangeably to describe the close proximity of humans through which TB might transmit. This project uses the estimate of 25 individuals per band, although death may reduce some bands to be smaller. Within band dynamics are often complex, but are not modeled in this project. Instead, within band dynamics are modeled as completely random, with transmission between any two members of the same band to be identical to any other two members.

3.2.4 The Basic Reproductive Number R_0

There are few ideas more central to infectious disease epidemiology than the basic reproductive number R_0 (Anderson and May, 2010). The value, R_0, has the intuition of the number of new infections caused by an infected individual. It has been widely used and is highly influential in epidemiology (Heesterbeek, 2002; Cushing and Diekmann, 2016).

Placing a particular point of the introduction of R_0 is challenging (Heesterbeek, 2002). However R_0 is one of the major methods by which the public health threat of an infectious disease is quantified, and its reduction is a major target for public health interventions (Anderson and May, 2010; Liu *et al.*, 2016).

In addition to the consideration of diseases in the present, it is important to look at both their evolutionary history and potential future trajectories (Nesse and Williams, 1996). Pathogens are in a co-evolutionary relationship with their hosts which is often described as an evolutionary arms race (Nesse and Williams, 1996; Gluckman *et al.*, 2009). The pathogens evolve strategies and weapons, while the host evolves defenses.

The naive assumption of the evolutionary trajectory of pathogens is perhaps that

they would always seek to increase their R_0 in the population. For many diseases, it is likely that they do experience an evolutionary pressure to increase R_0, however it may not be for all.

Early examination of the evolution of disease, even predating major theoretical developments such as R_0 and the field of evolutionary medicine, noted that diseases may in some circumstances evolve toward lessen the symptoms of infection (Gluckman *et al.*, 2009). A pathogen which kills its host also potentially kills itself. However R_0 is not necessarily tied to the symptoms of a disease. In a homogeneous population, the ideal infection from the pathogen's point of view is one which is easily transmitted but which lives for an extended period of time in the host.

Modern TB is also usually latent, which often does not transition to an active infection. One account of why TB is so often latent is the effectiveness of the human immune system, which is undoubtedly a proximate explanation; immunocompromised individuals are more likely to develop active tuberculosis on infection, and may transition from a latent to an active infection. However the proximate explanation may not be the ultimate, evolutionary explanation for this pattern in TB. Gagneux (2012) suggested qualitatively that a long latent period might also be expected of a pathogen transmitting on a hunter-gatherer contact network. What is shown below is that under some reasonable assumptions, TB in a hunter-gather population may evolve this longer latent period, even though it has the effect of lowering R_0.

3.3 Methods

To investigate the dynamics of TB in a hunter-gatherer community, a model simulates the introduction of TB to a completely susceptible population. The TB model simulates infection dynamics within 25 person bands, simulating daily contact with every other person in the band. Each contact has a probability of transmission.

There are separate random transmission between bands. For each band, the number of susceptible, latent, and active individuals are tracked. Birth processes are simulated within the band as a function of expected population size, while deaths occur randomly at a rate dependent on infection status.

A single active case of TB is introduced into the population, and the infection is simulated for 100 years. The final number of latent and active cases of TB is tabulated. This process is repeated 300 times per set of parameters.

3.3.1 Model Construction

The modeling framework used here is a Reed-Frost chain binomial type model on a pre-defined metapopulation type network. Each band i has a vector $(S, L, I)_{i,t}$ which tracks the number of susceptible S, latent L, and active I cases at time t. The time step is one day. In each time step, every individual with active TB interacts with every susceptible individual in the same band. This leads to SI interactions, each of which has a probability p_g of transmission. Thus, in every time step, the number of new within-group transmission events is a Binomial(SI, p_g) random variable. New infections have a p_a probability of becoming immediately active, and $1-p_a$ probability of becoming latent infections.

The latent infected individuals have a mortality rate m_l, and can also transition to active TB at a rate of a. Thus every time step, and in every band, Binomial(L, m_l) and Binomial(L, a) random variables are computed and subtracted from the latent class. In rare cases, these drive the L variable in a band negative, in which case it is set back to zero. The active TB cases have a higher mortality rate, m_a. Thus in every time step, a Binomial(I, m_a) random variable is calculated for each band and subtracted from the active class.

When the number of individuals in a group is less than 25, another random variable

is calculated. The birth rate is equal to the latent mortality rate m_l, and is computed every day for every band as a binomial$(25 - (S + L + I), m_l)$ random variable. All new individuals are added to the susceptible class.

Between group infections are modeled differently from within group infections. Each group is randomly paired with another group (repetition is permitted). If there are I_i active cases in group i, and S_j susceptible individuals group j, a Binomial$(I_i S_j, p_b)$ random variable is calculated. Never more than one new between-group infection is permitted, however. A new infection is assigned to be active with probability p_a.

Each of the parameters was kept constant (see below) except p_a, the probability a new TB case is active, which is varied between 0.1 and 1.0 with 0.1 increments. For each set of parameters, 100 years (36525 days) were simulated 300 times. In each simulation, one case of active TB was introduced into the population. The number of bands in the simulation was set to be 20, merely for convenience, leading to a total population size of approximately 500 individuals.

The model was coded and run in Python 2.7.12 using NumPy and MatPlotLib packages.

3.3.2 Parameters

Many of the parameters were drawn from existing estimates in the literature, as described in the subsections below. The parameters themselves are summarized in table 3.1.

Within Group Transmission Probability p_g

Transmission rates within the group is difficult to draw from the literature since the situation and model vary so widely. The model assumes one contact between every

Parameter	Symbol	Value	Unit
Within group transmission probability	p_g	0.002739726	day^{-1}
Between group transmission probability	p_b	0.0002739726	day^{-1}
Probability of becoming immediately active	p_a	See text	day^{-1}
Probability of transitioning from latent to active	a	0.00027397	day^{-1}
Latent mortality rate	m_l	0.00010140	day^{-1}
Active mortality rate	m_a	0.000169	day^{-1}

Table 3.1: The best estimate of the parameter values used in the model. See subsequent discussion for references and estimates of variability in these parameters.

individual per day, however clearly some individuals will be physically close more than once per day. Liu *et al.* (2010) estimated the transmission rate to be 0.5905 per contact between susceptible and infected individual, normalized by the size of the population. This would put the transmission rate at

$$1 - \exp\left(\frac{\ln(1 - 0.5905)}{365}\right) \approx 0.00244. \tag{3.1}$$

Outside of a formal set of circumstances defining contact, it is essentially impossible to obtain a single value for the transmissibility of an infectious disease. The more or less arbitrary value of $\frac{1}{365} = 0.002739726$ was chosen, and additional simulations were performed at $\frac{10}{365} = 0.02739726$ and $\frac{0.1}{365} = 0.002739726$, to ensure the same effect was observed.

Between Group Transmission Probability p_b

Between group transmission suffers from the same problems as within group transmission. Hunter-gatherer populations generally come into contact with members of other bands at much lower rates. As such, the transmission rate between groups was chosen to be one tenth of the within group transmission rate.

Probability of Becoming Immediately Active p_a

Several sources (Liu *et al.*, 2010; Ziv *et al.*, 2001) have used the figure of 5% of new cases become active within a year. The method used to arrive at this estimate is not clear, however. Another source estimated 2.4% of new infections become active within 5 years (Mancuso *et al.*, 2016; Sloot *et al.*, 2014). This modeling effort varies this parameter systematically, however, so a precise estimate is not needed. The literature supports a relatively small number of new infections becoming active, but a precise estimate is not needed.

The structure of this model also does not include a class of individuals who have been recently infected and will shortly become infectious, but are still latent. In the usage in the model, latent refers to generally long-term latent.

Probability of Transitioning from Latent to Active a

The reactivation of latent TB poses a significant source of new active TB cases in the United States and globally (Ai *et al.*, 2016; World Health Organization, 2017). Estimates of the frequency with which an individual transitions from latent to active is 5-10% over the course of their life. Using the 10% figure and a 73 year lifespan, the parameter estimate is

$$1 - \exp\left(\frac{\ln(1 - 0.10)}{365 * 73}\right) \approx 0.000004. \tag{3.2}$$

A wide range of factors influence the rate at which reactivation occurs (Ai *et al.*, 2016), many of which would be different in a hunter-gather population. Most notably, coinfections might increase the rate of transitioning to active TB, although this is speculative. Nevertheless, that is the rationale for choosing the higher end of the range. Experiments on this parameter were also performed to determine the

sensitivity of the model to this parameter.

Latent Mortality Rate m_l

Mortality rate and its corresponding lifespan, and other life history parameters in hunter-gatherer populations is probably key to understanding many facets of human evolution (Kaplan *et al.*, 2000), however there are few reliable studies on pre-contact mortality rates due to the challenges of collecting such data. The Hiwi, a hunter-gatherer community in Venezuela, has been studied pre-contact using extensive anthropological interviews (Hill *et al.*, 2007).

Estimates of pre-contact mortality rates show an estimated lifespan of 27 years among the Hiwi, which is used to set the parameter. The data also show a high infant mortality rate, so a separate simulation was performed with an estimated lifespan of 50 years, corresponding approximately to the lifespan at age 15 among the Hiwi, and the same as in other work (Ziv *et al.*, 2001).

In a stable population, birth rate and mortality rate should approximately balance, birth rate is set to be the latent mortality rate, although it only

Active Mortality Rate m_a

Active TB, unlike latent TB, causes a higher risk of death. Liu *et al.* (2010) estimated the disease-induced death rate to be 0.06 per year, which works out to be

$$1 - \exp\left(\frac{\ln(1 - 0.0.06)}{365}\right) \approx 0.0001695. \tag{3.3}$$

This is similar to the values found in other sources (Liu *et al.*, 2010).

3.3.3 Calculation of Active R_0

The classical meaning of R_0 is the number of new infections caused by an infected individual. However for the purposes of predicting the eventual outcome of an infectious disease, latent cases which do not transition to be active should not be counted. As such, a slightly different definition of R_0 is employed here, termed "active R_0," defined as the number of new active cases of TB caused by an infection in an otherwise entirely susceptible population.

First, it is necessary to find the probability that the infection is active or becomes active before the individual becomes active before dying. The probability a new infection is active initially is p_a, one of the model parameters. The probability it becomes active after x days is $(1-m_l)^x(1-a)^{x-1}a$. Summing over all x using the geometric series formula yields the following.

$$\sum_{x=1}^{\infty}(1-m_l)^x(1-a)^{x-1}a = a(1-m_l)\sum_{x=1}^{\infty}\left((1-m_l)(1-a)\right)^{x-1} \tag{3.4}$$

$$= \frac{a(1-m_l)}{1-(1-m_l)(1-a)} \tag{3.5}$$

Thus the probability that a newly infected individual is either active or becomes active is given by the following equation.

$$p_a + (1-p_a)\frac{a(1-m_l)}{1-(1-m_l)(1-a)} \tag{3.6}$$

Given that an individual has an active infection, the new infections they cause can be either within the band or in another band. While ordinary R_0 does not distinguish between secondary latent infections and secondary active infections, active R_0 only considers those secondary infections which are active or become active at some point. If an individual has active TB, they have an expected lifespan of $\frac{1}{m_a}$ days. Each individual in the band has a p_g probability of becoming infected per day, and there

are 24 individuals in the band. So the expected number of new infections in the band from this one individual is $\frac{1}{m_a}(24)p_g$. Of these new infections, p_a are active and $1-p_a$ are latent. So the new within-band active infections and, relying on equation 3.6, is

$$\frac{1}{m_a}(24)p_g\left(p_a+(1-p_a)\frac{a(1-m_l)}{1-(1-m_l)(1-a)}\right). \tag{3.7}$$

An infected individual also has a possibility of infecting someone outside of their own band. In each time step, there is a p_b probability of infecting each of 25 people in another band. This means on average, the expected number of infections per day is $25p_b$, so over the expected life of the active TB case, there will be $\frac{25p_b}{m_a}$ new TB cases in other groups. Not all of these are active, however, so again, the result from equation 3.6 is used. The number of new active cases caused in other bands is therefore

$$\frac{25p_b}{m_a}\left(p_a+(1-p_a)\frac{a(1-m_l)}{1-(1-m_l)(1-a)}\right) \tag{3.8}$$

Adding equation 3.7 to 3.8 yields the number of new cases caused by an active case of TB. This, multiplied by the probability the first case is or becomes an active case, given by equation 3.6, yields the formula for active R_0.

$$\left(\frac{24p_g}{m_a}+\frac{25p_b}{m_a}\right)\left(p_a+(1-p_a)\frac{a(1-m_l)}{1-(1-m_l)(1-a)}\right)^2 \tag{3.9}$$

3.4 Results

3.4.1 Main Result

The principle result follows from increasing the percent of new infections which are active from 10% to 100% in increments of 10%. In actual tuberculosis infections, this percent is rather small (see discussion above). As the percent of new active infections

increases, the end size of the epidemic decreases almost linearly. This is shown in figure 3.1.

Simulations were generally fairly consistent for each set of parameter values. Failed epidemics (simulations in which TB failed to spread at all) were rare. The maximum size of the epidemic was measured and a similar but smaller effect was seen: as the percent active increases, the maximum size of the epidemic decreased.

One important factor to note is the decrease in the overall population. Average total population in all categories decreased as the percent active increased.

3.4.2 Active R_0

The traditional definition of R_0 does not differentiate between latent and active infections, however it is useful to do so for the purposes of examining the potential for an epidemic. The ten values of p_a each provide a different value for active R_0, shown in table 3.2. Note that the traditional definition of R_0 is identical for every value of p_a.

3.4.3 Parameter Sensitivity

To assess the robustness of this result to parameter misspecification, the model was run 300 times at each of the following parameters.

3.5 Discussion

The transmission pattern of TB poses an interesting quandary: why is it a latent infection in many people? An infection which always becomes infectious is, in most circumstances, going to spread further. One possible explanation for the high

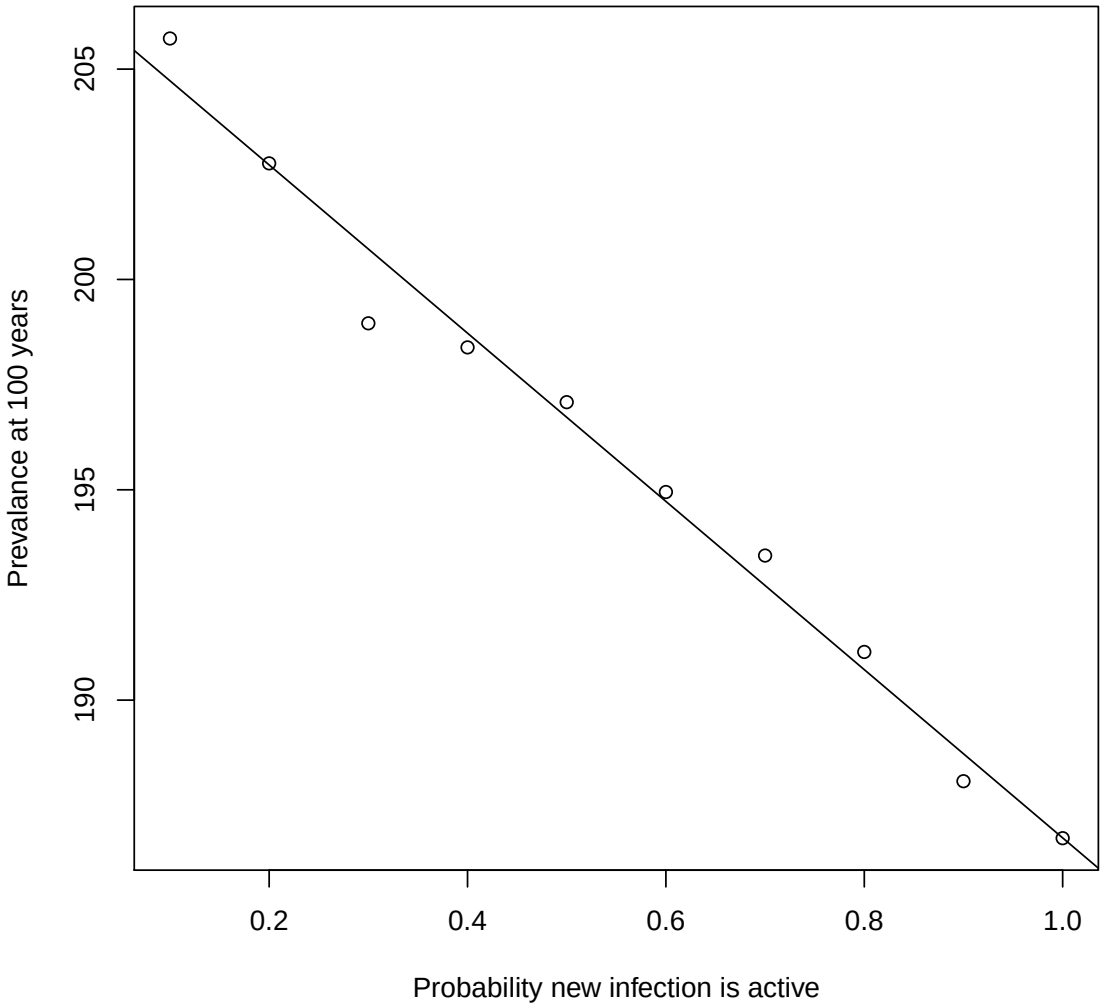

Figure 3.1: As the percent of active infections increases, the average size of the outbreak at 100 years decreases. Each point on this graph is the average of 300 simulations of the model with the same set of parameters.

p_a	R_0
0.1	235.88
0.2	253.02
0.3	270.76
0.4	289.10
0.5	308.04
0.6	327.59
0.7	347.73
0.8	368.48
0.9	389.82
1.0	411.77

Table 3.2: Active R_0 values for different values of p_a using the model parameters found above. These R_0 values indicate very high rates of transmission.

frequency of latent infections is simply the effectiveness of the human immune system. However an alternative explanation which is consistent with the data is that TB evolved to have a high rate of latent infection to be more effective at transmission on the unique network structure of hunter-gatherers.

Although there are exceptions, increases in parameters affecting R_0, in most network structures and circumstances, are predicted to increase the final size of an epidemic. This project uses realistic model parameters and a simple stochastic model to demonstrate that increasing p_a, the probability a new infection is active, increases the active R_0 while decreasing the final size of the epidemic. This may be suggestive of an evolutionary adaptation of TB to hunter-gatherer communities, which itself indicates a relatively distant origin of TB.

The TB complex of bacterial species infects a wide range of mammalian hosts. Bos *et al.* (2014) find a relatively recent origin of TB, finding a most recent common

	p_g	p_b	a	m_l	m_a	Mean $p_a = 0.1$	Mean $p_a = 0.3$
Baseline	0.00274	0.000274	0.10	0.000274	0.000101	204.11	201.01
Low p_g	0.000274	0.000274	0.10	0.000274	0.000101	197.95	196.33
High p_g	0.0274	0.000274	0.10	0.000274	0.000101	205.87	201.67
Low p_b	0.00274	0.0000274	0.10	0.000274	0.000101	203.78	201.04
High p_b	0.00274	0.00274	0.10	0.000274	0.000101	205.23	201.66
Low a	0.00274	0.000274	0.10	0.0000274	0.000101	227.56	228.15
High a	0.00274	0.000274	0.10	0.00274	0.000101	188.50	186.20
Low m_a	0.00274	0.000274	0.10	0.000274	0.000101	411.82	416.37
High m_a	0.00274	0.000274	0.10	0.000274	0.00101	100.59	88.94

Table 3.3: Testing the sensitivity of the parameters involved estimating the effect at a 10x increase and 10x decrease of each of the constant parameters.

ancestor date for human TB strains in the mid-Holocene, and being derived from pinniped strains. This is a similar time frame to the proposal that human TB originated with the domestication of cattle, however it is not consistent with the proposed non-human source, *Mycobacterium bovis* (Stead *et al.*, 1995).

If the modeling results indeed indicate an adaptation to hunter-gatherer contact networks, these late introductions of TB would appear to be incorrect. Others have suggested a much older, potentially ancient date for the introduction of TB to humans (Kappelman *et al.*, 2007; Roberts *et al.*, 2009). Since the hunter-gatherer lifestyle became common in the Holocene, but was exceedingly rare prior to the Holocene, it is unclear how old of a date the results support.

Spigelman *et al.* (2015) has proposed that the long latent period is the result of a long existence in humans, and the resulting evolution of the human immune system to cope with TB. This is at least consistent with the results of this project, however the focus of this project is on the benefits that the human immune response provides to the TB pathogen. By suppressing the pathogen, the human immune response may have actually provided an optimal method for transmission in the human hunter-gatherer social network.

This project also has some significant limitations, and raises several additional questions. From the modeling perspective, it is surprising that the R_0 values were as high as they were found to be. This may suggest an error or problem within the model or the parameter values. The values in this project are certainly well-above most published values Blower *et al.* (1995); Liu *et al.* (2010), which are generally between 1 and 2.

This project also did not look at changing the pathogenicity of a TB infection, even though the evolution of pathogenicity is widely discussed Frank (1996). Thus it is possible that the results in this model are the result of the well-known evolutionary

trajectory toward lower pathogenicity of infections Bull (1994), being driven indirectly by lowered infectivity rates.

Finally, the value of a model in the investigation of a hypothesis is ultimately driven by the confidence in the model and its accuracy in capturing the system. Definitive answers to questions on the origin of TB in humans will almost certainly driven by compelling data, rather than modeling. Nevertheless, modeling can provide some insight.

3.6 Conclusion

This modeling effort shows an inverse relationship between the frequency with which new TB infections become immediately active and the long term eventual size of the infection. Although higher R_0 may be associated with more widespread infections in a highly connected social network, it is not necessarily the case on a hunter-gatherer network. If this is the result of disease evolution, or human-disease co-evolution, this pattern seems to be consistent with a pre-Holocene introduction of TB to humans. However it is not definitive.

More broadly, this work demonstrates the importance of social networks to disease evolution. Diseases may evolve transmission parameters or patterns which are uniquely adapted to a particular contact network among the hosts. Identifying those patterns may yet yield more insight into the evolution of disease.

Chapter 4

CONCLUSION

The social interactions of humans are central to human behavior, and as the results of this dissertation demonstrate, play an important role in determining the spread of infectious disease in an epidemic. This dissertation is an examination of the effects of social networks on the transmission of infectious disease. The first project examined SARS, demonstrating that social networks–specifically superspreading on social networks–plays an important role in determining the case counts. This demonstration illustrates how social networks can be a confounding effect on the evaluation of infectious disease epidemics. The second project in this dissertation demonstrates that TB easily persists in hunter-gatherer populations, and may have evolved characteristics which facilitate spread in such a population. This suggests an ancient origin to TB in humans. Both of these projects illustrate the potential that mathematical tools can bring to understanding the causes and consequences of human variation, particular in social network structure.

This dissertation is an illustration of the unique role that an anthropologist can play in uncovering the dynamics of infectious disease transmission (DiGiacomo, 1999; Trostle and Sommerfeld, 1996). Social networks, firmly rooted in the study of human behavior, are an integral part of identifying the epidemic potential of an infectious pathogen and potentially an important influence on the evolution of a pathogen. Models of this sort have been used in a similar manner before; Khanna *et al.* (2014) and Rothenberg *et al.* (1998), for example, both investigate HIV dynamics using network models. Anthropologists contributed to both works, demonstrating the importance of this intersection between infectious disease epidemiology and social anthropology.

This dissertation is, in the same vein, a demonstration of the power that mathematical models of infectious disease can have in the investigation of the causes and consequences of social phenomena. Chapter 2 illustrates one of the consequences by demonstrating the social network can give rise to the superspreading phenomenon, which in turn can create relatively large subcritical epidemics. Chapter 3 by contrast shows that evolution of an infectious agent on a particular social network can result in a particular signature. In the case of TB, that signature appears to show evolution on a hunter-gatherer social network.

Broadly these projects demonstrate the wealth of knowledge which can be gained from an examination of infectious disease transmission in an anthropological context.

4.1 Anthropological Conclusions

Investigation into the spread of infectious disease, absent an understanding of social networks, may have some problems. Chapter 2 demonstrates that estimates of R_0 which ignore superspreading may be in error, and thus epidemiological data not be able to distinguish between a supercritical and subcritical epidemics. The main result in chapter 3 demonstrates that the structure of a social network can influence the evolution of an infectious disease. This can, then, be used as a signature to identify the social network on which a pathogen evolved. Both of these illustrate how social networks, mediated through mathematical models, cause important changes to the dynamics of infectious disease transmission.

Understanding the impact of social phenomena such as infectious disease contact networks is not new to anthropology (Inhorn and Brown, 1990; Klovdahl, 1985). Superspreading is one possibility for a social network structure. Superspreading is a particular pattern of heterogeneity in contacts or transmission rates in which a relatively few infected individuals account for a disproportionate number of secondary

cases (Paull *et al.*, 2012). The consequences for this pattern can be subcritical epidemics which appear to be supercritical. It is apparent from this work that epidemiological practice is dependent on understanding the anthropological phenomenon of social network structure.

The impact of social networks can also be seen in the evolution of infectious disease (Leventhal *et al.*, 2015). The timing of the introduction of tuberculosis to the human population is not well understood (Bos *et al.*, 2014). However this project adds a line of evidence showing that TB seems to have evolved properties which are conducive to transmission on hunter-gatherer social networks, suggesting a long period of co-evolution with humans and thus an early introduction.

These two projects underscore the importance of heterogeneity in social networks has for disease transmission dynamics. As anthropology understands the causes and consequences of human diversity, this project highlights the causes and consequences of specific types of social network diversity.

4.2 Infectious Disease Conclusions

Superspreading events force epidemiologists to think about the structure of networks and the context of those networks. The first project in this dissertation demonstrates that a social network can complicate the identification of supercritical and subcritical epidemics. In many epidemics, the goal of public health officials is to implement policies which force $R_0 < 1$, that is make the epidemic subcritical (WHO, 2003a). However the results of the SARS portion of this dissertation cast doubt on whether distinguishing between a supercritical and subcritical epidemic is achievable.

The project examining TB also demonstrates the importance of social networks to epidemiologists. The assumption that diseases might evolve toward a higher R_0 is demonstrably untrue. In hunter-gatherer social networks, it is the result of this

project that a TB can evolve toward lower R_0. This is interesting as a question about the evolution of disease in the past, but also opens the question of how other social networks can influence the evolutionary trajectory of modern infectious diseases.

4.3 Discussion

Broadly this dissertation is descriptive of the influence of social networks on disease transmission. This alone is not a new field of investigation (Inhorn and Brown, 1990; Zelner *et al.*, 2012; Keeling and Eames, 2005; Getz *et al.*, 2006). This project looks separately at the phenomenon of superspreading and disease evolution, both of which have been investigated before (Antia *et al.*, 2003; Lloyd-Smith *et al.*, 2005). However this dissertation is the first quantitative examination of subcritical superspreading, demonstrating the importance of social network structure to the evaluation of infectious disease epidemics. This is also the first attempt to find a signature a social network leaves on the transmission parameters of tuberculosis. This project demonstrates TB likely was introduced to the human population in the distant past, to provide it with a long co-evolutionary history. This is in contrast to some recent results such as Bos *et al.* (2014) which found a recent ancestry, but consistent with the qualitative conclusions of Gagneux (2012) and the timeline set out in Comas *et al.* (2013). Both of these projects add important information to their respective questions, but also have important limitations.

This project fits into a broader field of anthropological research. Medical anthropology has a long history at the intersection of medicine and anthropology (McElroy and Townsend, 1996), however some anthropologists have criticized epidemiologists specifically for under-utilizing anthropological insights (Inhorn, 1995; Ashan, 2016). DiGiacomo (1999) in particular draws a distinction between the intensively quantitative field of epidemiology and the often-qualitative field of social anthropology.

This project, which discusses a social phenomenon in quantitative terms and derives from them quantitative conclusions, is a contribution to a literature which is at that interface. The goal is not, of course, to devalue qualitative research, however by describing social phenomena in quantitative terms, it is possible to derive quantitative conclusions which are of interest to both anthropologists and epidemiologists.

4.3.1 Limitations

Although this dissertation provides insight into several questions, it is important to recognize the limitations of the work. The first project, investigating the subcritical superspreading phenomenon, demonstrates that differentiating between supercritical and subcritical epidemics is challenging given data on the size of epidemics. It is possible that some other data may be used to clearly delineate the difference between supercritical and subcritical epidemics, although such data has not been identified (Guttorp and Perlman, 2015).

Inhorn (1995) points to the richness and depth that anthropologists can bring to epidemiological questions. Traditional epidemiology quantifies humans extensively–from behaviors to risk factors–to better make predictions. While this is an important part of science, Inhorn points to a wealth of data that can be had through anthropological research. While the questions investigated in this dissertation are anthropological, it can certainly be argued that the approach was similar to the reductionist viewpoint Inhorn criticizes in epidemiology.

Specifically on the project related to the evolution of tuberculosis, the results found in chapter 3 are similar to those predicted qualitatively by Gagneux (2012). However the results should be considered in a constellation of evidence–paleopathological, genetic, genomic, and others–to draw a conclusion. Although the results shown here indicate an early introduction of TB to humans, it is possible that such a result will

be overturned by subsequent lines of more direct evidence.

4.4 Conclusion

The lessons that can be drawn from this dissertation are several. The first are the two results of the research, discussed above. However more broadly this dissertation can be seen as an example of how mathematical models can be used to connect epidemiology and anthropology. As with any inquiry into anthropology, it is an investigation into the causes of diversity and their consequences. Here that diversity is in social networks. Social networks remain an important area for research in both anthropology and epidemiology.

REFERENCES

Adjorlolo, G., J. Kouadio, M. Maran and K. M. Vetter, "Epidemiology and Transmission of HIV-2", JAMA **270**, 2083–2086 (1993).

Ai, J.-W., Q.-L. Ruan, Q.-H. Liu and W.-H. Zhang, "Updates on the risk factors for latent tuberculosis reactivation and their managements", Emerging Microbes & Infections **5**, 2, e10 (2016).

Allison, M. J., D. Mendoza and A. Pezzia, "Documentation of a Case of Tuberculosis in Pre-Columbian America", Annual Review of Respiratory Disease **107**, 985–991 (1973).

Anderson, R. M. and R. M. May, *Infectious diseases of humans: dynamics and control* (Oxford Univ. Press, Oxford, 2010), reprinted edn., oCLC: 844937291.

Antia, R., R. R. Regoes, J. C. Koella and C. T. Bergstrom, "The role of evolution in the emergence of infectious diseases", Nature **426**, 6967, 658–661 (2003).

Armbruster, B. and M. L. Brandeau, "Contact tracing to control infectious disease: when enough is enough", Health care management science **10**, 4, 341–355 (2007).

Ashan, A., "The relationship between epidemiology and anthropology", (2016).

Aunger, R., ed., *Darwinizing culture: the status of memetics as a science* (Oxford University Press, Oxford ; New York, 2000).

Bailey, N. T. J., *The Mathematical Theory of Epidemics* (Hafner Publishing Company, New York, 1957).

Bartko, J. J., "A Note on the Negative Binomial Distribution", Technometrics **4**, 4, 609–610 (1962).

Barton, N. H., ed., *Evolution* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y, 2007), oCLC: ocm86090399.

Bearman, P., J. Moody and K. Stovel, "Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks", American Journal of Sociology **110**, 1, 44–91 (2004).

Becker, N., "On Parametric Estimation for Mortal Branching Processes", Biometrika **61**, 3, 393–399 (1974).

Becker, N., "Estimation for Discrete Time Branching Processes with Application to Epidemics", Biometrics **33**, 3, 515–522 (1977).

Blackmore, S., "Evolution and Memes: The Human Brain as a Selective Imitation Device", Cybernetics and Systems **32**, 1-2, 225–255 (2001).

Blackmore, S., "Memes shape brains shape memes", Behavioral and Brain Sciences; New York **31**, 5, 513–513 (2008).

Blower, S. M., A. R. McLean, T. C. Porco, P. M. Small, P. C. Hopewell, M. A. Sanchez and A. R. Moss, "The intrinsic transmission dynamics of tuberculosis epidemics", Nature Medicine **4**, 8, 815–821 (1995).

Blumberg, S. and J. Lloyd-Smith, "Comparing methods for estimating R0 from the size distribution of subcritical transmission chains", Epidemics **5**, 3, 131–145 (2013a).

Blumberg, S. and J. O. Lloyd-Smith, "Inference of R0 and Transmission Heterogeneity from the Size Distribution of Stuttering Chains", PLoS Computational Biology **9**, 5, e1002993 (2013b).

Bond, V. and L. Nyblade, "The importance of addressing the unfolding TB-HIV stigma in high HIV prevalence settings", Journal of Community & Applied Social Psychology **16**, 6, 452–461 (2006).

Bos, K. I., "TBs Chinese travels", Nature Ecology & Evolution **2**, 12, 1842–1843 (2018).

Bos, K. I., K. M. Harkins, A. Herbig, M. Coscolla, N. Weber, I. Comas, S. A. Forrest, J. M. Bryant, S. R. Harris, V. J. Schuenemann, T. J. Campbell, K. Majander, A. K. Wilbur, R. A. Guichon, D. L. Wolfe Steadman, D. C. Cook, S. Niemann, M. A. Behr, M. Zumarraga, R. Bastida, D. Huson, K. Nieselt, D. Young, J. Parkhill, J. E. Buikstra, S. Gagneux, A. C. Stone and J. Krause, "Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis", Nature **514**, 7523, 494–497 (2014).

Bott, E., *Family and Social Network: Roles, Norms, External Relationships in Ordinary Urban Families* (Free Press, New York, NY, 1971), 2nd ed. edn.

Boyd, R. and P. J. Richerson, *Culture and the evolutionary process* (University of Chicago Press, Chicago, 1988), paperback edn.

Brauer, F., "Mathematical epidemiology: Past, present, and future", Infectious Disease Modelling **2**, 2, 113–127 (2017).

Brauer, F., P. Van den Driessche, J. Wu and L. J. S. Allen, eds., *Mathematical epidemiology*, no. 1945 in Mathematical biosciences subseries (Springer, Berlin, 2008), oCLC: ocn229446076.

Braveman, P., S. Egerter and D. R. Williams, "The Social Determinants of Health: Coming of Age", Annual Review of Public Health **32**, 1, 381–398 (2011).

Breugelmans, J. G., P. Zucs, K. Porten, S. Broll, M. Niedrig, A. Ammon and G. Krause, "SARS Transmission and Commercial Aircraft", Emerging Infectious Diseases **10**, 8, 1502–1503 (2004).

Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen and S. T. Cole, "A new evolutionary scenario for

the *Mycobacterium tuberculosis* complex", Proceedings of the National Academy of Sciences **99**, 6, 3684–3689 (2002).

Brown, H. and A. H. Kelly, "Material Proximities and Hotspots: Toward an Anthropology of Viral Hemorrhagic Fevers: Material Proximities and Hotspots", Medical Anthropology Quarterly **28**, 2, 280–303 (2014).

Bull, J. J., "Perspective: Virulence", Evolution **48**, 5, 1423–1437 (1994).

Casella, G. and R. L. Berger, *Statistical inference* (Cengage Learning & Wadsworth, Pacific Grove, CA, 2002), oCLC: 905616056.

Castellano, C. and R. Pastor-Satorras, "Thresholds for Epidemic Spreading in Networks", Physical Review Letters **105**, 21 (2010).

Cavalli-Sforza, L. L. and M. W. Feldman, *Cultural transmission and evolution: a quantitative approach*, no. 16 in Monographs in population biology (Princeton University Press, Princeton, N.J, 1981).

CDC, "Severe Acute Respiratory Syndrome — Taiwan, 2003", Morbidity and Mortality Weekly Report **52**, 20, 461–466 (2003).

Chambers, H. F., "The changing epidemiology of Staphylococcus aureus?", Emerging Infectious Diseases **7**, 2, 178–182 (2001).

Chen, M. I. and Y.-S. Leo, "Understanding the Super-spreading Events of SARS in Singapore", **35**, 6, 5 (2006).

Chowell, G. and F. Brauer, "The Basic Reproduction Number of Infectious Diseases: Computation and Estimation Using Compartmental Epidemic Models", in "Mathematical and Statistical Estimation Approaches in Epidemiology", edited by G. Chowell, J. M. Hyman, L. M. A. Bettencourt and C. Castillo-Chavez, pp. 1–30 (Springer Netherlands, Dordrecht, 2009).

Chowell, G., C. Castillo-Chavez, P. W. Fenimore, C. M. Kribs-Zaleta, L. Arriola and J. M. Hyman, "Model Parameters and Outbreak Control for SARS", Emerging Infectious Diseases **10**, 7, 1258–1263 (2004).

Chun, B. C., "Understanding and Modeling the Super-spreading Events of the Middle East Respiratory Syndrome Outbreak in Korea", Infection & Chemotherapy **48**, 2, 147–149 (2016).

Clark, G. A., M. A. Kelley, J. M. Grange, M. C. Hill, M. A. Katzenberg, L. L. Klepinger, N. C. Lovell, J. McGrath, M. S. Micozzi and R. T. Steinbock, "The Evolution of Mycobacterial Disease in Human Populations: A Reevaluation [and Comments and Reply]", Current Anthropology **28**, 1, 45–62 (1987).

Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. B. Iii, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy,

K. Oliver, J. Osborne, M. A. Quail, M.-A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead and B. G. Barrell, "Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence", Nature **393**, 6685, 537–544 (1998).

Comas, I., M. Coscolla, T. Luo, S. Borrell, K. E. Holt, M. Kato-Maeda, J. Parkhill, B. Malla, S. Berg, G. Thwaites, D. Yeboah-Manu, G. Bothamley, J. Mei, L. Wei, S. Bentley, S. R. Harris, S. Niemann, R. Diel, A. Aseffa, Q. Gao, D. Young and S. Gagneux, "Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans", Nature Genetics **45**, 10, 1176–1182 (2013).

Cremer, S., S. A. O. Armitage and P. Schmid-Hempel, "Social Immunity", Current Biology **17**, 16, R693–R702 (2007).

Cronk, L., N. A. Chagnon and W. Irons, eds., *Adaptation and human behavior: an anthropological perspective*, Evolutionary foundations of human behavior (Aldine de Gruyter, New York, 2000).

Cushing, J. M. and O. Diekmann, "The many guises of R0 (a didactic note)", Journal of Theoretical Biology **404**, 295–302 (2016).

Daszak, P., A. A. Cunningham and A. D. Hyatt, "Emerging Infectious Diseases of Wildlife– Threats to Biodiversity and Human Health", Science **287**, 5452, 443–449 (2000).

Davies, C. R., J. M. Ayres, C. Dye and L. M. Deane, "Malaria Infection Rate of Amazonian Primates Increases with Body Weight and Group Size", Functional Ecology **5**, 5, 655–662 (1991).

Dawkins, R., *The selfish gene* (Oxford University Press, Oxford ; New York, 1989), new ed edn.

Day, S., "Sexual behaviour and networking: anthropological and socio-cultural studies on the transmission of HIV", Journal of Epidemiology and Community Health **48**, 2, 214 (1994).

De Serres, G., N. J. Gay and C. P. Farrington, "Epidemiology of Transmissible Diseases after Elimination", American Journal of Epidemiology **151**, 11, 1039–1048 (2000).

Didelot, X., A. S. Walker, T. E. Peto, D. W. Crook and D. J. Wilson, "Within-host evolution of bacterial pathogens", Nature reviews. Microbiology **14**, 3, 150–162 (2016).

Diekmann, O., J. A. P. Heesterbeek and J. A. J. Metz, "On the definition and the computation of the basic reproductive ratio R_0 in models for infectious diseases in heterogeneous populations", Journal of Mathematical Biology **28**, 365–382 (1990).

DiGiacomo, S. M., "Can There Be a 'Cultural Epidemiology'?", Medical Anthropology Quarterly **13**, 4, 436–457 (1999).

Din-Dzietham, R., W. N. Nembhard, R. Collins and S. K. Davis, "Perceived stress following race-based discrimination at work is associated with hypertension in African-Americans. The metro Atlanta heart disease study, 19992001", Social Science & Medicine **58**, 3, 449–461 (2004).

Dobson, A. J., *An introduction to generalized linear models*, Chapman & Hall/CRC texts in statistical science series (Chapman & Hall/CRC, Boca Raton, 2002), 2nd ed edn.

Donnelly, C. A., A. C. Ghani, G. M. Leung, A. J. Hedley, C. Fraser, S. Riley, L. J. Abu-Raddad, L.-M. Ho, T.-Q. Thach, P. Chau, K.-P. Chan, T.-H. Lam, L.-Y. Tse, T. Tsang, S.-H. Liu, J. H. Kong, E. M. Lau, N. M. Ferguson and R. M. Anderson, "Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong", The Lancet **361**, 9371, 1761–1766 (2003).

Dressler, W. W., "Hypertension and Perceived Stress: A St. Lucian Example", Ethos **12**, 3, 265–283 (1984).

Durkheim, E., "What is a social fact?", in "Anthropological History: An Introductory History", pp. 87–93 (Mayfield Publishing, Mountain View, CA, 2000), 2nd ed edn.

Eames, K. T. D. and M. J. Keeling, "Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases", Proceedings of the National Academy of Sciences of the United States of America **99**, 20, 13330–13335 (2002).

Farrington, C. P. and A. D. Grant, "The Distribution of Time to Extinction in Subcritical Branching Processes: Applications to Outbreaks of Infectious Disease", Journal of Applied Probability **36**, 3, 771–779 (1999).

Feller, W., "On a General Class of "Contagious" Distributions", The Annals of Mathematical Statistics **14**, 4, 389–400 (1943).

Fenner, L., M. Egger, T. Bodmer, H. Furrer, M. Ballif, M. Battegay, P. Helbling, J. Fehr, T. Gsponer, H. L. Rieder, M. Zwahlen, M. Hoffmann, E. Bernasconi, M. Cavassini, A. Calmy, M. Dolina, R. Frei, J.-P. Janssens, S. Borrell, D. Stucki, J. Schrenzel, E. C. Bttger, S. Gagneux and f. t. S. H. C. a. M. E. o. T. S. Groups, "HIV Infection Disrupts the Sympatric Host-Pathogen Relationship in Human Tuberculosis", PLOS Genetics **9**, 3, e1003318 (2013).

Frank, S. A., "Models of Parasite Virulence", The Quarterly Review of Biology **71**, 1, 37–78 (1996).

Fujie, R. and T. Odagaki, "Effects of superspreaders in spread of epidemic", Physica A: Statistical Mechanics and its Applications **374**, 2, 843–852 (2007).

Gagneux, S., "Host-pathogen coevolution in human tuberculosis", Philosophical Transactions of the Royal Society B: Biological Sciences **367**, 1590, 850–859 (2012).

Gagneux, S., "Ecology and evolution of Mycobacterium tuberculosis", Nature Reviews Microbiology **16**, 4, 202–213 (2018).

Galvani, A. P. and R. M. May, "Dimensions of superspreading", Nature **438**, 293–295, 3 (2005).

Garske, T. and C. Rhodes, "The effect of superspreading on epidemic outbreak size distributions", Journal of Theoretical Biology **253**, 2, 228–237 (2008).

Getz, W. M., J. O. Lloyd-Smith, P. C. Cross, S. Bar-David, P. L. Johnson, T. C. Porco and M. S. Sanchez, "Modeling the Invasion and Spread of Contagious Diseases in Heterogeneous Populations", in "DIMACS: Series in Discrete Mathematics and Theoretical Computer Science", (American Mathematical Society, 2006).

Gluckman, P. D., A. Beedle and M. A. Hanson, *Principles of evolutionary medicine*, Oxford biology (Oxford University Press, Oxford ; New York, 2009), oCLC: ocn313664861.

Gopinath, S., J. S. Lichtman, D. M. Bouley, J. E. Elias and D. M. Monack, "Role of disease-associated tolerance in infectious superspreaders", Proceedings of the National Academy of Sciences **111**, 44, 15780–15785 (2014).

Greenwood, M. and G. U. Yule, "An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents", Journal of the Royal Statistical Society **83**, 2, 255 (1920).

Grenfell, B. T., P. K. Das, P. K. Rajagopalan and D. A. Bundy, "Frequency distribution of lymphatic filariasis microfilariae in human populations: population processes and statistical estimation", Parasitology **101 Pt 3**, 417–427 (1990).

Gu, J., S. Lee, J. Saramki and P. Holme, "Ranking influential spreaders is an ill-defined problem", EPL (Europhysics Letters) **118**, 6, 68002, arXiv: 1703.05644 (2017).

Gurland, J., "Some Applications of the Negative Binomial and Other Contagious Distributions", American Journal of Public Health and the Nations Health **49**, 10, 1388–1399 (1959).

Gustafsson, L., "Poisson SimulationA Method for Generating Stochastic Variations in Continuous System Simulation", Simulation **74**, 5, 264–274 (2000).

Gustafsson, L. and M. Sternad, "Bringing consistency to simulation of population models Poisson Simulation as a bridge between micro and macro simulation", Mathematical Biosciences **209**, 2, 361–385 (2007).

Gutierrez, M. C., S. Brisse, R. Brosch, M. Fabre, B. Omas, M. Marmiesse, P. Supply and V. Vincent, "Ancient Origin and Gene Mosaicism of the Progenitor of Mycobacterium tuberculosis", PLOS Pathogens **1**, 1, e5 (2005).

Guttorp, P. and M. D. Perlman, "Predicting extinction or explosion in a GaltonWatson branching process with power series offspring distribution", Journal of Statistical Planning and Inference **167**, 193–215 (2015).

Harris, T. E., *The Theory of Branching Processes*, vol. 119 of *Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen* (1963).

Hassell, J. M., M. Begon, M. J. Ward and E. M. Fvre, "Urbanization and Disease Emergence: Dynamics at the WildlifeLivestockHuman Interface", Trends in Ecology & Evolution **32**, 1, 55–67 (2017).

Heesterbeek, J. A. P., "A brief history of R0 and a recipe for its calculation", Acta Biotheoretica **50**, 189–204 (2002).

Herring, A. and A. C. Swedlund, eds., *Plagues and epidemics: infected spaces past and present*, Wenner-Gren international symposium series (Berg, Oxford ; New York, 2010), english ed edn., oCLC: ocn457155512.

Hethcote, H. W., J. A. Yorke and A. Nold, "Gonorrhea modeling: a comparison of control methods", Mathematical Biosciences **58**, 1, 93–109 (1982).

Hill, K., M. Barton and A. M. Hurtado, "The emergence of human uniqueness: Characters underlying behavioral modernity", Evolutionary Anthropology: Issues, News, and Reviews **18**, 5, 187–200 (2009).

Hill, K., A. Hurtado and R. Walker, "High adult mortality among Hiwi hunter-gatherers: Implications for human evolution", Journal of Human Evolution **52**, 4, 443–454 (2007).

Hill, K. and A. M. Hurtado, *Ache life history: the ecology and demography of a foraging people*, Foundations of human behavior (Aldine de Gruyter, New York, 1996).

Hill, K. R., R. S. Walker, M. Bozicevic, J. Eder, T. Headland, B. Hewlett, A. M. Hurtado, F. Marlowe, P. Wiessner and B. Wood, "Co-Residence Patterns in Hunter-Gatherer Societies Show Unique Human Social Structure", Science **331**, 6022, 1286–1289 (2011).

Hirsh, A. E., A. G. Tsolaki, K. DeRiemer, M. W. Feldman and P. M. Small, "Stable association between strains of Mycobacterium tuberculosis and their human host populations", Proceedings of the National Academy of Sciences of the United States of America **101**, 14, 4871–4876 (2004).

Hsieh, Y.-H., C. W. Chen and S.-B. Hsu, "SARS Outbreak, Taiwan, 2003", Emerging Infectious Diseases **10**, 2, 201–206 (2004).

Hsieh, Y.-H., C.-C. King, C. W. Chen, M.-S. Ho, S.-B. Hsu and Y.-C. Wu, "Impact of quarantine on the 2003 SARS outbreak: A retrospective modeling study", Journal of Theoretical Biology **244**, 4, 729–736 (2007).

Hurtado, A. M., K. R. Hill, W. Rosenblatt, J. Bender and T. Scharmen, "Longitudinal study of tuberculosis outcomes among immunologically naive Ach natives of Paraguay: Longitudinal Study of Tuberculosis Among Ach", American Journal of Physical Anthropology **121**, 2, 134–150 (2003).

Inhorn, M. C., "Medical anthropology and epidemiology: Divergences or convergences?", Social Science & Medicine **40**, 3, 285–290 (1995).

Inhorn, M. C. and P. J. Brown, "The Anthropology of Infectious Disease", Annual Review of Anthropology **19**, 1, 89–117 (1990).

Jacob, C., "Branching Processes: Their Role in Epidemiology", International Journal of Environmental Research and Public Health **7**, 3, 1186–1204 (2010).

Jacobsen, K. H., *Introduction to global health* (Jones and Bartlett Pub, Sudbury, Mass, 2008).

Jones, P. C. T., J. E. Mollison and M. H. Quenouille, "A technique for the quantitative estimation of soil micro-organisms", Journal of General Microbiology **2**, 54–69 (1948).

Kaplan, H., K. Hill, J. Lancaster and A. M. Hurtado, "A theory of human life history evolution: Diet, intelligence, and longevity", Evolutionary Anthropology: Issues, News, and Reviews **9**, 4, 156–185 (2000).

Kappelman, J., M. C. Aliek, N. Kazanc, M. Schultz, M. zkul and . en, "First Homo erectus from Turkey and implications for migrations into temperate Eurasia", American Journal of Physical Anthropology **135**, 1, 110–116 (2007).

Keeling, M. J. and K. T. Eames, "Networks and epidemic models", Journal of the Royal Society Interface **2**, 4, 295–307 (2005).

Kermack, W. O. and A. G. McKendrick, "A Contribution to the Mathematical Theory of Epidemics", Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **115**, 772, 700–721 (1927).

Keyfitz, B. L. and N. Keyfitz, "The McKendrick partial differential equation and its uses in epidemiology and population study", Mathematical and Computer Modelling **26**, 6, 1–9 (1997).

Khan, H. A., F. K. Baig and R. Mehboob, "Nosocomial infections: Epidemiology, prevention, control and surveillance", Asian Pacific Journal of Tropical Biomedicine **7**, 5, 478–482 (2017).

Khanna, A. S., D. T. Dimitrov and S. M. Goodreau, "What can mathematical models tell us about the relationship between circular migrations and HIV transmission dynamics?", Mathematical Biosciences and Engineering **11**, 5, 1065–1090 (2014).

Kimmel, M. and D. E. Axelrod, *Branching processes in biology*, no. volume 19 in Interdisciplinary applied mathematics (Springer, New York, 2015), second edition edn., oCLC: ocn910273492.

Kingsland, *Modeling Nature* (University of Chicago Press, 1995).

Kitsak, M., L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley and H. A. Makse, "Identification of influential spreaders in complex networks", Nature Physics **6**, 11, 888–893 (2010).

Klovdahl, A., J. Potterat, D. Woodhouse, J. Muth, S. Muth and W. Darrow, "Social networks and infectious disease: The Colorado Springs study", Social Science & Medicine **38**, 1, 79–88 (1994).

Klovdahl, A. S., "Social networks and the spread of infectious diseases: The AIDS example", Social Science & Medicine **21**, 11, 1203–1216 (1985).

Klovdahl, A. S., E. A. Graviss, A. Yaganehdoost, M. W. Ross, A. Wanger, G. J. Adams and J. M. Musser, "Networks and tuberculosis: an undetected community outbreak involving public places", Social Science & Medicine **52**, 5, 681–694 (2001).

Kot, M., *Elements of mathematical ecology* (Cambridge University Press, Cambridge ; New York, NY, 2001).

Kroeber, A. L., *Anthropology: culture patterns & processes* (Harcourt Brace Jovanovich, San Diego, 1988), oCLC: 80563037.

Kucharski, A. J. and C. L. Althaus, "The role of superspreading in Middle East respiratory syndrome coronavirus (MERS-CoV) transmission", Eurosurveillance **20**, 25, 21167 (2015).

Kucharski, A. J., R. M. Eggo, C. H. Watson, A. Camacho, S. Funk and W. J. Edmunds, "Effectiveness of Ring Vaccination as Control Strategy for Ebola Virus Disease", Emerging Infectious Diseases **22**, 1, 105–108 (2016).

Lau, M. S. Y., B. D. Dalziel, S. Funk, A. McClelland, A. Tiffany, S. Riley, C. J. E. Metcalf and B. T. Grenfell, "Spatial and temporal dynamics of superspreading events in the 20142015 West Africa Ebola epidemic", Proceedings of the National Academy of Sciences **114**, 9, 2337–2342 (2017).

Leventhal, G. E., A. L. Hill, M. A. Nowak and S. Bonhoeffer, "Evolution and emergence of infectious diseases in theoretical and real-world networks", Nature Communications **6** (2015).

Li, W., Z. Shi, M. Yu, W. Ren, C. Smith, J. H. Epstein, H. Wang, G. Crameri, Z. Hu, H. Zhang, J. Zhang, J. McEachern, H. Field, P. Daszak, B. T. Eaton, S. Zhang and L.-F. Wang, "Bats Are Natural Reservoirs of SARS-Like Coronaviruses", **310**, 5 (2005).

Liang, W., Z. Zhu, J. Guo, Z. Liu, X. He, W. Zhou, D. P. Chin and A. Schuchat, "Severe Acute Respiratory Syndrome, Beijing, 2003", Emerging Infectious Diseases **10**, 1, 25–31 (2004).

Lindenbaum, S., *Kuru sorcery: disease and danger in the New Guinea highlands*, Explorations in world ethnology (Mayfield Pub. Co, Palo Alto, Calif, 1979), 1st ed edn.

Lindsay, J. A., "Staphylococcus aureus genomics and the impact of horizontal gene transfer", International Journal of Medical Microbiology **304**, 2, 103–109 (2014).

Liu, L., X.-Q. Zhao and Y. Zhou, "A Tuberculosis Model with Seasonality", Bulletin of Mathematical Biology **72**, 4, 931–952 (2010).

Liu, S., Y. Li, Y. Bi and Q. Huang, "Mixed vaccination strategy for the control of tuberculosis: A case study in China", Mathematical Biosciences and Engineering **14**, 3, 695–708 (2016).

Lloyd-Smith, J. O., S. J. Schreiber, P. E. Kopp and W. M. Getz, "Superspreading and the effect of individual variation on disease emergence", Nature **438**, 7066, 355–359 (2005).

Low, D. E., *SARS: Lessons from Toronto* (National Academies Press (US), 2004).

Luo, T., I. Comas, D. Luo, B. Lu, J. Wu, L. Wei, C. Yang, Q. Liu, M. Gan, G. Sun, X. Shen, F. Liu, S. Gagneux, J. Mei, R. Lan, K. Wan and Q. Gao, "Southern East Asian origin and coexpansion of *Mycobacterium tuberculosis* Beijing family with Han Chinese", Proceedings of the National Academy of Sciences **112**, 26, 8136–8141 (2015).

Maddocks, I., "Cultural Epidemiology", in "When Culture Impacts Health", pp. 167–179 (Elsevier, 2013).

Mancuso, J. D., J. M. Diffenderfer, B. J. Ghassemieh, D. J. Horne and T.-C. Kao, "The Prevalence of Latent Tuberculosis Infection in the United States", American Journal of Respiratory and Critical Care Medicine **194**, 4, 501–509 (2016).

Martin, G. and M. Boland, "Planning and preparing for public health threats at airports", Globalization and Health **14** (2018).

McElroy, A. and P. K. Townsend, *Medical anthropology in ecological perspective* (Westview Pr, Boulder, 1996), 3rd ed edn., oCLC: 833203637.

McGee, R. J. and R. L. Warms, *Anthropological theory: an introductory history* (Mayfield Pub. Co, Mountain View, Calif, 2000), 2nd ed edn.

McGrath, J. W., "Social networks of disease spread in the lower illinois valley: A simulation approach", American Journal of Physical Anthropology **77**, 4, 483–496 (1988).

McKendrick, A. G., "Applications of Mathematics to Medical Problems", Proceedings of the Edinburgh Mathematical Society **44**, 98 (1925).

Mehndiratta, M. M., P. Mehndiratta and R. Pande, "Poliomyelitis", The Neurohospitalist **4**, 4, 223–229 (2014).

Milner, G. R., "Epidemic Disease in the Postcontact Southeast: A Reappraisal", Midcontinental Journal of Archaeology **5**, 1, 39–56 (1980).

Mitchell, J. C., "Social Networks", Annual Review of Anthropology **3**, 1, 279–299 (1974).

Moran-Mendoza, O., M. C. Tello-Zavala, M. Rivera-Camarillo and Y. Ros-Meza, "Comparison of different methods and times for reading the tuberculin skin test", International Journal of Tuberculosis and Lung Disease **17**, 10, 1273–1278 (2013).

Morris, M., "Epidemiology and Social Networks: Modeling structured diffusion", Sociological Methods and Research **22**, 1, 99–126 (1993).

Morse, D., "Prehistoric Tuberculosis in America", American Journal of Respiratory Diseases **83**, 4, 489–502 (1961).

Muller, R., C. A. Roberts and T. A. Brown, "Complications in the study of ancient tuberculosis: Presence of environmental bacteria in human archaeological remains", Journal of Archaeological Science **68**, 5–11 (2016).

Mummert, T. M. A., "Modeling Super-spreading Events for Infectious Diseases: Case Study SARS", p. 7 (2011).

Murray, J. D., *Mathematical biology*, Interdisciplinary applied mathematics (Springer, New York, 2002), 3rd ed edn.

Nesse, R. M. and G. C. Williams, *Why we get sick: the new science of Darwinian medicine*, Vintage books (Vintage Books, New York, 1996).

Newman, M. E. J., "Spread of epidemic disease on networks", Physical Review E **66**, 1 (2002).

Ng, S. K., "Possible role of an animal vector in the SARS outbreak at Amoy Gardens", The Lancet **362**, 9383, 570–572 (2003).

Nishiyama, A., N. Wakasugi, T. Kirikae, T. Quy, L. D. Ha, H. T. Long, N. Keicho, T. Sasazuki and T. Kuratsuji, "Risk Factors for SARS Infection within Hospitals in Hanoi, Vietnam", Japanese Journal of Infectious Disease **61**, 388–390 (2008).

Nunn, C. L., F. Jordn, C. M. McCabe, J. L. Verdolin and J. H. Fewell, "Infectious disease and group size: more than just a numbers game", Philosophical Transactions of the Royal Society B: Biological Sciences **370**, 1669 (2015).

Oh, M.-d., P. G. Choe, H. S. Oh, W. B. Park, S.-M. Lee, J. Park, S. K. Lee, J.-S. Song and N. J. Kim, "Middle East Respiratory Syndrome Coronavirus Superspreading Event Involving 81 Persons, Korea 2015", Journal of Korean Medical Science **30**, 11, 1701–1705 (2015).

Otto, M., "Community-associated MRSA: what makes them special?", International journal of medical microbiology : IJMM **303**, 0, 324–330 (2013).

Paull, S. H., S. Song, K. M. McClure, L. C. Sackett, A. M. Kilpatrick and P. T. Johnson, "From superspreaders to disease hotspots: linking transmission across hosts and space", Frontiers in Ecology and the Environment **10**, 2, 75–82 (2012).

Petticrew, M. P., K. Lee and M. McKee, "Type A Behavior Pattern and Coronary Heart Disease: Philip Morriss Crown Jewel", American Journal of Public Health **102**, 11, 2018–2025 (2012).

Posa, A., F. Maixner, B. G. Mende, K. Khler, A. Oszts, C. Sola, O. Dutour, M. Masson, E. Molnr, G. Plfi and A. Zink, "Tuberculosis in Late Neolithic-Early Copper Age human skeletal remains from Hungary", Tuberculosis **95**, S18–S22 (2015).

Potterat, J. J., R. B. Rothenberg, D. Woodhouse, J. B. Muth, C. I. Pratts and J. F. Fogle, "Gonorrhea as a Social Disease", Sexually Transmitted Diseases **12**, 1, 25–32 (1985).

Richardson, T. O. and T. E. Gorochowski, "Beyond contact-based transmission networks: the role of spatial coincidence", Journal of The Royal Society Interface **12**, 111, 20150705 (2015).

Richerson, P. J., R. Boyd and R. L. Bettinger, "Was Agriculture Impossible during the Pleistocene but Mandatory during the Holocene? A Climate Change Hypothesis", American Antiquity **66**, 03, 387–411 (2001).

Riley, S., C. Fraser, C. A. Donnelly, A. C. Ghani, L. J. Abu-Raddad, A. J. Hedley, G. M. Leung, L.-M. Ho, T.-H. Lam, T. Q. Thach, P. Chau, K.-P. Chan, S.-V. Lo, P.-Y. Leung, T. Tsang, W. Ho, K.-H. Lee, E. M. C. Lau, N. M. Ferguson and R. M. Anderson, "Transmission Dynamics of the Etiological Agent of SARS in Hong Kong: Impact of Public Health Interventions", Science **300**, 5627, 1961–1966 (2003).

Riska, E., "The rise and fall of Type A man", Social Science & Medicine **51**, 11, 1665–1674 (2000).

Roberts, C. A. and J. E. Buikstra, *The bioarchaeology of tuberculosis: a global view on a reemerging disease* (University Press of Florida, Gainesville, 2003), oCLC: 694890650.

Roberts, C. A., L.-A. Pfister and S. Mays, "Letter to the editor: was tuberculosis present in Homo erectus in Turkey?", American Journal of Physical Anthropology **139**, 3, 442–444 (2009).

Rothenberg, R. B., J. J. Potterat, D. E. Woodhouse, S. Q. Muth, W. W. Darrow and A. S. Klovdahl, "Social network dynamics and HIV transmission:", AIDS **12**, 12, 1529–1536 (1998).

Salathe, M., M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, J. H. Jones and A. Raftery, "A high-resolution human contact network for infectious disease transmission", Proceedings of the National Academy of Sciences of the United States of America **107**, 51, 22020–22025 (2010).

Salo, W. L., A. C. Aufderheide, J. Buikstra and T. A. Holcomb, "Identification of Mycobacterium tuberculosis DNA in a pre-Columbian Peruvian mummy.", Proceedings of the National Academy of Sciences of the United States of America **91**, 6, 2091–2094 (1994).

Sattenspiel, L., "Modeling the spread of infectious disease in human populations", American Journal of Physical Anthropology **33**, S11, 245–276 (1990).

Schaller, M. and J. H. Park, "The Behavioral Immune System (and Why It Matters)", Current Directions in Psychological Science **20**, 2, 99–103, URL http://journals.sagepub.com/doi/10.1177/0963721411402596 (2011).

Shaban, N., M. Andersson, . Svensson and T. Britton, "Networks, epidemics and vaccination through contact tracing", Mathematical Biosciences **216**, 1, 1–8 (2008).

Sharp, P. M. and B. H. Hahn, "The evolution of HIV-1 and the origin of AIDS", Philosophical Transactions of the Royal Society B: Biological Sciences **365**, 1552, 2487–2494 (2010).

Sloot, R., M. F. Schim van der Loeff, P. M. Kouw and M. W. Borgdorff, "Risk of tuberculosis after recent exposure. A 10-year follow-up study of contacts in Amsterdam", American Journal of Respiratory and Critical Care Medicine **190**, 9, 1044–1052 (2014).

Smith, E. A., "Three styles in the evolutionary analysis of human behavior", in "Adaptation and human behavior", pp. 27–46 (Aldine de Gruyter, Hawthorne, NY, 1999).

Spigelman, M., H. D. Donoghue, Z. Abdeen, S. Ereqat, I. Sarie, C. L. Greenblatt, I. Pap, I. Szikossy, I. Hershkovitz, G. K. Bar-Gal and C. Matheson, "Evolutionary changes in the genome of Mycobacterium tuberculosis and the human genome from 9000 years BP until modern times", Tuberculosis **95**, S145–S149 (2015).

Stead, W. W., K. D. Eisenach, M. D. Cave, M. L. Beggs, G. L. Templeton, C. O. Thoen and J. H. Bates, "When did Mycobacterium tuberculosis infection first occur in the New World? An important question with public health implications.", American Journal of Respiratory and Critical Care Medicine **151**, 4, 1267–1268 (1995).

Stein, R. A., "Super-spreaders in infectious diseases", International Journal of Infectious Diseases **15**, 8, e510–e513 (2011).

Stockman, L. J., R. Bellamy and P. Garner, "SARS: Systematic Review of Treatment Effects", PLoS Medicine **3**, 9, e343 (2006).

Stroeymeyt, N., A. V. Grasse, A. Crespi, D. P. Mersch, S. Cremer and L. Keller, "Social network plasticity decreases disease transmission in a eusocial insect", Science **362**, 6417, 941–945 (2018).

Strong, P., "Epidemic psychology: a model", Sociology of Health & Illness **12**, 3, 249–259 (1990).

Student, "On the Error of Counting with a Haemacytometer", Biometrika **5**, 3, 351 (1907).

Sundaram, N., C. Schaetti, V. Purohit, A. Kudale and M. G. Weiss, "Cultural epidemiology of pandemic influenza in urban and rural Pune, India: a cross-sectional, mixed-methods study", BMJ Open **4**, 12 (2014).

Taylor, L. H., S. M. Latham and M. E. Woolhouse, "Risk factors for human disease emergence.", Philosophical Transactions of the Royal Society B: Biological Sciences **356**, 1411, 983–989 (2001).

Tibayrenc, M., "A molecular biology approach to tuberculosis", Proceedings of the National Academy of Sciences of the United States of America **101**, 14, 4721–4722 (2004).

Trostle, J. A. and J. Sommerfeld, "Medical Anthropology and Epidemiology", Annual Review of Anthropology **25**, 1, 253–274 (1996).

Vysakh, P. R. and M. Jeya, "A Comparative Analysis of Community Acquired and Hospital Acquired Methicillin Resistant Staphylococcus Aureus", Journal of Clinical and Diagnostic Research : JCDR **7**, 7, 1339–1342 (2013).

Weiss, M., "Explanatory Model Interview Catalogue (EMIC): Framework for Comparative Study of Illness", Transcultural Psychiatry **34**, 2, 235–263 (1997).

Weiss, M. G., "Cultural epidemiology: An introduction and overview", Anthropology & Medicine **8**, 1, 5–29 (2001).

Weiss, M. G., "Stigma and the Social Burden of Neglected Tropical Diseases", PLoS Neglected Tropical Diseases **2**, 5 (2008).

Weiss, M. G., D. Somma, F. Karim, A. Abouihia, C. Auer, J. Kemp and M. S. Jawahar, "Cultural epidemiology of TB with reference to gender in Bangladesh, India and Malawi", International Journal of Tuberculosis and Lung Disease **12**, 7, 837–847 (2008).

WHO, "Consensus document on the epidemiology of severe acute respiratory syndrome (SARS)", Tech. rep., Department of Communicable Disease Surveilance and Response (2003a).

WHO, "Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003", URL `https://www.who.int/csr/sars/country/table2003_09_23/en/` (2003b).

Wilbur, A. K., A. S. Bouwman, A. C. Stone, C. A. Roberts, L.-A. Pfister, J. E. Buikstra and T. A. Brown, "Deficiencies and challenges in the study of ancient tuberculosis DNA", Journal of Archaeological Science **36**, 9, 1990–1997 (2009).

Willem, L., F. Verelst, J. Bilcke, N. Hens and P. Beutels, "Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006-2015)", BMC Infectious Diseases **17** (2017).

Williams, B. G. and C. Dye, "Dynamics and control of infections on social networks of population types", Epidemics **23**, 11–18 (2018).

Wilson, E. O., *Sociobiology: the new synthesis* (Belknap Press of Harvard University Press, Cambridge, Mass, 2000), 25th anniversary ed edn.

Wirth, T., F. Hildebrand, C. Allix-Bguec, F. Wlbeling, T. Kubica, K. Kremer, D. van Soolingen, S. Rsch-Gerdes, C. Locht, S. Brisse, A. Meyer, P. Supply and S. Niemann, "Origin, Spread and Demography of the Mycobacterium tuberculosis Complex", PLoS Pathogens **4**, 9, e1000160 (2008).

Wong, S. S. and K. Y. Yuen, "The severe acute respiratory syndrome (SARS)", Journal of Neurovirology **11**, 5, 455–468 (2005).

Woolhouse, M. E. J., C. Dye, J.-F. Etard, T. Smith, J. D. Charlwood, G. P. Garnett, P. Hagan, J. L. K. Hii, P. D. Ndhlovu, R. J. Quinnell, C. H. Watts, S. K. Chandiwana and R. M. Anderson, "Heterogeneities in the transmission of infectious agents: Implications for the design of control programs", Proceedings of the National Academy of Sciences **94**, 1, 338–342 (1997).

World Health Organization, *Global tuberculosis report 2017* (2017), oCLC: 1030441030.

Xu, R.-H., J.-F. He, M. R. Evans, G.-W. Peng, H. E. Field, D.-W. Yu, C.-K. Lee, H.-M. Luo, W.-S. Lin, P. Lin, L.-H. Li, W.-J. Liang, J.-Y. Lin and A. Schnur, "Epidemiologic Clues to SARS Origin in China", Emerging Infectious Diseases **10**, 6, 1030–1037 (2004).

Yang, Y.-A., A. Chong and J. Song, "Why Is Eradicating Typhoid Fever So Challenging: Implications for Vaccine and Therapeutic Design", Vaccines **6**, 3 (2018).

Yorke, J. A., H. W. Hethcote and A. Nold, "Dynamics and Control of the Transmission of Gonorrhea", Sexually Transmitted Diseases **5**, 2, 51 (1978).

Ypma, R. J. F., H. K. Altes, D. van Soolingen, J. Wallinga and W. M. van Ballegooijen, "A Sign of Superspreading in Tuberculosis: Highly Skewed Distribution of Genotypic Cluster Sizes", Epidemiology **24**, 3, 395–400 (2013).

Yu, I. T., Z. H. Xie, K. K. Tsoi, Y. L. Chiu, S. W. Lok, X. P. Tang, D. S. Hui, N. Lee, Y. M. Li, Z. T. Huang, T. Liu, T. W. Wong, N. S. Zhong and J. J. Sung, "Why Did Outbreaks of Severe Acute Respiratory Syndrome Occur in Some Hospital Wards but Not in Others?", Clinical Infectious Diseases **44**, 8, 1017–1025 (2007).

Zelner, J. L., J. Trostle, J. E. Goldstick, W. Cevallos, J. S. House and J. N. Eisenberg, "Social Connectedness and Disease Transmission: Social Organization, Cohesion, Village Context, and Infection Risk in Rural Ecuador", American Journal of Public Health **102**, 12, 2233–2239 (2012).

Zielhuis, G. A. and L. A. Kiemeney, "Social epidemiology? No way", International Journal of Epidemiology **30**, 1, 43–44 (2001).

Ziv, E., C. L. Daley and S. M. Blower, "Early Therapy for Latent Tuberculosis Infection", American Journal of Epidemiology **153**, 4, 381–385 (2001).