SOCIAL STRUCTURE, CONTACT NETWORKS, AND THE SPREAD OF

RESPIRATORY INFECTIONS

By

PAIGE B. MILLER

(Under the Direction of JOHN M. DRAKE)

ABSTRACT

Our current pandemic has illustrated how infectious diseases, particularly respiratory

transmitted infections, can have devastating effects on human life. These pathogens

typically require close contact between a susceptible and infected host for successful

transmission. As a result, the nature of contacts among populations of hosts is of

extreme importance to the spread of disease. Traditional models are based on a

compartmentalization of hosts according to their disease status (i.e., susceptible,

infected, recovered) and make strong assumptions about homogenous mixing within the

population. This assumption can be inappropriate for populations which have

heterogeneous contact patterns. Heterogeneous contact patterns can be modeled using

contact networks (graphs where nodes represent individuals and edges represent

contacts which can facilitate disease spread). In recent years, many interesting

hypotheses about contact networks and disease spread have been put forward in

network and public health literature. In this dissertation, I use empirical data analyses

and simulations to investigate these hypotheses in the context of two globally important

pathogens, tuberculosis and COVID-19. In chapter 2, I investigate whether network

centrality and mixing patterns are correlated with male-bias in tuberculosis in a large,

social network from Kampala, Uganda. Next, in chapter 3, I focus specifically on the

effects of preferential social mixing by sex and whether it can facilitate higher infection

rates among males, as recently proposed in public health literature. In chapter 4, I

investigate the effects of core-periphery contact networks on disease spread and

discuss implications for populations with this contact structure. Lastly, in chapter 5, I

model counter-factual scenarios of the spread of COVID-19 inside care, correctional,

and meat-packing facilities without interventions and compare the distribution of

outbreak sizes to the actual distributions and argue that interventions these facilities

have taken have significantly mitigated spread. The work in this thesis is unique because

it tests theories derived from analytical methods or proposed in the public health

literature using simulations and data analyses. This work extends our understanding of

how human contact patterns alter disease spread, provides general insights into

infectious disease ecology, and has practical public health recommendations.

INDEX WORDS: Human social network, community structure, contact
network, assortativity, random graphs, sociality, infectious
diseases, mathematical model, tuberculosis, COVID-1

SOCIAL STRUCTURE, CONTACT NETWORKS, AND THE SPREAD OF

RESPIRATORY INFECTIONS

by

PAIGE B. MILLER
B.A., Gustavus Adolphus College, 2015

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in
Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA
2020

SOCIAL STRUCTURE, CONTACT NETWORKS, AND THE SPREAD OF

RESPIRATORY INFECTIONS

by

Paige B. Miller

Major Professor: John M. Drake

Committee: Nicole Lazar
                    Pejman Rohani
                    Andrew W. Park
                    Christopher C. Whalen

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
December 2020

Dedication

This dissertation is dedicated to my family and especially my dad, whose love of science, technology, and history is *infectious.*

ACKNOWLEDGEMENTS

It is difficult to overstate my gratitude to advisor, Dr. John Drake. John, from my first undergraduate research projects in your lab to finishing this dissertation, you have been incredibly generous with opportunities and wholehearted support. You always guided me in the right direction and now you have helped me achieve my goals of working for the CDC. In particular, I deeply admire your passion for science, dazzling intelligence, and patience with your students.

I also want to thank my dissertation committee members for their support. Chris, you helped me merge my love of public health into my degree in ecology. Thank you for helping me understand the world of tuberculosis and giving me another perspective throughout graduate school. Dr. Park, especially, thank you for taking me on as an undergraduate researcher in 2013 and introducing me to mathematical models of infectious diseases. I've always admired your creativity and hope to channel it someday.

Besides my committee members, I have countless science mentors to thank. Dr. Margaret Bloch-Qazi, my undergraduate biology advisor, gave me an opportunity to pursue research in 2012. Thank you for trusting me and teaching me to stay organized and especially for being my first role model as a female scientist. Dr. TJ Morrison, I wish I could have thanked you for your humor and support in guiding me through my undergraduate math major. Dr. Gordon

Mansergh, thank you making my dreams of working at the CDC come true, and thank you for continuing to be a mentor to me. Dr. John Lammert, you have inspired generations of Gustavus students with your Intro Biology courses, and I was definitely one of them. Dr. Sarah Bowden, thank you for always lending me advice and for being an amazing role model for me in this journey. Each of you have been so important to me over the years and I couldn't have done it without you.

At last, I would like to thank my family and especially my parents, Mike and Rose, for shaping me into the person I am today and for being a constant support system. I thank them for being my first teachers and always giving me the best a daughter could ask for. Finally, I am especially grateful to my partner, Philip, for keeping me grounded, for being my support and comfort, and for being the one I can share this with.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

Chapter 1

INTRODUCTION

Background

Disease ecology is motivated by the need to understand the spread and persistence of infectious diseases, which impose burdens on human and animal populations alike. Many infectious diseases are spread through direct contact of infectious and susceptible hosts (Hay et al., 2013). Consequently, epidemic outcomes are inherently tied to the organization of interactions among hosts. These interactions and resulting transmission from infected to susceptible hosts can be dependent on host parameters -- age, sex, and contact rate -- and disease parameters -- susceptibility to disease, infectiousness, and recovery rate (Anderson & May, 1991; Baggaley et al., 2010; Horton et al., 2020; Mossong et al., 2008). The theoretical framework for mathematical models that explore the consequences of heterogeneities in host and disease parameters on epidemics was pioneered in the twentieth century with the work of Hamer (1906), Ross (1908), Kermack and Mckendrick (1927) and Anderson and May (1992).

Traditional transmission models start with the basic premise that a population can be divided into a set of distinct classes depending on their disease status. The simplest of these models classifies individuals as susceptible (S), infectious (I), or recovered (R) and is called an SIR model. All individuals in

the population start in the susceptible class, where they can be infected by the disease at rate λ. Once infected, individuals can recover at rate γ. Recovered individuals are assumed to be fully immune for life. The familiar SIR model can be described more precisely with a system of differential equations, one for the proportion of individuals in each disease class:

$$\frac{dS}{dt} = -\lambda S$$
$$\frac{dI}{dt} = \lambda S - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

Many biologically motivated modifications can be made to this basic framework, involving the further subdivision of the S, I, and R classes to reflect either more complex host (e.g., subgroups based on age and sex (Hethcote, 2000)) or disease traits (e.g., a pathogen with an incubation period (Anderson, 1988)). Birth and death rates can also be introduced for diseases where progression spans timescales over which these demographic changes are important. Although the differential equation SIR model and its extensions have had a long and successful history, it is critically dependent on the assumptions of mass-action and homogenous mixing within compartments. In recent years, network analytic approaches have gained momentum as they capture the variation in individual behavior as well as the local, and global patterns of contacts present in host populations (Meyers et al., 2005; Keeling & Eames, 2005; Newman et al., 2006; Bansal et al., 2007; Craft 2015).

Multiscale features of contact networks and implications for disease spread

In network models, nodes that interact with each other are linked by edges. Nodes can represent countries, cities, or locations, but in this dissertation, I consider nodes to be individual hosts. Formally, a contact network explicitly represents interactions among hosts that can mediate disease spread. We contrast this with social networks which include interactions among hosts which cannot mediate disease spread (e.g., online communication). Node degree is the number of edges attached to it (i.e., number of contacts) and is one measure of a node's centrality within a network. Other measures of node centrality include betweenness, closeness, and eigencentrality and each measures a node's "importance" according to different criteria (Christley et al., 2005). Separate statistics describe network-level features, such as the mean degree of a network. These different scales of network structure, spanning node to network-level features, can affect disease spread (Figure 1.1). In this section, I review how disease spread is affected by multiple scales of network structures and then discuss emerging hypotheses and problems that this dissertation addresses.

At a local scale, nodes with higher network centrality can be infected faster and more often than less central nodes (Christley et al., 2005). This idea has been investigated in a diverse range of infectious diseases including influenza (Christakis & Fowler, 2010) and HIV (Rothenberg et al., 1995) in human populations and *Mycobacterium bovis* in captive possums (Corner et al., 2003). In these examples, centrality measures were used to identify which

individuals would be pivotal to the spread of infection. In a study of influenza in a college population, infections among more central individuals were detected nearly two weeks before less central individuals (Christley et al., 2005). In another study, predictions that more central individuals would be infected more often and faster than less central individuals were confirmed by the experimental introduction of *M. bovis* into a possum population (Corner et al., 2003). Generally, previous studies of network centrality and infection focused on outbreaks in closed populations of mostly susceptible individuals. An open question I address is whether the same expectations regarding centrality and infection apply to endemic infectious diseases which do not spread amongst a fully susceptible network of individuals as invading influenza strains or sexually transmitted infections.

In addition to local statistics, a remarkable number of studies have focused on characterizing how disease spreads in idealized networks with different degree distributions, clustering patterns, and other network-level structures, such as lattice grids (Barbour & Mollison, 1990; Pastor-Satorras & Vespignani, 2001; Shirley & Rushton, 2005). Researchers have paid special attention to small-world networks (Watts & Strogatz, 1998) – characterized by high levels of local clustering and global connectivity – and scale-free networks (Barabasi & Albert, 1999) – defined by having degree distributions that follow a power-law giving rise to a small fraction of highly connected nodes. To study the effects of scale-free and small-world structure on disease spread, researchers

4

compare outbreaks on "null" networks which can be lattice, regular, or Poisson (also called Erdos-Renyi) random graphs (Bansal et al., 2007). Modelers have long known that the epidemic threshold (i.e., the critical transmission rate above which disease may spread and persist) varies across these archetype networks. For example, the epidemic threshold is low, or non-existent, in scale-free networks meaning outbreaks are always possible (Pastor-Satorras & Vespignani, 2001). More generally, as variation in contact rates (i.e., degree distribution) increases, the epidemic threshold decreases (Woolhouse et al., 1997) and as clustering increases, the final outbreak size decreases (Badham & Stocker, 2010). Real-world contact networks are composed of various idealized network features. For example, a well-known study found human mobility networks to follow scale-free distributions while social networks followed small-world distributions (Eubank et al., 2004). However, identifying which of these idealized networks resemble specific populations given incomplete sampling is a challenging methodological area that I tackle in this dissertation.

In between local and network-scale network structures are those at the "meso-scale" (Rombach et al., 2014) which describe organization of groups of similar individuals within the network. The most common meso-scale statistic is network modularity (i.e., assortativity), which is a measure of how often a network's nodes attach to others in the same group (Newman, 2006). Modularity can reduce the final outbreak size for immunizing infectious diseases (e.g., measles), but may only play a large role when it is relatively strong (Sah et al.,

2017; Salathé & Jones, 2010). For example, an analysis of animal social networks found that most animal populations are not modular enough to have protective effects of limiting spread (Sah et al., 2017). This study did, however, show that intermediate levels of modularity could affect timing of infection spread to different groups highlighting how modularity may alter individual likelihood and timing of infection based on group membership. In human contact networks, there are a number of subtle, modular patterns proposed to affect the distribution of infections across groups including age, race, and sex-assortative mixing patterns. Patterns of age-assortative mixing are the most widely studied; these mixing patterns are important for explaining incidence and mortality rates for pathogens including measles (Schenzle, 1984), pertussis (Rohani et al., 2010), and tuberculosis (Arregui et al., 2018). Patterns of racial assortativity in sexual contact networks have been identified in models to explain why Black populations of men who have sex with men have a higher prevalence of HIV (Mustanski et al., 2015), but other modeling studies have found factors such as the care continuum to be more important (Goodreau et al., 2017). Lastly, preferential mixing in social contact networks have been proposed to drive sex-disparities in infectious diseases (Horton et al., 2020), which I test with a modeling study in this dissertation. In addition to modularity, I explore the disease consequences "core-periphery" networks, another meso-scale structure.

Two Study Systems

Social contact networks, tuberculosis, and sex-disparities in infection

Tuberculosis (TB) is a suitable disease for studying the effects of network centrality and social network assortativity on prevalence of endemic diseases. It is endemic in many low-income countries and up to one-quarter of the world's population has latent tuberculosis infection (Dye, 2015; Houben & Dodd, 2016). As a respiratory transmitted infectious disease, it relies on contact networks, spreading more often during prolonged, intimate, frequent contacts (Castellanos et al., 2020; Yates et al., 2016). Indeed, having a household member with TB is the largest risk factor for getting infected in endemic settings despite most transmission being community-spread (reviewed in (Yates et al., 2016)), underscoring why contact tracing remains an important but insufficient prevention strategy in many places. In addition, infection risk is also strongly related to factors including sex, age, co-infections, and cigarette smoking, among others (Dye, 2015). Sex as a risk factor for TB is particularly not well-understood and part of this dissertation focuses on understanding whether and how contact patterns help explain sex-disparities in prevalence. To investigate this, I use network models and data analyses to examine mechanisms for explaining sex-disparities in TB which fall into "biological" and "social" categories. Biological mechanisms for male-bias of TB typically discuss higher male susceptibility and transmissibility (Nhamoyebonde & Leslie, 2014). Social factors are generally discussed in terms of gender roles and preferences in mixing patterns driving

higher male exposure to TB cases through work outside their home, travel, or mixing with other men (Guerra-Silveira & Abad-Franch, 2013). These suspected differences in gender roles and preferences might manifest in social contact networks as higher male centrality and sex-assortativity. The overarching question for Chapters 2 and 3 is how social and biological factors may combine to explain sex-disparities in TB.

In Chapter 2, I characterized social factors that could drive sex-disparities in TB by comparing patterns of centrality and sex-assortativity among TB cases and males in Uganda, where TB is endemic (Organization, 2018). I used data from a large, social network survey of recently diagnosed cases and community controls. I estimated various centrality statistics for TB cases and controls including degree, closeness, betweenness, eigencentrality, and network distance to other TB cases. Despite there being correlations between centrality and infection in other respiratory transmitted diseases and in past studies of TB, testing this correlation for human TB and in this population is novel. The endemic nature of TB in Kampala could mean the remaining susceptible individuals, from which recently diagnosed cases predominantly emerge in high incidence countries (Cohn & O'Brien, 1998), are *less central* than the general population (Ferrari et al., 2006). To understand whether network centrality could help explain sex-disparities in TB, I compared centrality statistics between males and females. To understand how underlying network structures and network sampling protocol may affect estimates of centrality, I conducted multiple simulation

studies. Finally, to quantify assortative mixing patterns, I estimated proportion of within-sex, between-sex, and age-assortative contact stratified by disease status and sex. Understanding whether TB cases and males are more central in their networks and whether TB cases differ in preferences in social contacts has implications for case-finding and prioritization of contact tracing.

The subject of Chapter 3 is comparing the effects of social and biological factors on sex-disparities in TB. Specifically, I investigated whether sex-assortative social interactions are strong enough to contribute to sex disparities of TB or if biological differences in sex-traits are needed to explain observed male-bias. Hypothesized biological mechanisms for male-bias in TB are primarily related to a suspected higher male susceptibility to infection due to genetic (chromosomal genes), hormonal (effects of sex-hormones), or behavioral (smoking) factors (Nhamoyebonde & Leslie, 2014). Other than susceptibility, lesser studied biological mechanisms that could plausibly lead to male-bias include higher male transmissibility (Dodd et al., 2016) or sex differences in the disease progression rates (Holmes et al., 1998). Each of these different sex-traits -- susceptibility, transmissibility, and infectious period -- could lead to male-bias in TB. I used mathematical models of disease spread on networks to examine how sex-disparities in infection could emerge in the presence of each of these sex-traits with varying levels of network assortativity. I was also interested in understanding factors that allow subtle patterns in assortative contact patterns to shape epidemic outcomes. Specifically, I investigated whether the unique life

9

history of human TB, with its long and variable latent period and slow rates of transmission, increases the effects of sex-traits and assortative mixing on male-bias. To study this, I compared simulations of models with multiple transmission patterns (SIR, SLIR, SIRS, and SLIRS) and transmission rates. Although this study focused on TB, many infectious diseases are male-biased and most populations have social mixing patterns marked by sex-assortativity. Results from this study thus highlight how behavior and contact rates amplify the consequences of sex-differences in immune investment.

Contact network structure, interventions, and spread of SARS-CoV-2

Our current pandemic of Coronavirus Infectious Disease (COVID) is illustrating how contact patterns and human behavior alter spread and prevention of infectious diseases. On December 31st, 2019, China first reported that a cluster of patients with pneumonia-like symptoms in Wuhan, Hubei Province (WHO Archived Timeline). Subsequent investigation identified severe acute respiratory syndrome coronavirus-2 (SARS-Cov-2) as the causative agent. Chinese authorities ordered multiple containment measures in January 2020, but the virus quickly spread globally throughout the winter and spring of 2020. SARS-Cov-2 spread so widely in the United States that it now accounts for one-quarter of all cases (nearly 5 million) and 22% of all deaths (160,000) despite comprising only 4% of the world's population (WHO Situation Report). Cases and deaths in the US have not been distributed randomly but early on were concentrated in populations that live or work nursing homes (Barnett & Grabowski, 2020), prisons

(Hawks et al., 2020), and cruise ships (Emery et al., 2020) which have highly heterogeneous contact patterns. I was motivated to study how contact patterns specific to these settings may affect spread of infectious diseases.

In contrast to other high-transmission settings which are thought to be more well-mixed (e.g., markets, worship settings, concerts, rallies, sports games, and funerals), interactions in nursing homes, prisons, and cruise ships vary greatly depending on the roles individuals play in the population. For example, on board the Diamond Princess Cruise Ship during its 3-week quarantine, crew workers were largely unconfined whereas passengers rarely left their cabins. This outbreak resulted in nearly 700 passengers being infected, 9 deaths, and an estimated initial R_0 of 9.3 (Emery et al., 2020). Whether this quarantine, and contact structure, prevented spread is unknown but clearly highlights the need to understand spread of disease in closed settings with structured populations. In addition to cruise ships, nursing homes and prisons have instated similar isolation policies for residents during lockdown, who then form a sort of periphery. In both nursing homes and prisons, staff maintain their duties and may serve as the only daily interactions for residents. Together, we hypothesized, staff and resident interactions form a sort of "core-periphery" contact network in these settings.

In Chapter 4, my goal was to understand how contact networks structured into densely connected "cores" and sparsely connected "periphery" nodes affect spread of infectious diseases. To do this, I simulated core-periphery contact

networks, described their unique structural properties, and compared disease spread on core-periphery networks with homogenous networks. Specifically, I examined the probability distribution of outbreak sizes, and compared when the probability distribution transitions from unimodal, with rare large outbreaks, to bimodal, with frequent large outbreaks in core-periphery and homogenous networks. I additionally characterized various epidemic dynamics on core-periphery networks that affect containability including the peak timing, peak size, and outbreak duration. Lastly, I quantified estimation biases for R_0 in core-periphery networks, which may be important to understand for accurate estimations of vaccination thresholds required for herd immunity in these vulnerable populations. The effects of core-periphery network structure on disease dynamics represents a gap in the contact network epidemiology literature and improved understanding of disease spread in these networks may inform interventions based on modifying contact network structures.

In Chapter 5, I reasoned that the outbreak size distribution of COVID clusters in US hotspots might resemble distributions of outbreak sizes on core-periphery networks, unless interventions facilities took substantially curbed spread. I modeled the unmitigated spread of COVID in core-periphery networks under a variety of scenarios, including different number of introduced infected individuals and varying reproductive ratios. I compared simulated distributions to outbreak size distributions in US long-term care facilities, correctional facilities, and food processing facilities. To understand how interventions may have altered the

spread of COVID in these facilities, I further characterized the distribution of outbreak sizes in care facilities over time and described differences in outbreak sizes when facilities had access to two different interventions. Understanding whether and how interventions have mitigated spread in vulnerable populations is important as the current pandemic is far from over.

Intellectual Significance

This dissertation analyzes when individual-, group-, and population-level patterns of human mixing affect pathogen spread and therefore provides theoretical insight into when contact patterns could be useful for accurately predicting incidence and developing control strategies. In the second chapter, I asked whether network centrality is higher among recently diagnosed, active TB cases than community controls in an endemic population. While previous studies focused on network centrality and infection in mostly susceptible populations, nearly half of Kampala residents have latent TB infection and the remaining contact network of susceptible individuals may be different from the contact network overall. Therefore, this analysis elucidates how population-level differences in disease endemicity control when individual-level variation in centrality is predicted to affect infection risk. I additionally determined whether males are more central in their contact network than females and whether there are differences in mixing patterns by sex that could help explain male-bias in TB notifications, a hypothesis presented in a recent review (Horton et al., 2020). Understanding social factors leading to increased infections among men has

implications for improving models of disease spread and design of new control strategies. In the third chapter, I further investigated the hypothesis that sex-assortative mixing patterns drive sex-disparities in infectious diseases using a mathematical model and simulations of disease spread on networks. This chapter highlights disease-specific factors that make mixing patterns more important for predicting disease incidence and how behavior can amplify the consequences of sex-differences in immune investment. In the fourth chapter, I investigated effects of core-periphery group structure on infectious disease dynamics, which is a current gap in the extensive contact network epidemiology literature. Specifically, I compared the outbreak size distribution in core-periphery and homogenous networks and studied the pathogen characteristics that increase the likelihood of large outbreaks in core-periphery networks. Lastly, in the fifth chapter, I reasoned that the distribution of outbreak sizes in COVID19 hotspots should resemble simulated outbreak size distributions, unless interventions have significantly mitigated spread. To test this, I compared simulated and actual outbreak size distributions in United States care, correctional, and food processing facilities. Understanding whether interventions dampened large outbreaks of COVID19 in vulnerable populations is important for future control programs. Throughout, I highlight why studying human mixing patterns improves our understanding of disease spread -- a central goal of disease ecologists.

References

Anderson, R. M. (1988). The epidemiology of HIV infection: variable incubation plus infectious periods and heterogeneity in sexual activity. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 151(1), 66-93.

Anderson, R. M., & May, R. M. (1992). *Infectious Diseases of Humans*.

Arregui, S., Iglesias, M. J., Samper, S., Marinova, D., Martin, C., Sanz, J., & Moreno, Y. (2018). Data-driven model for the assessment ofMycobacterium tuberculosistransmission in evolving demographic structures. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(14), E3238–E3245. https://doi.org/10.1073/pnas.1720606115

Badham, J., & Stocker, R. (2010). The impact of network clustering and assortativity on epidemic behaviour. *Theoretical Population Biology*, *77*(1), 71–75. https://doi.org/10.1016/j.tpb.2009.11.003

Baggaley, R. F., White, R. G., & Boily, M.-C. (2010). HIV transmission risk through anal intercourse: systematic review, meta-analysis and implications for HIV prevention. *International Journal of Epidemiology*, *39*(4), 1048–1063. https://doi.org/10.1093/ije/dyq057

Bansal, S., Grenfell, B. T., & Meyers, L. A. (2007). When individual behaviour matters: homogeneous and network models in epidemiology. *Journal of The Royal Society Interface*, *4*(16), 879–891. https://doi.org/10.1098/rsif.2007.1100

Barabasi, A., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512.

Barbour, A., & Mollison, D. (1990). *Epidemics and random graphs* (Vol. 86, pp. 86–89). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-10067-7_8

Barnett, M. L., & Grabowski, D. C. (2020). Nursing Homes Are Ground Zero for COVID-19 Pandemic. *JAMA Health Forum*, *1*(3), e200369–e200369. https://doi.org/10.1001/jamahealthforum.2020.0369

Castellanos, M. E., Zalwango, S., Kakaire, R., Ebell, M. H., Dobbin, K. K., Sekandi, J., Kiwanuka, N., & Whalen, C. C. (2020). Defining adequate contact for transmission of Mycobacterium tuberculosis in an African urban environment. *Bmc Public Health*, *20*(1), 892. https://doi.org/10.1186/s12889-020-08998-7

Christakis, N. A., & Fowler, J. H. (2010). Social Network Sensors for Early Detection of Contagious Outbreaks. *PLoS ONE*, *5*(9). https://doi.org/10.1371/journal.pone.0012948

Christley, R. M., Pinchbeck, G. L., Bowers, R. G., Clancy, D., French, N. P., Bennett, R., & Turner, J. (2005). Infection in social networks: using network analysis to identify high-risk individuals. *American Journal of Epidemiology*, *162*(10), 1024–1031. https://doi.org/10.1093/aje/kwi308

Cohn, D. L., & O'Brien, R. J. (1998). The use of restriction fragment length polymorphism (RFLP) analysis for epidemiological studies of tuberculosis in developing countries. *Int. J. Tuberc. Lung Dis.*, *2*(1), 16–26.

Corner, L. A. L., Pfeiffer, D. U., & Morris, R. S. (2003). Social-network analysis of Mycobacterium bovis transmission among captive brushtail possums (Trichosurus vulpecula). *Preventive Veterinary Medicine*, *59*(3), 147–167. https://doi.org/10.1016/s0167-5877(03)00075-8

Craft, M. E. (2015). Infectious disease transmission and contact networks in wildlife and livestock. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1669). https://doi.org/10.1098/rstb.2014.0107

Dodd, P. J., Looker, C., Plumb, I. D., Bond, V., Schaap, A., Shanaube, K., Muyoyeta, M., Vynnycky, E., Godfrey-Faussett, P., Corbett, E. L., Beyers, N., Ayles, H., & White, R. G. (2016). Age- and Sex-Specific Social Contact Patterns and Incidence of Mycobacterium tuberculosis Infection. *American Journal of Epidemiology*, *183*(2), 156–166. https://doi.org/10.1093/aje/kwv160

Dye, C. (2015). *Population Biology of Tuberculosis*. Princeton University Press.

Emery, J. C., Russell, T. W., Liu, Y., Hellewell, J., Pearson, C. A., Group, C. C.-19 W., Atkins, K. E., Klepac, P., Endo, A., Jarvis, C. I., Davies, N. G., Rees, E. M., Meakin, S. R., Rosello, A., Zandvoort, K. van, Munday, J. D., Edmunds, W. J., Jombart, T., Auzenbergs, M., … Houben, R. M. (2020). The contribution of asymptomatic SARS-CoV-2 infections to transmission on the Diamond Princess cruise ship. *Elife*, *9*, 2081. https://doi.org/10.7554/elife.58699

Eubank, S., Guclu, H., Kumar, V., Marathe, M. V., Srinivasan, A., Toroczkai, Z., & Wang, N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, *429*(6988), 180–184. https://doi.org/10.1038/nature02541

Ferrari, M. J., Bansal, S., Meyers, L. A., & Bjørnstad, O. N. (2006). Network frailty and the geometry of herd immunity. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1602), 2743–2748. https://doi.org/10.1098/rspb.2006.3636

Goodreau, S. M., Rosenberg, E. S., Jenness, S. M., HIV, N. L. T. L., & 2017. (2017). Sources of racial disparities in HIV prevalence in men who have sex with men in Atlanta, GA, USA: a modelling study. *The Lancet HIV*, *4*(7), e311–e320. https://doi.org/10.1016/s2352-3018(17)30067-x

Guerra-Silveira, F., & Abad-Franch, F. (2013). Sex bias in infectious disease epidemiology: patterns and processes. *PLoS ONE*, *8*(4), e62390. https://doi.org/10.1371/journal.pone.0062390

Hawks, L., Woolhandler, S., & McCormick, D. (2020). COVID-19 in Prisons and Jails in the United States. *JAMA Internal Medicine*, *180*(8), 1041–1042. https://doi.org/10.1001/jamainternmed.2020.1856

Hay, S. I., Battle, K. E., Pigott, D. M., Smith, D. L., Moyes, C. L., Bhatt, S., Brownstein, J. S., Collier, N., Myers, M. F., George, D. B., & Gething, P. W. (2013). Global mapping of infectious disease. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1614), 20120250. https://doi.org/10.1098/rstb.2012.0250

Hethcote, H. W. (2000). The mathematics of infectious diseases. *SIAM review*, *42*(4), 599-653.

Holmes, C. B., Hausler, H., & Nunn, P. (1998). A review of sex differences in the epidemiology of tuberculosis. *The International Journal of Tuberculosis and Lung Disease*, *2*(2), 96–104.

Horton, K. C., Hoey, A. L., Béraud, G., Corbett, E. L., & White, R. G. (2020). Systematic Review and Meta-Analysis of Sex Differences in Social Contact Patterns and Implications for Tuberculosis Transmission and Control. *Emerging Infectious Diseases*, *26*(5), 910–919. https://doi.org/10.3201/eid2605.190574

Houben, R. M. G. J., & Dodd, P. J. (2016). The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS Med*, *13*(10), e1002152. https://doi.org/10.1371/journal.pmed.1002152

Keeling, M. J., & Eames, K. T. (2005). Networks and epidemic models. Journal of the Royal Society Interface, 2(4), 295-307.Meyers, L. A., Pourbohloul, B., Newman, M. E., Skowronski, D. M., & Brunham, R. C. (2005). Network theory and SARS: predicting outbreak diversity. *Journal of theoretical biology*, *232*(1), 71-81.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., & Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*, *5*(3), e74. https://doi.org/10.1371/journal.pmed.0050074

Mustanski, B., Birkett, M., Kuhns, L. M., Latkin, C. A., & Muth, S. Q. (2015). The Role of Geographic and Network Factors in Racial Disparities in HIV Among Young Men Who have Sex with Men: An Egocentric Network Study. *AIDS and Behavior*, *19*(6), 1037–1047. https://doi.org/10.1007/s10461-014-0955-0

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, *103*(23), 8577–8582. https://doi.org/10.1073/pnas.0601602103

Newman, M. E., Barabási, A. L. E., & Watts, D. J. (2006). *The structure and dynamics of networks*. Princeton university press.

Nhamoyebonde, S., & Leslie, A. (2014). Biological Differences Between the Sexes and Susceptibility to Tuberculosis. *Journal of Infectious Diseases*, *209*(suppl 3), S100–S106. https://doi.org/10.1093/infdis/jiu147

Organization, W. H. (2018). *Global tuberculosis report*. https://www.who.int/tb/publications/global_report/en/

Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic dynamics and endemic states in complex networks. *Physical Review E*, *63*(6 Pt 2), 066117. https://doi.org/10.1103/physreve.63.066117

Rohani, P., Zhong, X., & King, A. A. (2010). Contact Network Structure Explains the Changing Epidemiology of Pertussis. *Science*, *330*(6006), 982–985. https://doi.org/10.1126/science.1194134

Rombach, M. P., Porter, M. A., Fowler, J. H., & Mucha, P. J. (2014). Core-Periphery Structure in Networks. *SIAM Journal on Applied Mathematics*, *74*(1), 167–190. https://doi.org/10.1137/120881683

Rothenberg, R. B., Potterat, J. J., Woodhouse, D. E., Darrow, W. W., Muth, S. Q., & Klovdahl, A. S. (1995). Choosing a centrality measure: Epidemiologic correlates in the Colorado Springs study of social networks. *Social Networks*, *17*(3–4), 273–297. https://doi.org/10.1016/0378-8733(95)00267-r

Sah, P., Leu, S. T., Cross, P. C., Hudson, P. J., & Bansal, S. (2017). Unraveling the disease consequences and mechanisms of modular structure in animal social networks. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(16), 4165–4170. https://doi.org/10.1073/pnas.1613616114

Salathé, M., & Jones, J. H. (2010). Dynamics and Control of Diseases in Networks with Community Structure. *PLoS Computational Biology*, *6*(4), e1000736. https://doi.org/10.1371/journal.pcbi.1000736

Schenzle, D. (1984). An age-structured model of pre- and post-vaccination measles transmission. *IMA Journal of Mathematics Applied in Medicine and Biology*, *1*(2), 169–191. https://doi.org/10.1093/imammb/1.2.169

Shirley, M. D. F., & Rushton, S. P. (2005). The impacts of network topology on disease spread. *Ecological Complexity*, *2*(3), 287–299. https://doi.org/10.1016/j.ecocom.2005.04.005

Waltenburg, M. A., Victoroff, T., Rose, C. E., Butterfield, M., Jervis, R. H., Fedak, K. M., Gabel, J. A., Feldpausch, A., Dunne, E. M., Austin, C., Ahmed, F. S., Tubach, S., Rhea, C., Krueger, A., Crum, D. A., Vostok, J., Moore, M. J., Turabelidze, G., Stover, D., … Team, C.-19 R. (2020). COVID-19 Among Workers in Meat and Poultry Processing Facilities - United States, April-May 2020. *MMWR. Morbidity and Mortality Weekly Report*, *69*(27), 887–892. https://doi.org/10.15585/mmwr.mm6927e2

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Woolhouse, M. E., Dye, C., Etard, J. F., Smith, T., Charlwood, J. D., Garnett, G. P., Hagan, P., Hii, J. L., Ndhlovu, P. D., Quinnell, R. J., Watts, C. H., Chandiwana, S. K., & Anderson, R. M. (1997). 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.

Yates, T. A., Khan, P. Y., Knight, G. M., Taylor, J. G., McHugh, T. D., Lipman, M., White, R. G., Cohen, T., Cobelens, F. G., Wood, R., Moore, D. A. J., & Abubakar, I. (2016). The transmission of Mycobacterium tuberculosis in high burden settings. *The Lancet. Infectious Diseases*, *16*(2), 227–238. https://doi.org/10.1016/s1473-3099(15)00499-5
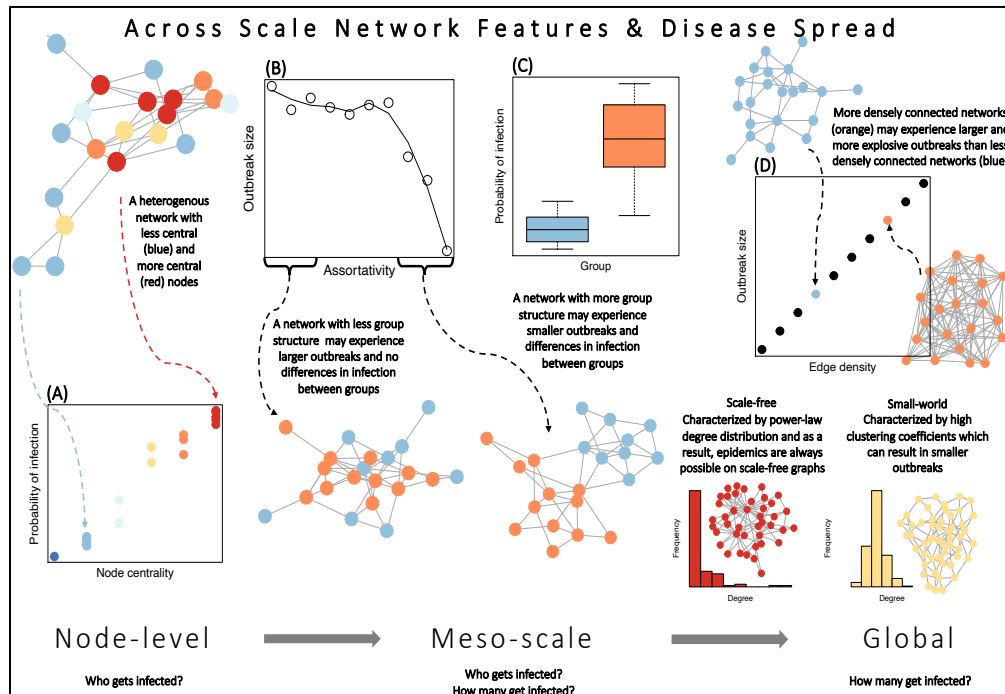
Figure 1.1 Multi-scale network structure and disease spread.

Network features across multiple scales of organization and those that measure a node's position or centrality in the network such as node degree. (A) Higher centrality nodes are generally expected to become infected faster and more often than lower centrality nodes. Meso-scale statistics, such as assortativity, describe the organization of nodes into groups. (B) With high levels of assortativity, outbreaks can be reduced in size and (C) group membership may affect the probability of a node getting infected. Finally, global features are summary statistics of the network as a whole such as mean degree, the degree distribution, and the clustering coefficient. (D) As network connectivity increases, outbreak size increases. Generally, node-level statistics determine "who gets infected" while global features determine "how many individuals get infected". Meso-scale statistics may affect both "who" and "how many individuals" get infected.

Chapter 2

ASSOCIATION BETWEEN MALE-BIAS IN TUBERCULOSIS CASES AND

SOCIAL NETWORK STRUCTURE IN KAMPALA, UGANDA[1]

Abstract

Globally, Tuberculosis disease (TB) is more common among men than women. Recent research proposes that differences in social mixing by sex could alter infection patterns in TB. We examine evidence for two mechanisms by which social-mixing could increase men's contact rates with TB cases. First, men could be positioned in social networks such that they contact more people or social groups. Second, preferential mixing by sex could expose men to more TB cases. We compared the networks of male and female TB cases and controls living in Rubaga Division in Kampala, Uganda. Specifically, we estimated their node position (including distance to TB cases, node degree, betweenness, and closeness) and assortativity patterns (mixing with adult males, females, and children inside and outside the household). The observed network consisted of 11,840 individuals. There were few differences in node position by sex. We found distinct mixing patterns by sex and TB disease status including that TB cases having a proportionally more adult male contacts and fewer contacts with children. This analysis uses a network approach to study which social mixing patterns are associated with TB disease. Understanding these mechanisms has implications for designing targeted intervention strategies in high-burden populations.

Introduction

Although tuberculosis (TB) is both a treatable and preventable disease, it remains one of the leading causes of death worldwide. Each year an estimated 1

million people die among 9 million new TB cases (WHO, 2018). The global

burden of TB is even greater as 25% of the world's population has a latent

infection with *M. tuberculosis (Mtb)* and is at risk of progressing to active disease

(Horton et al., 2016). Notification of TB disease is more common in men than in

women with an average of 1.8 cases notified among men for each woman

globally in 2017 (WHO, 2018). One explanation for the excess of cases among

men is that they have greater access to health care than women. Although this

may contribute to disparity among men and women in some places, TB

prevalence surveys, which control for access to care, find the same male-biased

pattern (Borgdorff et al., 2000). Given that the current paradigm for TB control

depends on case detection, understanding how and why TB disease differs by

sex is critical to informing control programs.

Many explanations for male-bias in TB propose that men have greater

susceptibility to infection or more frequent opportunities for exposure (Neyrolles

& Quintana-Murci, 2009; Nhamoyebonde & Leslie, 2014). A number of factors

have been advanced as possible mechanisms for heightened susceptibility in

men. In most countries, men smoke more cigarettes than women, and per capita

smoking rates explain roughly one-third of the variation in country-level male-bias

in case reports (Watkins & Plant, 2006), perhaps due to toxic lung injury and

reduced immune cell function (Arcavi & Benowitz, 2004) leaving them more

susceptible to infection. Alcohol use is also identified as a risk factor for TB

disease as it may have immunosuppressive effects (Lönnroth et al., 2008).

These behavioral factors and other hormonal and physiological factors likely play a role in determining sex-specific susceptibility to Mtb (Neyrolles & Quintana-Murci, 2009; Nhamoyebonde & Leslie, 2014), though we do not know the combined extent to which they explain male-bias across populations nor do we know the full spectrum of possible mechanisms.

Apart from susceptibility, men may be exposed to undetected, infectious TB cases more often than women. This exposure may be determined in part by the social role men fill and how social roles influence mixing with others in their community (Nhamoyebonde & Leslie, 2014). For instance, in Uganda adult men travel more often than women who mostly identify as housewives (Waroux et al., 2018); the mobility of men may lead to higher exposure rates due to their centrality within social networks. Alternatively, exposure to TB cases could be perpetuated because men preferentially interact with other men who are more likely to be infected than women (Dodd et al., 2016). Sex-assortativity could further magnify spread among men if men were more likely than women to transmit infection to their close contacts, as some household contact studies have found (Hector et al., 2017). Compared with biological differences in susceptibility, few studies have directly examined whether differences in network centrality and mixing patterns by sex can account for the predominance of tuberculosis incidence among men.

To examine the role of network position and social mixing patterns on TB by sex, we performed a cross-sectional study of a large social network in

Kampala, Uganda, that included component social networks of both TB cases and appropriate controls in an area endemic for TB. In this large network, we determined whether TB cases were more central in their networks than index controls (Christakis & Fowler, 2010; Kawatsu et al., 2015) and whether men were more central than women in their social networks. For mixing patterns, we assessed assortativity by sex stratified by TB disease.

Methods

Data collection

We performed a cross-sectional, social network survey from 2012 to 2016 in the Rubaga Division of Kampala, Uganda. Rubaga is an urban area where approximately 300,000 people reside. The burden of TB disease in Uganda is one of the highest in the world (WHO, 2018) and, in Rubaga Division, nearly one half of the population may be latently infected (Kizza et al., 2015). This area's urban landscape, high male:female case bias (2.4:1) (WHO, 2018), and high prevalence of infection make it a relevant place to study the factors affecting TB spread in endemic populations.

To characterize the social networks of people living in Kampala, Uganda, we recruited 123 index cases and 124 index controls without TB and then delineated their social networks through systematic interviews. The study enrolled sputum-smear positive or culture-confirmed pulmonary TB cases who presented to the National Control Programme at Mulago Hospital Complex or were found through active case-finding of reported contacts. Index controls were

recruited by frequency-matching index cases according to age-group, sex, and parish (proxy for residence). For both index cases and controls, the inclusion procedures for enrollment were identical: index participants had to be 15 years or older and reside primarily in Rubaga.

The social networks were ascertained in a two-step process. In the first step, index participants listed members of their households and all individuals living outside their household with whom they had close contact, defined as being within talking distance for more than 4 hours during one or more contact episodes. These first-level contacts were then traced and evaluated for signs of latent TB infection or active TB disease. In the second step, first-level contacts were asked to list their household and extra-household contacts (i.e., second-level contacts of the index participants). Unless there was a suspicion of active TB, field nurses did not trace the second-degree contacts. This sampling methodology was an extended form of egocentric sampling, which we will refer to as "second-level egocentric sampling" in the remaining sections. Second-level egocentric sampling differs from classic egocentric network sampling in the additional layer of contacts collected (Figure 2.1).

Social network analysis

We described the large-scale features of the social network including its size (number of members), component distribution (number of sub-networks connected to each other through common contacts), and mean degree (average number of contacts per individuals). Additionally, we also compared the degree

distribution of index participants, first-level, and second-level contacts and used a two-way ANOVA to compare the degree distribution of index individuals and that of first-level contacts. We compared estimates of node position among index participants by index type (TB case or control) and sex. Centrality statistics included node degree, betweenness (Brandes, 2001), and closeness (Freeman, 1979) (definitions in Table 2.1). To determine whether men were more clustered with TB cases than women, we calculated the network distance to a TB case. We used two-way ANOVAs to determine whether dependent variables were associated with independent variables sex (male or female) and index type (TB case or control).

To assess patterns of mixing by sex and age in the network, I compared the proportion of contacts occurring within-sex, between-sex, and with children in the network for index individuals. I additionally quantified sex-assortative mixing with the assortativity coefficient (Newman, 2003). These coefficients are based on the matrix, E_{ij}, describing the fraction of all edges that connect a node of type i to type j, such that the diagonal E_{ii} represents within-group edges, the off-diagonal represents between-group edges, and $\sum_{ij} E_{ij} = 1$. If $a_i = \sum_j E_{ij}$ (i.e., the proportion of all edges connecting to nodes in each group i) and the network is undirected, the assortativity coefficient is defined as $r = \frac{\sum_i E_{ii} - \sum_i a_i^2}{1 - \sum_i a_i^2}$. When edges occur only between individuals in the same group, only the diagonal of the matrix E_{ij} will be non-zero, e.g., $E_{ij} = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$, and it is straightforward to see that

28

$$r = \frac{(0.5 + 0.5) - (0.5^2 + 0.5^2)}{1 - (0.5^2 + 0.5^2)} = 1,$$

representing perfectly assortative mixing. In contrast, when edges are

concentrated between groups, the diagonal of the matrix will be zero, e.g., $E_{ij} =$

$\begin{pmatrix} 0 & 0.5 \\ 0.5 & 0 \end{pmatrix}$ and the assortativity coefficient is

$$r = \frac{0 - (0.5^2 + 0.5^2)}{1 - (0.5^2 + 0.5^2)} = -1,$$

representing perfectly disassortative mixing. As a last example, the matrix $E_{ij} =$

$\begin{pmatrix} 0.27 & 0.24 \\ 0.24 & 0.25 \end{pmatrix}$ represents approximately random mixing. Here, 51% of edges

attach to group i and 49% attach to group j. The assortativity coefficient for this

matrix illustration would be

$$r = \frac{(0.27 + 0.25) - (0.51^2 + 0.49^2)}{1 - (0.51^2 + 0.49^2)} = \frac{0.52 - 0.5002}{1 - 0.5002} = 0.04,$$

which confirms approximately random mixing between these two groups.

Sensitivity analyses

I examined the sensitivity of estimated centrality statistics to biases that

might be introduced by egocentric network sampling and variation in the

underlying network structure (small-world and scale-free) and population size

(N=50,000, 75,000, 100,000, 125,000 and 250,000). These population sizes

were chosen to represent a sub-population of the Rubaga Division in Kampala.

Specifically, I randomly selected 240 nodes from each unsampled network and

estimated their betweenness, closeness, and degree statistics (true centrality

statistics). Then, I sampled each network with one of three types of egocentric

network sampling (Figure 2.1) using the 240 nodes as egos. Lastly, I estimated

the betweenness, closeness, and degree from the sampled network (sampled

centrality statistics). For each set of network structures (small-world and scale-

free), sizes, and sampling types, I calculated the correlation between the true

node centrality and sampled node centrality estimates. To understand how the

Kampala network related to simulated networks, we calculated the clustering

coefficient (i.e., the probability that neighbors of a node are also connected

(Wasserman & Faust, 1994)) and fit to a power-law degree distribution (Clauset

et al., 2009) because small-world networks are characterized by high clustering

coefficients and scale-free networks by a power-law degree distributions

(Barabasi & Albert, 1999; Watts & Strogatz, 1998). We performed a second

sensitivity analysis focusing on estimates of the assortativity coefficients. We

simulated networks with varying levels of assortativity, randomly chose 240

nodes to represent egos, and sampled the network with a second-level

egocentric design. We then calculated the correlation between sampled

assortativity and true assortativity of the underlying network to assess robustness

to egocentric sampling. Full details of these analyses are provided in Appendix I.

All network analyses and simulations were completed in R (4.0.0) using the

package igraph (Csardi & Nepusz, 2006). R scripts for sensitivity analyses are

available on https://github.com/DrakeLab/miller-tb-centrality.

Ethics considerations

The study was approved by the University of Georgia Institutional Review Board, the Higher Degrees Research and Ethics Committee at Makerere University School of Public Health, and approved by the Uganda National Council for Science and Technology.

Results

Overall network structure

We recruited 247 index participants (123 cases, 124 controls) of which 169 were men and 78 were women. Index participants listed 2,418 contacts (first-level contacts) of which 1,930 agreed to enroll in the study and subsequently identified 9,175 second-level individuals. Index participants identified on average two more contacts than first-level individuals (10.4 vs. 8.2) (Mann-Whitney U-test, p<0.0001). Thus, 2,177 individuals (index participants and enrolled first-level individuals) reported 14,307 edges. Control indexes were almost twice as likely to be in monogamous or polygamous relationships than TB cases (χ^2=13.9, df=1, P=0.0002) (Table 2.2). Male index participants were six years older than female index participants on average ($F_{1, 246}$=26.8, P=4.68$\cdot 10^{-7}$) (Table 2.3).

The resulting network of 11,840 individuals represented 6,507 men, 5,333 women, 9,720 adults (at least 15 years old), and 2,002 children less than 15 years old (age was not identified for 118 individuals). Average overall degree, including second-level contacts, was 2.4 (± 0.03, SE). Despite low connectivity of second-level contacts, all 247 index networks were distributed in one of only 47

31

network components after joining networks with common contacts. One

component connected 9,974 (84%) individuals and 187 (75%) index participants

(102 controls and 85 TB cases).

Node position of index individuals

In our analyses of network centrality, there was little variation in node

position among index participants stratified by type (Table 2.2) and sex (Figure

2.2, Table 2.3). Network betweenness was the only node position metric that

differed between index cases and their controls, whereby controls had higher

betweenness than cases ($F_{1,183}$ = 12.73, P= 0.0005). TB cases tended to have a

higher degree distribution with their first-level contacts than index controls (10.7

versus 10.2, Table 2.2); index males also tended to be closer in network distance

to other index cases than were females ($F_{1,183}$ = 2.78, P= 0.096). Closeness of

index participants was not higher among TB cases compared with index controls

nor among men compared with women. Since a higher proportion of controls

were in monogamous or polygamous relationships, we stratified network position

statistics by relationship status (Table S1.1). There was no difference in network

position (degree, betweenness, closeness, or distance) between index

participants that were single and those in a relationship.

Social mixing patterns of index individuals

The distribution of contacts reported by index participants indicated

preferential mixing by sex and higher levels of contact between women and

children than men and children (Figure 2.3, Tables 2.2 and 2.3). Across the

observed network, same-sex edges were almost twice as common as different-sex edges (9,079 vs. 5,228), with a sex-assortativity coefficient of 0.26 (± 0.01, SE). Of the contacts reported by index participants, the proportion that were adult men varied by index sex ($F_{1, 246}=37.1$, P=$4.23 \cdot 10^{-9}$) and index type ($F_{1,246}=12.3$, P=0.0006) with TB cases and men having higher proportion of their reported contacts with men. The proportion of contacts with adult women varied by index sex ($F_{1, 246}=4.213$, P=0.04) but not index type. Overall, index women reported approximately two times more contact with children than index men (0.30 and 0.15, respectively; $F_{1, 246}=36.5$, P=$5.67 \cdot 10^{-9}$). In addition, index controls reported a higher proportion of contacts with children than TB cases (0.33 vs. 0.26; $F_{1, 246}=11.4$, P= 0.0008). Relationship status was not associated with any mixing variables (Appendix I, Figure S1.1).

Sensitivity analyses

Overall, estimates of node centrality statistics were affected by network type, sampling type, and underlying network size (Figure 2.4). When a scale-free network was sampled, all centrality estimates from egocentric samples were correlated with the true centrality in second-level samples ($\rho > 0.4$), indicating some identifiability of central individuals. In contrast to scale-free networks, when sampling from small-world networks, correlation coefficients were lower indicating less identifiability of central individuals. The social network we analyzed had low clustering $(C = 0.1)$ and a degree distribution consistent with either a power-law (i.e., matching that of scale-free networks) or log-normal

distribution (i.e., sum of multiple normal-distributions), but we could not

distinguish between these two distributions (Appendix I, Figure S1.1, S1.2). In

general, second-level egocentric sampling, as was used in this study, was

superior to ego-only and first-level egocentric sampling. Estimates of node

position from egocentric samples were not highly sensitive to the range of

network sizes. In a separate sensitivity analysis of network assortativity,

estimated assortativity from egocentric networks was highly correlated ($\rho = $

0.999) with true assortativity of underlying networks (Appendix I, Figure S1.3).

Discussion

Social contact networks can alter transmission patterns of infectious

diseases (Arregui et al., 2018; Mossong et al., 2008; Rohani et al., 2010). Here,

we analyzed the structure of a large social network in an urban population in

Kampala, Uganda, to examine the male-bias of TB disease. We observed

assortative mixing patterns that differed by sex and age category. We also found

subtle yet important differences in the social networks of TB cases and

community controls that may account, in part, for the male predominance of TB

disease.

Within the 247 individual ego-centric social networks for each index (both

case and control), we found two inter-related assortative mixing patterns by index

type and sex. In the first pattern, regardless of index type, men listed a greater

proportion of men in their networks, and women listed a greater proportion of

women. In the second pattern, stratified by sex, male and female TB index cases

both listed a greater proportion of male network members. This first pattern of

sex assortment is a common feature of social networks across many different

cultures (Dodd et al., 2016; Horton et al., 2020; Waroux et al., 2018). The

preferential assortment of TB cases with men, however, is a new finding that has

implications for transmission of *M. tuberculosis*. It suggests that there may be

unknown component networks comprising mostly men that perpetuate

transmission of TB in the community. This is an interesting extension to the idea

that that men are source cases to a disproportionate amount of new infections

(Dodd et al., 2016).

When we stratified further by household and extra-household contacts, we

found TB cases, both male and female cases, reported a lower proportion of

household contacts compared with index controls. This finding demonstrates that

extra-household mixing within a network may be an important path for

transmission of *M. tuberculosis* in the community and is consistent with current

models of TB transmission that indicate extra-household transmission

predominates (Kizza et al., 2015; Martinez et al., 2017; Verver et al., 2004). All

together, these observations suggest that the sex-specific social mixing, which is

consistently found among adults in many populations (Horton et al., 2020), may

amplify exposure among men.

When considering category of age, another pattern of assortative mixing

emerged. We found that male indexes had proportionally fewer contacts with

children than female indexes, consistent with other studies of age- and sex-

specific contact patterns observed in Zambia and South Africa (Dodd et al., 2016). Further, within the household, one-quarter to one-third of contacts are with children, thereby creating an environment of intense transmission to children who likely have fewer community interactions. This finding corroborates the known risk of household infection among children (Cruz & Starke, 2010).

As for network position, we found that TB index cases had lower betweenness-centrality than index controls, whereas degree and closeness centrality were similar for both types of indexes. This pattern was present in both male and female indexes. These measures indicate that while TB cases appear to be well connected to others in their networks, they less often bridge relations between two other members in the network. We speculate that this pattern may arise because network members avoid TB cases because of their illness and thereby bypass the role of the case as a 'bridge' in the network. It is also possible that the index controls were more frequently connected to their community and consequently had higher betweenness measures. We found evidence for this as the controls were more numerous than TB index cases in the largest component network.

Our findings are based on partially-sampled social networks. To understand how sampling could affect the estimates of network metrics, we performed sensitivity analyses. We found a higher correlation of centrality estimates from egocentric network samples taken from scale-free networks than from small-world networks, which we expect to be due to higher variation in the

degree distribution of scale-free networks and consequently, increased ability to distinguish between more and less central individuals. Since the social network in the study was not fully sampled, we cannot determine whether the full, underlying network was a small-world or scale-free network. The observed degree distribution, however, suggests a closer alignment with scale-free graphs. Thus, we believe our centrality measures in our egocentric samples are valid, though our findings about node position should be interpreted with caution. Importantly, the sensitivity analyses showed that assortativity statistics are robust to egocentric sampling. Although our simulated networks were not subject to mechanisms of data incompleteness (e.g., imperfect recall) other than the sampling protocol itself (this has been analyzed elsewhere (Kossinets, 2006)) our sensitivity analyses aid in the interpretation of our findings.

In the study, we enrolled prevalent cases of TB from city clinics and a local hospital. It is possible that there is referral bias if these prevalent cases had more advanced disease at presentation than cases at other medical practices in the city. We cannot judge the effect of this bias on our findings, but the results on assortativity and network position should be interpreted in light of our enrolment strategy. Furthermore, our analysis focused on the structure of the social network of the indexes without including the infection status of the contacts. In future analyses, we will test whether increased network centrality is associated with tuberculous infection among contacts (Christley et al., 2005; Kawatsu et al., 2015).

Conclusions

The findings of this study provide novel insight into how social mixing patterns could contribute to the male-bias seen in TB disease. Although men with TB were not more central in their social network than women with TB, the men did have preferential and higher levels of contact with other men in their networks. These findings suggest that contact tracing routines for TB disease and infection should be adapted differently for men and for women to optimize screening of contacts. The findings also raise questions about the patterns of mixing that bring together men with or without TB. Further research is needed to delineate the geographic locations and social settings for this mixing, so that targeted interventions for screening and treatment may be developed.

Author Contributions

PM conceptualized network analysis, cleaned and analyzed data, wrote manuscript; SZ, RK, JS, CW, and NK secured project funding and oversaw data collection; RG, RK, JS, LS, JD, and CW assisted with data cleaning, analysis, and edited manuscript.

References

Arcavi, L., & Benowitz, N. L. (2004). Cigarette Smoking and Infection. Archives of Internal Medicine, 164(20), 2206–2216. https://doi.org/10.1001/archinte.164.20.2206

Arregui, S., Iglesias, M. J., Samper, S., Marinova, D., Martin, C., Sanz, J., & Moreno, Y. (2018). Data-driven model for the assessment ofMycobacterium tuberculosistransmission in evolving demographic structures. Proceedings of the National Academy of Sciences of the United States of America, 115(14), E3238–E3245. https://doi.org/10.1073/pnas.1720606115

Barabasi, A., & Albert, R. (1999). Emergence of scaling in random networks. Science, 286(5439), 509–512.

Borgdorff, M. W., Nagelkerke, N. J., Dye, C., & Nunn, P. (2000). Gender and tuberculosis: a comparison of prevalence surveys with notification data to explore sex differences in case detection. Int. J. Tuberc. Lung Dis., 4(2), 123–132.

Brandes, U. (2001). A faster algorithm for betweenness centrality. Journal of Mathematical Sociology, 25(2), 163–177. https://doi.org/10.1080/0022250x.2001.9990249

Christakis, N. A., & Fowler, J. H. (2010). Social Network Sensors for Early Detection of Contagious Outbreaks. PLoS ONE, 5(9). https://doi.org/10.1371/journal.pone.0012948

Christley, R. M., Pinchbeck, G. L., Bowers, R. G., Clancy, D., French, N. P., Bennett, R., & Turner, J. (2005). Infection in social networks: using network analysis to identify high-risk individuals. American Journal of Epidemiology, 162(10), 1024–1031. https://doi.org/10.1093/aje/kwi308

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. SIAM Review, 51(4), 661–703. https://doi.org/10.1137/070710111

Cruz, A. T., & Starke, J. R. (2010). Pediatric tuberculosis. Pediatrics in Review, 31(1), 13-25-quiz 25–26. https://doi.org/10.1542/pir.31-1-13

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. International Journal of Complex Systems, 1695(5), 1–9.

Dodd, P. J., Looker, C., Plumb, I. D., Bond, V., Schaap, A., Shanaube, K., Muyoyeta, M., Vynnycky, E., Godfrey-Faussett, P., Corbett, E. L., Beyers, N., Ayles, H., & White, R. G. (2016). Age- and Sex-Specific Social Contact Patterns and Incidence of Mycobacterium tuberculosis Infection. American Journal of Epidemiology, 183(2), 156–166. https://doi.org/10.1093/aje/kwv160

Freeman, L. (1979). Centrality in Social Networks I: Conceptual Clarification. Social Networks, 1, 215–239.

Hanneman, R. A., & Riddle, M. (2005). Introduction to social network methods.

Hector, J., Anderson, S. T., Banda, G., Kamdolozi, M., Jefferys, L. F., Shani, D., Garton, N. J., Mwale, A., Jobe, A., Davies, G. R., & Sloan, D. J. (2017). TST positivity in household contacts of tuberculosis patients: a case-contact study in Malawi. BMC Infectious Diseases, 17(1), 259. https://doi.org/10.1186/s12879-017-2348-2

Horton, K. C., Hoey, A. L., Béraud, G., Corbett, E. L., & White, R. G. (2020). Systematic Review and Meta-Analysis of Sex Differences in Social Contact Patterns and Implications for Tuberculosis Transmission and Control. Emerging Infectious Diseases, 26(5), 910–919. https://doi.org/10.3201/eid2605.190574

Horton, K. C., MacPherson, P., Houben, R. M. G. J., White, R. G., & Corbett, E. L. (2016). Sex Differences in Tuberculosis Burden and Notifications in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis. PLoS Med, 13(9), e1002119. https://doi.org/10.1371/journal.pmed.1002119

Kawatsu, L., Izumi, K., Uchimura, K., Urakawa, M., Ohkado, A., & Takahashi, I. (2015). Can social network analysis assist in the prioritisation of contacts in a tuberculosis contact investigation? Int. J. Tuberc. Lung Dis., 19(11), 1293–1299. https://doi.org/10.5588/ijtld.15.0378

Kizza, F. N., List, J., Nkwata, A. K., Okwera, A., Ezeamama, A. E., Whalen, C. C., & Sekandi, J. N. (2015). Prevalence of latent tuberculosis infection and associated risk factors in an urban African setting. BMC Infectious Diseases, 15(1). https://doi.org/10.1186/s12879-015-0904-1

Kossinets, G. (2006). Effects of missing data in social networks. Social Networks, 28(3), 247–268. https://doi.org/10.1016/j.socnet.2005.07.002

Lönnroth, K., Williams, B. G., Stadlin, S., Jaramillo, E., & Dye, C. (2008). Alcohol use as a risk factor for tuberculosis - a systematic review. Bmc Public Health, 8(1), 289. https://doi.org/10.1186/1471-2458-8-289

Martinez, L., Shen, Y., Mupere, E., Kizza, A., Hill, P. C., & Whalen, C. C. (2017). Transmission of Mycobacterium Tuberculosis in Households and the Community: A Systematic Review and Meta-Analysis. American Journal of Epidemiology, 185(12), 1327–1339. https://doi.org/10.1093/aje/kwx025

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., & Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med, 5(3), e74. https://doi.org/10.1371/journal.pmed.0050074

Newman, M. (2003). Mixing patterns in networks. Physical Review E, 67(2). https://doi.org/10.1103/physreve.67.026126

Neyrolles, O., & Quintana-Murci, L. (2009). Sexual Inequality in Tuberculosis. PLoS Med, 6(12), e1000199. https://doi.org/10.1371/journal.pmed.1000199

Nhamoyebonde, S., & Leslie, A. (2014). Biological Differences Between the Sexes and Susceptibility to Tuberculosis. Journal of Infectious Diseases, 209(suppl 3), S100–S106. https://doi.org/10.1093/infdis/jiu147

Rohani, P., Zhong, X., & King, A. A. (2010). Contact Network Structure Explains the Changing Epidemiology of Pertussis. Science, 330(6006), 982–985. https://doi.org/10.1126/science.1194134

Verver, S., Warren, R. M., Munch, Z., Vynnycky, E., Helden, P. D. van, Richardson, M., Spuy, G. D. van der, Enarson, D. A., Borgdorff, M. W., Behr, M. A., & Beyers, N. (2004). Transmission of tuberculosis in a high

incidence urban community in South Africa. International Journal of Epidemiology, 33(2), 351–357. https://doi.org/10.1093/ije/dyh021

Waroux, O. le P. de, Cohuet, S., Ndazima, D., Kucharski, A. J., Juan-Giner, A., Flasche, S., Tumwesigye, E., Arinaitwe, R., Mwanga-Amumpaire, J., Boum, Y., Nackers, F., Checchi, F., Grais, R. F., & Edmunds, W. J. (2018). Characteristics of human encounters and social mixing patterns relevant to infectious diseases spread by close contact: a survey in Southwest Uganda. BMC Infectious Diseases, 18(1), 172. https://doi.org/10.1186/s12879-018-3073-1

Wasserman, S., & Faust, K. (1994). Social Network Analysis: Methods and Applications (1st ed., Vol. 24). Cambridge University Press.

Watkins, R. E., & Plant, A. J. (2006). Does smoking explain sex differences in the global tuberculosis epidemic? Epidemiology and Infection, 134(2), 333–339. https://doi.org/10.1017/s0950268805005042

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. Nature, 393(6684), 440–442. https://doi.org/10.1038/30918

WHO. (2018). Global tuberculosis report. World Health Organization. https://www.who.int/tb/publications/global_report/en/

Table 2.1 Social Network Estimates used to describe the Node Position of Index Individuals within the Kampala Network.

Statistic	Definition	Equation	Notation
Node degree, $k_{s \in 1,2,...n}$	Number of adjacent edges	$\sum_{j=1}^{N} A_{s,j}$	Adjacency matrix, $A_{ij} = 1$, if we identified contact between i, j
Betweenness, $b_{s \in 1,2,...n}$	Number of times node is on shortest path between pairs of other nodes[1]	$\sum_{u \neq s \neq v} \frac{\sigma_{uv}(s)}{\sigma_{uv}}$	σ_{uv} is the total number of shortest paths from node u to v and $\sigma_{uv}(s)$ is the number of those paths that pass through s
Closeness, $c_{s \in 1,2,...n}$	Inverse of the average length of shortest path to all other nodes[1]	$\frac{1}{\sum_{i \neq s} d_{si}}$	d_{si} is the network distance between nodes s and i
Distance to TB case, $y_{s \in 1,2,...n}$	Network distance to a TB case[1]	$\min(d_{st, t \neq s})$	t is the set of TB cases

[1]Network distance, closeness, and betweenness were calculated within the giant component because path length is not defined for disconnected graphs.

Table 2.2 Social network statistics among index individuals

Demographics and Social Network Estimates for Index Individuals stratified by Index type. Values indicate the number of individuals (proportion) or mean ($\pm$ standard errors) for each variable.

	Case n=123	**Control** n=124	**Sig.**
Age	30.6 ($\pm$0.90)	32.0 ($\pm$0.85)	
Monogamous or polygamous relationship	40 (0.33)	70 (0.56)	◊
Node position			
Degree	10.7 ($\pm$0.36)	10.2 ($\pm$0.36)	
Closeness	0.076 ($\pm$0.001)	0.077 ($\pm$0.001)	
Betweenness	0.009 ($\pm$0.002)	0.02 ($\pm$0.002)	◊
Distance to case	3.2 ($\pm$0.2)	3.5 ($\pm$0.2)	
Mixing variables			
Proportion of all contacts with adult men	0.47 ($\pm$0.02)	0.36 ($\pm$0.03)	◊
Proportion of all contacts with adult women	0.37 ($\pm$0.02)	0.40 ($\pm$0.02)	
[1]Proportion of all contacts with children	0.16 ($\pm$0.02)	0.23 ($\pm$0.02)	◊
[2]Proportion of all contacts occurring within HH	0.23 ($\pm$0.02)	0.37 ($\pm$0.03)	◊
Proportion of HH contacts occurring with children	0.28 ($\pm$0.02)	0.32 ($\pm$0.02)	

[1] Adults $\geq$ 15 years old, children <15
[2] HH: Household

Table 2.3 Social network statistics among index individuals by sex

Demographics and Social Network Estimates for Index Individuals stratified by Index Type and Sex. Values indicate the number of individuals (proportion) or mean ($\pm$ standard errors) for each variable.

	Female		Male		
	Case n=39	**Control** n=39	**Case** n=84	**Control** n=85	**Sig.**
Age	25.6 ($\pm$0.89)	28.0 ($\pm$1.11)	32.9 ($\pm$1.16)	33.8 ($\pm$1.08)	*
Monogamous or polygamous relationship	12 (0.31)	22 (0.56)	28 (0.33)	48 (0.57)	◊
Node position					
Degree	10.4 ($\pm$0.70)	9.95 ($\pm$0.4)	10.8 ($\pm$0.4)	10.3 ($\pm$0.5)	
Closeness	0.077 ($\pm$0.002)	0.076 ($\pm$0.003)	0.075 ($\pm$0.002)	0.078 ($\pm$0.002)	
Betweenness	0.005 ($\pm$0.001)	0.015 ($\pm$0.003)	0.010 ($\pm$0.002)	0.022 ($\pm$0.003)	◊
Distance to case	3.4 ($\pm$0.3)	3.8 ($\pm$0.3)	3.0 ($\pm$0.2)	3.3 ($\pm$0.2)	
Mixing variables					
Proportion of all contacts with adult men	0.34 ($\pm$0.04)	0.22 ($\pm$0.02)	0.53 ($\pm$0.03)	0.43 ($\pm$0.03)	*◊
Proportion of all contacts with adult women	0.41 ($\pm$0.04)	0.44 ($\pm$0.03)	0.35 ($\pm$0.02)	0.38 ($\pm$0.02)	*
[1]Proportion of all contacts with children	0.26 ($\pm$0.03)	0.34 ($\pm$0.03)	0.12 ($\pm$0.02)	0.19 ($\pm$0.02)	*◊
[2]Proportion of all contacts occurring within HH	0.31 ($\pm$0.04)	0.45 ($\pm$0.04)	0.19 ($\pm$0.02)	0.33 ($\pm$0.03)	*◊
Proportion of HH contacts occurring with children	0.32 ($\pm$0.04)	0.39 ($\pm$0.03)	0.25 ($\pm$0.02)	0.2 ($\pm$0.02)	*

[1] Adults $\geq$ 15 years old, children <15

[2] HH: Household

∗ Significant difference (p <0.05) between means by index sex (male, female)

◇ Significant difference (p <0.05) between means by index type (case, control)

Ego–only   First–level egocentric   Second–level egocentric
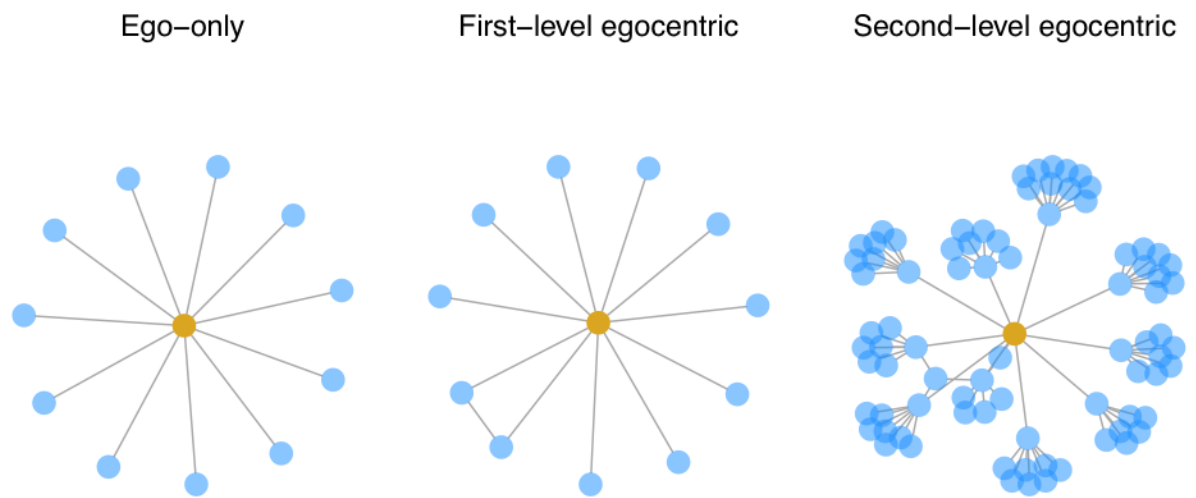
Figure 2.1 Types of egocentric network sampling.

Types of egocentric network sampling varies in the amount of information collected outward from the index individual. Index individuals, shown in yellow, list their contacts in ego-only sampling. In first-level egocentric sampling, index individuals additionally indicate whether their contacts are also contacts. In second-level egocentric sampling, contacts of index individuals list their own contacts. The social network in the Kampala study utilized a second-level egocentric design. Sensitivity analyses compare network statistics across different types of egocentric sampling.
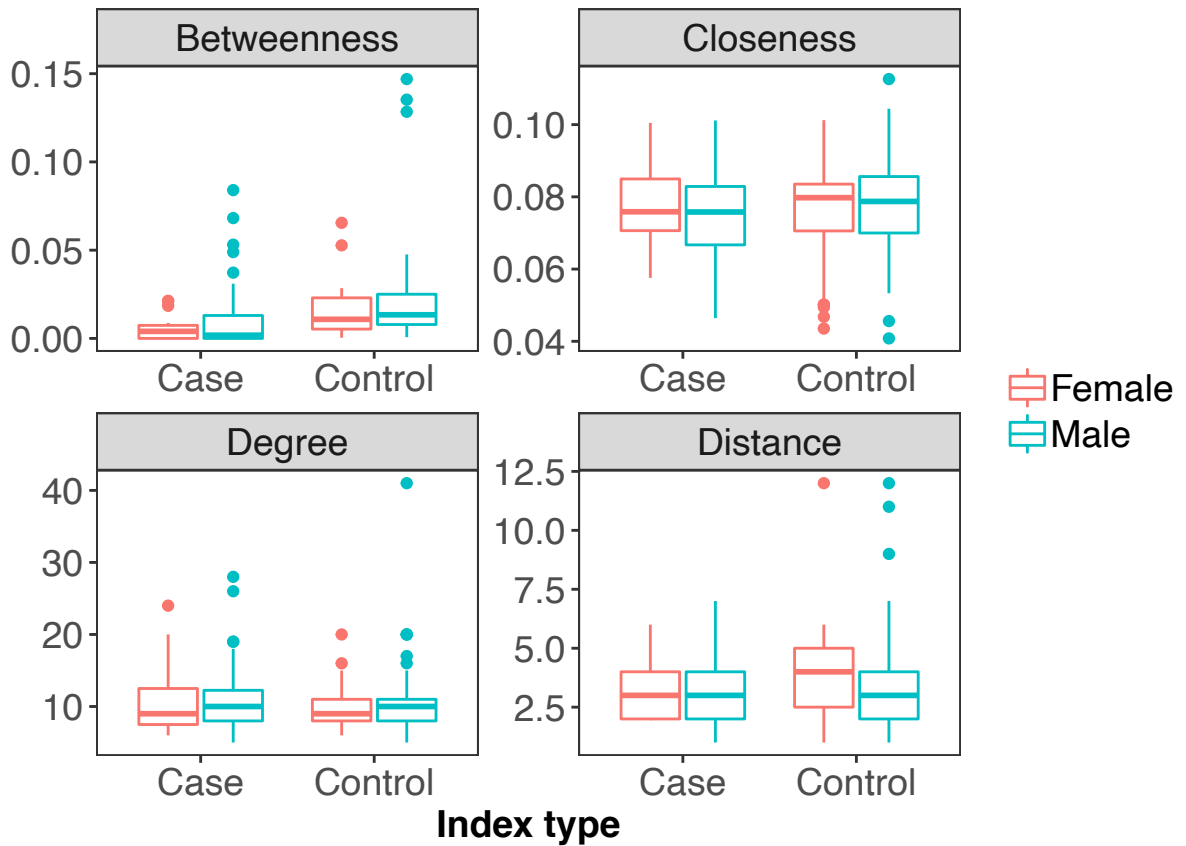
Figure 2.2 Individual-level node statistics among index cases.

We failed to detect a relationship between individual-level characteristics (TB infection and sex) and statistics describing a node's position in social networks (degree, betweenness, closeness, and network distance to TB cases). Closeness, betweenness, and network distance to TB cases were calculated for index individuals in the giant component (i.e., the largest connected network component). Boxplots show interquartile regions and outliers.
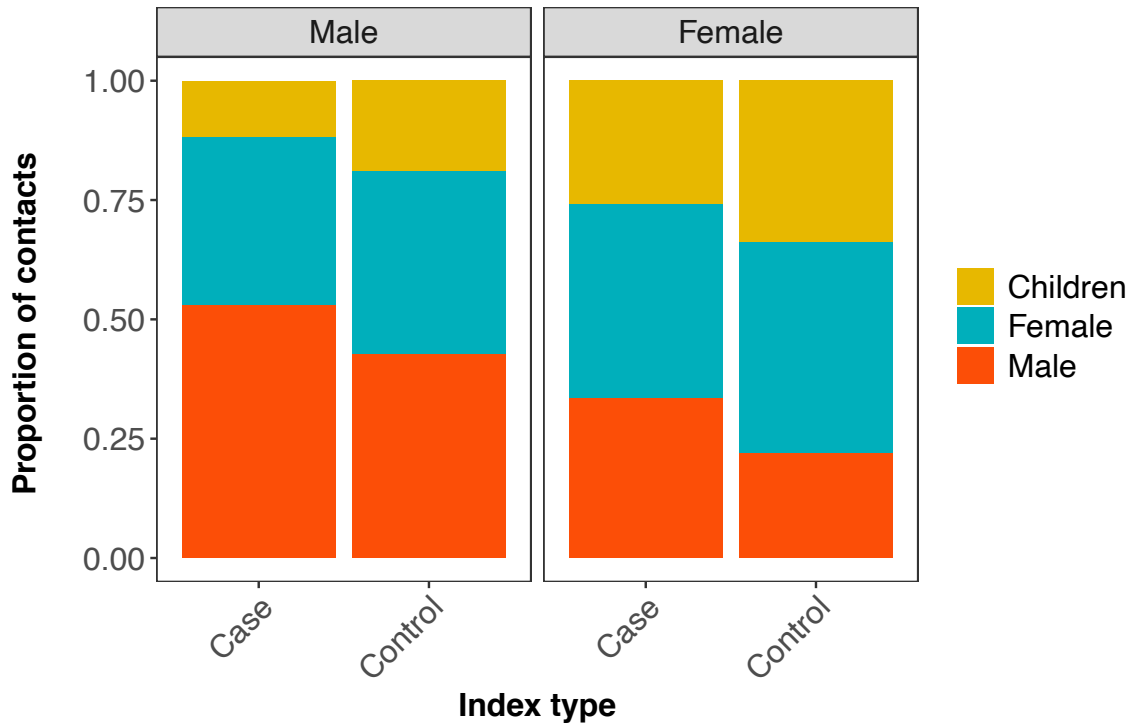
Figure 2.3 Sex-assortative mixing among index cases.

Proportion of all contacts of index individuals that are with adults and children shows there to be high assortativity in the Kampala network. There were more within-sex edges than between-sex edges. There were also more edges between adult females and children than adult males and children.
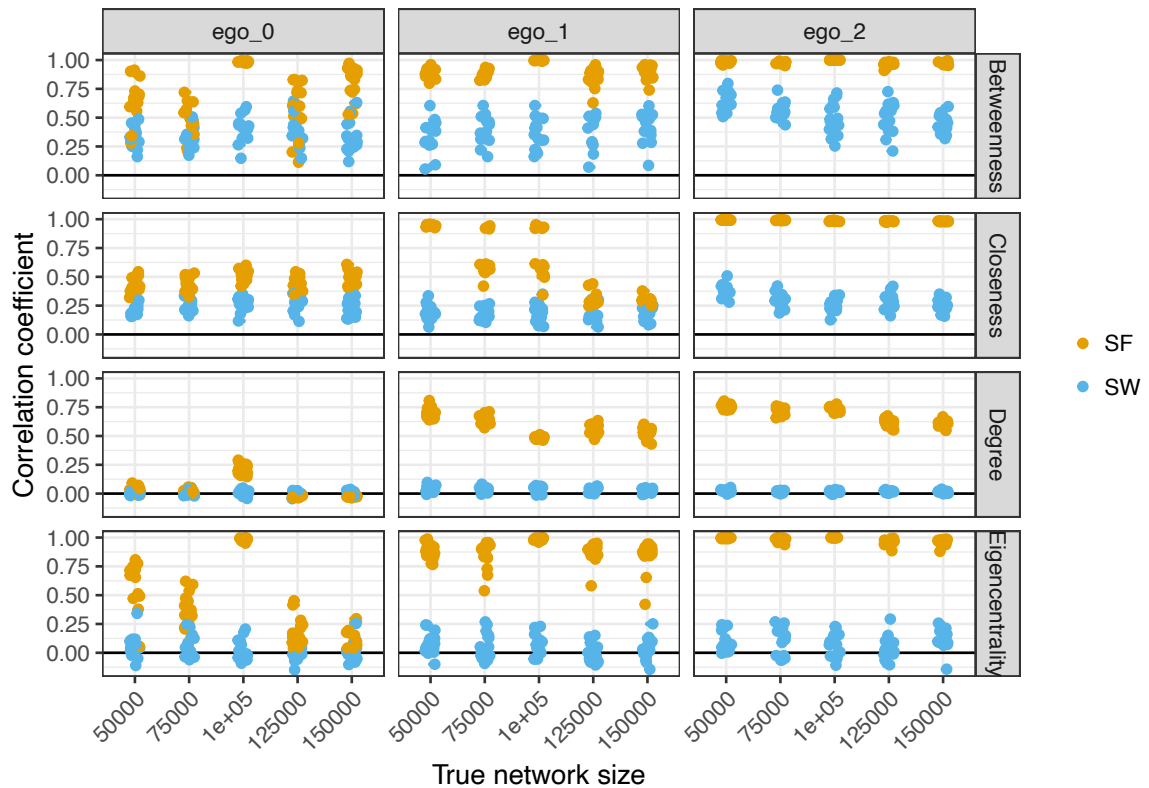
Figure 2.4 Effects of network sampling on centrality statistics.

The correlation between true node centrality and estimated node centrality depends on the underlying network being sampled and the type of egocentric sampling used. We simulated 15 replicates of each network type (scale-free and small-world) across a range of network sizes $(5 \cdot 10^4, 7.5 \cdot 10^4, 1 \cdot 10^5, 1.25 \cdot 10^5, 1.5 \cdot 10^5)$ all with mean node degree of 10. We then simulated the process of three types of egocentric sampling (ego_0, ego_1, and ego_2) and calculated the correlation of estimated centrality with true centrality. The black line indicates no correlation between sampled node statistics and true node statistics and the red triangle shows the mean across all replicates. Since we assumed perfect recall, we calculated the correlation in sampled degree on all nodes in the sampled network (i.e., not just the ego). Betweenness of egos was estimated from simulated networks by capping the number of search algorithm of shortest paths to 25.

Chapter 3

EFFECTS OF ASSORTATIVE MIXING AND SEX-TRAITS ON MALE-BIAS IN

TUBERCULOSIS: A MODELLING STUDY[2]

Abstract

Globally, Tuberculosis disease (TB) is more common among males than females. Recent research proposes that differences in preferential social mixing by sex, or sex-assortativity, can alter infection patterns in TB. We conducted a simulation study to see whether sex-assorted mixing patterns can explain the global ratio of male:female TB cases and what factors might cause sex-disparities in infectious diseases to be sensitive to assortative mixing. Simulations showed sex-assortativity alone cannot cause sex-bias in TB. However, we find an effect of interaction between assortativity and sex-traits that suggests a role for behaviour to influence sex-specific epidemiology of infectious diseases. In our study, the role of sex-assortativity was especially apparent for slower spreading infectious diseases, like TB. We also examined how assortativity and sex-traits affect the final outbreak size and other epidemic dynamics. These results are important for understanding when sex-assortativity, a common feature across human populations, can change epidemiological patterns.

Introduction

Tuberculosis (TB) is now the leading cause of death due to infectious diseases globally, and notification data show that, on average, 1.8 male cases are reported for every female case (WHO, 2018). This pattern is strikingly consistent across all regions of the world with male:female ratios below 1 being extremely rare (Neyrolles & Quintana-Murci, 2009). Male-bias is also seen in

52

adults of all ages but does not seem to apply to children (Guerra-Silveira & Abad-Franch, 2013). Differences in access to healthcare are not associated with this pattern as male-bias is observed in surveys using active case-finding (Borgdorff et al., 2000; Salim et al., 2004). Moreover, male-bias is observed in low and high income countries alike (Neyrolles & Quintana-Murci, 2009), further reducing the likelihood that male-bias is primarily driven by differences in access to healthcare since access to healthcare should be more equal in higher income countries. In fact, TB is not unique in being male-biased (among adults, 9 out of 11 infectious diseases were found to be male-biased (Guerra-Silveira & Abad-Franch, 2013). Understanding why sex-bias arises, in both TB and other infectious diseases, has widespread implications for basic research on sex-differences in disease and treatments (Clayton, 2016), public health (Organization, 2007), and more realistic models and predictions of disease burden.

What causes male-bias in infectious diseases, and in TB, specifically? Proposed mechanisms are often categorized into "biological" or "social" (Guerra-Silveira & Abad-Franch, 2013; Organization, 2007). Hypothesized biological mechanisms for male-bias in TB are primarily related to a suspected higher male susceptibility to infection. For instance, female cells have two X chromosomes, which encode genes involved with both the innate and adaptive immune system and are thought to reduce susceptibility of females to a number of pathogens (Schurz et al., 2019). In addition, the female hormone estradiol enhances, while testosterone downregulates, macrophage activation which is an important

pathway for initiating the innate immune response and consequently detecting M.

tuberculosis (reviewed in (Schurz et al., 2019)). Another common explanation for

increased male susceptibility to TB is smoking, which is more common among

men (Islami et al., 2015) and can lead to damaged lung tissue (reviewed in

(Prevention, 2010)). Indeed at the country-level, adult smoking rates explain up

to one-third of variation in male-bias (Watkins & Plant, 2006). Other than

susceptibility, there are additional, lesser studied biological mechanisms that

could plausibly lead to male-bias. For instance, males are more likely to spread

infection to their contacts than females (Dodd et al., 2016; Hector et al., 2017),

indicating higher male transmissibility. Finally, the length of time from disease to

treatment can vary by sex, with males generally delaying care for longer period

than females (Meintjes et al., 2008) suggesting males are infectious in the

community for a longer period. Plausibly, therefore, these different sex-traits --

susceptibility, transmissibility, and infectious period -- could lead to male-bias in

TB.

Gender-roles and preferences in social contacts may also cause males

and females to have different exposure patterns (Dodd et al., 2016; Horton et al.,

2020; Nhamoyebonde & Leslie, 2014). For example, one study found adult males

travelled outside their village 75% more often than females and more than one-

quarter of females identified as housewives, although there was no difference in

the total number of contacts by sex (Waroux et al., 2018). This suggests if social

networks play a role in male-bias, it's not the number of social contacts but the

pattern of social contacts that matters for transmission. Assortative mixing by sex is a common phenomenon across many cultures and because males are a higher incident demographic group than females, this social network structure may be important for understanding the basis for male-bias of TB (Dodd et al., 2016; Horton et al., 2020). Whether biological sex-traits or assortative-mixing by sex, have an outsized effect on male-bias at the population-level is the focus of this modelling study.

Infectious disease transmission models can help sort out the importance of various biological and social factors on sex-bias in infection. In this study, we use mathematical models of disease spread on social networks to examine the relative differences in sex-traits and preferential mixing by sex (i.e., assortativity), independently and in combination, required to give rise to observed levels of male-bias as seen in TB. We were also interested in whether the unique life history of human TB, with its long and variable latent period and endemic levels of infection in some regions, mediates the effects of sex-traits and assortative mixing on male-bias. To investigate these questions, we conducted a comparative simulation study of multiple transmission patterns (SIR, SLIR, SIRS, and SLIRS) spreading on contact networks that varied from random to extremely sex-assortative. Sex-traits investigated were sex-specific susceptibility, transmissibility, and infectious period.

Methods

General approach

      We modelled the effects sex-assortativity (r, defined below) and sex-traits on male-bias. The sex-traits we considered were susceptibility, transmissibility, and infectious period. Susceptibility was defined as the rate of becoming infected given contact with an infected neighbour in the network. Transmissibility was the rate of infecting a susceptible neighbour in the network. Infectious period was defined as the period of time spent in the infected class before recovering. We quantified male-bias as the number of infected males divided by the number of infected females.

Network simulation

      In simulated social networks, nodes represent individuals and edges between them represent repeated interactions between nodes on which infection can spread. To measure assortativity of simulated networks, we used Newman's discrete assortativity coefficient (Newman, 2003). As in Chapter 2, these coefficients are based on the matrix, E_{ij}, describing the fraction of all edges that connect a node of type i to type j we use $a_i = \sum_j E_{ij}$ (i.e., the proportion of all edges connecting to nodes in each group i). The assortativity coefficient is defined as $r = \frac{\sum_i E_{ii} - \sum_i a_i^2}{1 - \sum_i a_i^2}$. Because perfectly disassortative networks would be disconnected components, r ranges from approximately -1 (nearly perfectly disassortative) to 1 (perfectly assortative) with zero representing random mixing.

To simulate social networks with varying levels of sex-assortativity, we used an algorithm presented in (Sah et al., 2014). The Sah algorithm is designed to simulate assorted networks that maintain network structures which alter epidemic dynamics including average clustering, path length, and degree assortativity (Badham & Stocker, 2010). We simulated Sah networks with geometric degree distributions because the algorithm was unstable with other degree distributions (e.g., Poisson and power-law) and two modules (i.e. sexes) which was the focus of this study. To determine whether networks with other degree distributions affected results, we conducted sensitivity analyses using networks generated with a simple rewiring scheme whereby we rewired between-sex edges of small-world and scale-free networks until desired levels of assortativity were reached and made sure the resulting networks were simple (i.e., no multiple edges or self-loops) and connected (i.e., only one component). Additional details on the rewiring algorithm are given in Appendix II. We chose these networks because they represent realistic human interaction networks (Eubank et al., 2004). All networks were initialized with 1000 nodes (500 male, 500 female) and had a final mean degree of 10. We compared how network structural characteristics were affected by increases in sex-assortativity for both network algorithms.

Disease model

To study disease processes affecting sex-bias in TB, we varied parameters within a Susceptible-Latent-Infectious-Recovered-Susceptible

(SLIRS) model framework corresponding to different assumptions about disease transmission (Table 3.1). For example, to incorporate latent tuberculosis infections, we turned on/off latent infection by changing the σ parameter ($\sigma \ll \infty$ leads to a SLIR model whereas $\sigma \to \infty$ leads to a SIR model). For models with latent infection (SLIR and SLIRS), we set $\sigma = 0.1$ corresponding to an average latent period of 10 months. The duration of latent infection in tuberculosis is highly variable but the majority of individuals that develop contagious forms of tuberculosis progress within a year (reviewed in Behr et al. 2018). Similarly, to represent endemic levels of infection where "new" susceptibles reenter contact networks over longer time periods, we varied the ϕ parameter ($\phi \to 0$ leads to a SIR model whereas $\phi > 0$ leads to a SIRS model). In SIRS and SLIRS models, we assumed new susceptibles were encountered at rate $\phi = 0.1$, corresponding to a new contact every 10 months. We assumed the infectious period lasted 6 months ($\gamma = 0.5$) representing a typical treatment delay of 1-3 months (Sendagire et al. 2010) and the period of time to complete the intensive phase of typical tuberculosis treatment regimens of 2-3 months (Nahid et al. 2016). Finally, to understand how overall pathogen transmissibility and corresponding R_0 affect results, we varied the overall transmission rate, τ. In SIR models, the analytical solution for the epidemic threshold (i.e., when $R_0 = 1$) is given by $\frac{\tau <K>}{\gamma <K^2 - K>} > 1$ where K is the set of all node degree values in a network, K^2 is the set of all node degree values squared, and brackets indicate the mean of values in the set. Reproductive estimates for TB range from 0.24 to 4.3 (Ma et al., 2018). In

simulations, varied values of R_0 from 0.5 to 3.5. We confirmed the epidemic threshold numerically (Appendix II, Figure S2.1). Thus, sensitivity analyses investigate different pathogen life histories and transmission rates.

To study how sex-bias could be generated though differences in male and female sex traits we varied the strength (α) of each sex-trait. For susceptibility, we multiplied male and female transmission rates, τ_m and τ_f, depending on the sex of the *target* node in the S-I edge pair. Specifically, for the male:female susceptibility ratio, α_s, we solved the following equations

$$\tau_m = \alpha_s \tau_f$$
$$\frac{\tau_m + \tau_f}{2} = \tau,$$

which hold the overall susceptibility rate constant and results in the following solution:

$$\tau_m = \frac{2\alpha_s \tau}{1 + \alpha_s}$$
$$\tau_f = \frac{2\tau}{1 + \alpha_s}.$$

Thus, when $\alpha_s = 1$, $\tau_m = \tau_f = \tau$ and when $\alpha_s > 1$, the average susceptibility is still τ. Sex differences in transmissibility were modeled in a similar way, adjusting the rates depending on the sex of the *source* node in the S-I edge pair. Sex differences in the duration of the infectious period (inverse of the recovery rate) were modeled by changing the male and female γ parameters and holding the average infectious period constant.

We replicated each network type and disease-model parameter combination 250 times and initiated each simulation with one randomly chosen infected node. We ran SIR and SLIR simulations until there were zero infected individuals and SIRS and SLIRS simulations for 300 time steps. Simulations were implemented with a continuous-time Gillespie algorithm with exponentially-distributed waiting times using the Epidemics on Networks (Miller & Ting, 2020) and Networkx packages (https://networkx.github.io/) in Python (Version 2.7.17). The Gillespie algorithm is a common method to simulate stochastic processes exactly without approximation (Kiss et al., 2017) and while it can be inefficient for large systems, implementation with the Epidemics on Networks package was computationally tractable in this situation. Additional details about how disease spread was simulated in networks are provided in Appendix II.

Analysis

To measure male-bias, we calculated the number of males infected over the course of the epidemic for SIR and SLIR model structures and as the equilibrium ratio of male to female infections in the SIRS and SLIRS model structures. For the SIR and SLIR models, simulations were run until there were no infected individuals left. For the SIRS and SLIRS models, simulations were run until there were no infected individuals left or for 250 timesteps (whichever came first). For the SIRS and SLIRS models, we performed preliminary analyses to determine when simulations reached endemic levels of infection and parameters required to lead to 25% of the population having latent infection at

equilibrium (Houben & Dodd, 2016). To compare the effects of assortativity and heterogeneity in individual-level infection on epidemic dynamics, we calculated the final size, epidemic duration, equilibrium latent and infected prevalence for each simulation.

We used R Version 4.0.0 for analyses and visualizations. All Python and R scripts are available at github.com/drakelab/miller-tb-assortativity.

Results

In simulated networks, the proportion of within-sex contact increased with r, from 45% when r =0 to 77% when r =0.6 (Figure 3.1). Results from a meta-analysis of the proportion of within-sex mixing among adults (Horton et al., 2020), correspond to values of sex-assortativity from r =0.2 to 0.3, which also aligns with an independent estimate of sex-assortativity for a social network in Uganda (Chapter II of this dissertation).

Effects of sex-traits and assortativity on male-bias

In simulations, sex-assortativity did not lead to male-bias. This result was not sensitive to model type (SIR, SLIR, SIRS, SLIRS) or network type (Sah networks, rewired small-world, rewired scale-free) (Figure 3.2, Figure 3.3, Appendix II, Figures S2, S3).

Combined with sex-traits (susceptibility, transmissibility and infectious period), however, assortativity increased male-bias (Figure 3.2). The first sex-trait that we investigated, increased male susceptibility (SUS), led to male-bias in the absence of sex-assortativity but epidemics on assorted networks had higher

male-bias than networks without sex-assortativity (Figure 3.2). This result was not sensitive to inclusion of a latent class (Appendix II, Figure S2.3). The interactive effect of sex-assortativity and sex-traits on male-bias grew with the strength of male:female susceptibility. This was observed across all model types: without sex-assortativity median values of male-bias are below 1.8 while median values of male-bias with sex-assortativity can exceed 1.8. The amplification effect of sex-assortativity on male-bias was not as pronounced in rewired scale-free networks as rewired small-world networks or Sah networks (Appendix II, Figure S2.2).

The interaction of assortativity and sex-traits on male-bias was especially notable in the case of increased male transmissibility (TRA). Without sex-assortativity, even when males had more than three-fold higher transmissibility, male infections were no more likely than female infections (Figure 3.2, Appendix II, Figure S2.2). Overall, however, higher male transmissibility rarely resulted in ratios of male-bias observed in global TB data.

Longer male infectious periods (IP), similar to higher male transmissibility, did not lead to male-bias in SIR and SLIR epidemics unless taking place on sex-assorted networks (Figure 3.2, Appendix II, Figure S2.3). In the parameter ranges investigated here, median values of male-bias for simulations of longer male infectious periods in SIR and SLIR models were all below 1.8. However, when recovered individuals can re-enter the susceptible population (i.e., SIRS and SLIRS models), longer male infectious periods can result in male-bias and

there is a slight increase in male-bias on assorted networks. In SIRS and SLIRS

models, male-bias due to longer male infectious periods was possible. These

results were not sensitive to network type (Appendix II, Figure S2.2).

Transmission rate increased the effect of assortativity on male-bias

(Figure 3.3). For a slow spreading pathogen ($\tau = 0.04, R_0 \approx 1.5$) and male

susceptibility twice that of female susceptibility, male-bias increased from a

median of 1.48 without assortative mixing to 1.85 with strong assortativity (r

=0.6) indicating a 25% increase in male-bias with assortative mixing. For a faster

spreading pathogen ($\tau = 0.1, R_0 \approx 3.5$) with the same level of higher male

susceptibility, male-bias only increased from a median of 1.28 to 1.34 indicating a

5% increase in male-bias with assortative mixing. Similar relationships were

observed for the other sex-traits, transmissibility and infection period.

Effects of sex-traits and assortativity on epidemic outcomes

In general, on Sah contact networks epidemic dynamics were not affected

by sex-assortativity or sex-traits (infectious periods, susceptibility, and

transmissibility) including the peak size, final size, and duration for SIR or SLIR

epidemics (Appendix II, Figure S2.4). Higher male susceptibility slightly reduced

the final size of epidemics for faster spreading pathogens but the effect was

small (Figure 3.4). In contrast to results on Sah networks, assortativity was

associated with changes in peak size, final size, and duration on rewired network

(Figure 3.5). We note, however, that while networks generated with the Sah

algorithm had stable network structures as sex-assortativity increased, rewired

networks did not (Appendix II, Figure S2.5). As assortativity increased from r =0 to r =0.6, clustering increased by approximately 10% in scale-free networks and decreased by approximately 60% in small-world networks. With increasing assortativity, average network path length increased by about 10% in scale-free networks and decreased by approximately 25% in small-world networks. In both small-world and scale-free networks, degree-assortativity increased as sex-assortativity increased.

Discussion

Social mixing patterns can alter transmission patterns of infectious diseases (Arregui et al., 2018; Mossong et al., 2008; Rohani et al., 2010). We conducted a comparative simulation study to see whether sex-assorted mixing patterns can explain the global ratio of male:female TB cases. Simulations showed sex-assortativity alone does not cause sex-bias in TB. However, an interaction between assortativity and sex-traits does affect the ratio of male to female infections suggesting a role for behaviour to influence sex-specific epidemiology of infectious diseases. The role of sex-assortativity was especially apparent for slower spreading infectious diseases, like TB (Blower et al., 1995). We also examined the role of sex-assortativity and sex-traits on the final outbreak size and other overall epidemic dynamics.

Our main result showed that subtle but widespread patterns in sex assortativity may shape sex-specific epidemiological patterns. Approximately 55-65% of human social interactions occur within-sex (Dodd et al., 2016; Horton et

al., 2020), which we showed corresponds to assortativity coefficients of 0.2 to

0.4. Within this range, our simulations suggested that assortative mixing can

change sex-specific epidemiological patterns when there are underlying

heterogeneities in sex-traits, especially for slower spreading infectious diseases.

For example, SIR simulations showed longer male infectious periods and higher

male transmissibility only produced male-biased infection patterns in the

presence of moderate (greater than 0.3) values of sex-assortativity. On the other

hand, higher male susceptibility could lead to male-bias in infection alone.

However, male-bias was higher in outbreaks on sex-assorted networks

compared with non-assorted networks. Our finding that slower spreading

infectious diseases were more sensitive to differences in assortativity is similar to

previous results, which suggested slower spreading pathogens experience

structural trapping (i.e., stochastic extinction) while faster spreading pathogens

experience structural delay (i.e., spread between subgroups is merely delayed)

(Sah et al., 2017). Overall these results suggest that sex-assortativity can

increase the effects of infection differences between males and females on sex-

disparities in infectious diseases, especially for slow spreading pathogens, like

TB.

  Our conclusion that sex-assortativity can increase sex disparities in TB

hinges on there being relatively large differences in sex-traits, defined here as

susceptibility to infection, length of the infectious period, and rate of

transmissibility to contacts. Multiple authors present strong evidence that there

may be differences in immunity and infection for TB between males and females (Guerra-Silveira & Abad-Franch, 2013; Nhamoyebonde & Leslie, 2014; Rhines, 2013; Salim et al., 2004) but the relative difference in specific rates is a complex question. The first of the three traits investigated here, higher male susceptibility, has been studied experimentally. In one study, castration reduced infection following exposure by half in male mice but doubled infection following exposure in female mice (reviewed in (Nhamoyebonde & Leslie, 2014)). Susceptibility is also linked to male-dominated risk factors such as smoking (Bates et al., 2007) and alcohol use (Lönnroth et al., 2008). However, some household studies have found no difference in incidence of TB within households of infectious cases (Guwatudde et al., 2003). These converging lines of evidence make it difficult to ascertain the overall difference in male:female susceptibility, but it is likely a crucial determinant of male bias in TB.

We also explored the effects of higher male transmissibility and longer male infectious periods, although there is less evidence for these mechanisms than for differences in susceptibility. With regards to transmissibility, the proportion of infections caused by males was estimated to be 1.3 to 1.8 times higher than infections caused by females in South Africa and Zambia (Dodd et al., 2016). Additionally, household contacts of male cases had a higher prevalence of latent infection than female cases (ORs of 4.05 in univariate analyses; 7.62 in multivariate analyses) (Hector et al., 2017). However, our simulations showed no evidence that transmissibility can generate sex-

disparities, unless the relative difference in male and female transmissibility was large, similar to modelling results in (Perkins et al., 2008). The last trait, which assumes there could be variation of the infectious period of TB, is a developing area of research. Males are more likely to delay care and diagnosis (Meintjes et al., 2008), possibly resulting in longer periods of infectiousness in the community. Our simulations indicated for sex differences in the infectious period to cause sex disparities in infection, the difference in infectious period would have to be large and sex assortative mixing would be required. In reality, there are still many unknowns about the nature of the infectious period in TB cases generally (Drain et al., 2018; Xu et al., 2019) and whether and how the infectious period could vary between sexes is an open area of work. In addition, a combination of these traits may culminate to produce the consistently male-biased case notification data we see for TB. Future experimental and epidemiological studies are needed to better quantify the differences immunity and infection rates by sex because they all could have different implications for control programs.

Overall epidemic dynamics, such as final outbreak size, peak timing, and outbreak duration, can also be effected by assortativity (also known as modularity, or social grouping) in some situations (Nadini et al., 2018; Sah et al., 2017; Salathé & Jones, 2010). In our simulations, however, with only two groups (male and female) and moderate assortativity, there are few differences in overall dynamics. Previous studies that have found assortativity to alter the final outbreak size have mostly examined the situation where there are many groups

67

with high levels of assortativity (0.8-0.95) (Sah et al., 2017; Salathé & Jones, 2010). Differences in the direction of change attributed to assortativity can be explained by different assumptions about immunity (Nadini et al., 2018) and also whether high levels of assortativity in realistic contact networks are associated with increased network clustering (Salathé & Jones, 2010) which has the effect of lower overall outbreak size (Badham & Stocker, 2010). We find similar results in our simulations. For example, our rewired scale-free networks increased in clustering with increased assortativity and found assortativity to decrease final size. Thus, our simulations further aid in understanding the situations when assortativity can affect important outcomes of outbreaks.

While our simulations tested multiple model assumptions, and our results were not sensitive to parameters chosen, these simplistic models do not fully capture the complexities of TB, especially in light of recent advances in our understanding of the spectrum of infection and disease (Drain et al., 2018). In addition, our social contact networks omit age-specific infection rates and age-specific mixing patterns, which are important for accurately estimating TB burden in a population (Arregui et al., 2018). Because the aim of this study was to test a general phenomenon, our models were not parameterized for specific populations. Epidemiologically-relevant demographic variables, such as the reproductive rate, vary across populations (Ma et al., 2018), which may influence the applicability of these results to specific populations. The goal of our modelling study was to offer qualitative insight into whether assortativity may play a role in

causing sex-disparities in TB and other possible pathogen scenarios that might be affected by sex assorted mixing.

Conclusions

These results provide insight into how behaviour can amplify the consequences of evolutionary trade-offs between sex and immunity to infection. Although we focused on TB, many infectious diseases are male-biased (Guerra-Silveira & Abad-Franch, 2013) and most populations have social mixing patterns marked by sex-assortativity (Horton et al., 2020). We conclude that heterogeneity in sex rates, especially differing susceptibility, is more important to sex-disparity in infectious diseases than sex-assortativity, but mixing patterns can amplify the effects of sex-traits in some cases. For TB, important questions arise about whether differences in susceptibility and other sex-traits, are similar to levels analyzed here or if there are remaining factors driving sex-disparities in TB. For practical purposes, results from this study shed light on when it could be inappropriate or misleading to extrapolate infection risk or rates across sexes for different models.

Author Contributions

PM designed and implemented simulation study, analyzed data, wrote manuscript; JD designed simulation study and edited manuscript; CW edited manuscript.

References

Arregui, S., Iglesias, M. J., Samper, S., Marinova, D., Martin, C., Sanz, J., & Moreno, Y. (2018). Data-driven model for the assessment ofMycobacterium tuberculosistransmission in evolving demographic structures. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(14), E3238–E3245. https://doi.org/10.1073/pnas.1720606115

Badham, J., & Stocker, R. (2010). The impact of network clustering and assortativity on epidemic behaviour. *Theoretical Population Biology*, *77*(1), 71–75. https://doi.org/10.1016/j.tpb.2009.11.003

Bates, M. N., Khalakdina, A., Pai, M., Chang, L., Lessa, F., & Smith, K. R. (2007). Risk of Tuberculosis From Exposure to Tobacco Smoke: A Systematic Review and Meta-analysis. *Archives of Internal Medicine*, *167*(4), 335–342. https://doi.org/10.1001/archinte.167.4.335

Behr, M. A., Edelstein, P. H., & Ramakrishnan, L. (2018). Revisiting the timetable of tuberculosis. *Bmj*, *362*, k2738.

Blower, S. M., McLean, A. R., Porco, T. C., Small, P. M., Hopewell, P. C., Sanchez, M. A., & Moss, A. R. (1995). The intrinsic transmission dynamics of tuberculosis epidemics. *Nature Medicine*, *1*(8), 815–821. https://doi.org/10.1038/nm0895-815

Borgdorff, M. W., Nagelkerke, N. J., Dye, C., & Nunn, P. (2000). Gender and tuberculosis: a comparison of prevalence surveys with notification data to explore sex differences in case detection. *Int. J. Tuberc. Lung Dis.*, *4*(2), 123–132.

Clayton, J. A. (2016). Studying both sexes: a guiding principle for biomedicine. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *30*(2), 519–524. https://doi.org/10.1096/fj.15-279554

Dodd, P. J., Looker, C., Plumb, I. D., Bond, V., Schaap, A., Shanaube, K., Muyoyeta, M., Vynnycky, E., Godfrey-Faussett, P., Corbett, E. L., Beyers,

N., Ayles, H., & White, R. G. (2016). Age- and Sex-Specific Social Contact Patterns and Incidence of Mycobacterium tuberculosis Infection. *American Journal of Epidemiology*, *183*(2), 156–166. https://doi.org/10.1093/aje/kwv160

Drain, P. K., Bajema, K. L., Dowdy, D., Dheda, K., Naidoo, K., Schumacher, S. G., Ma, S., Meermeier, E., Lewinsohn, D. M., & Sherman, D. R. (2018). Incipient and Subclinical Tuberculosis: a Clinical Review of Early Stages and Progression of Infection. *Clinical Microbiology Reviews*, *31*(4), e00021-18. https://doi.org/10.1128/cmr.00021-18

Eubank, S., Guclu, H., Kumar, V., Marathe, M. V., Srinivasan, A., Toroczkai, Z., & Wang, N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, *429*(6988), 180–184. https://doi.org/10.1038/nature02541

Guerra-Silveira, F., & Abad-Franch, F. (2013). Sex bias in infectious disease epidemiology: patterns and processes. *PLoS ONE*, *8*(4), e62390. https://doi.org/10.1371/journal.pone.0062390

Guwatudde, D., Nakakeeto, M., Jones-Lopez, E. C., Maganda, A., Chiunda, A., Mugerwa, R. D., Ellner, J. J., Bukenya, G., & Whalen, C. C. (2003). Tuberculosis in Household Contacts of Infectious Cases in Kampala, Uganda. *American Journal of Epidemiology*, *158*(9), 887–898. https://doi.org/10.1093/aje/kwg227

Hector, J., Anderson, S. T., Banda, G., Kamdolozi, M., Jefferys, L. F., Shani, D., Garton, N. J., Mwale, A., Jobe, A., Davies, G. R., & Sloan, D. J. (2017). TST positivity in household contacts of tuberculosis patients: a case-contact study in Malawi. *BMC Infectious Diseases*, *17*(1), 259. https://doi.org/10.1186/s12879-017-2348-2

Horton, K. C., Hoey, A. L., Béraud, G., Corbett, E. L., & White, R. G. (2020). Systematic Review and Meta-Analysis of Sex Differences in Social Contact Patterns and Implications for Tuberculosis Transmission and Control. *Emerging Infectious Diseases*, *26*(5), 910–919. https://doi.org/10.3201/eid2605.190574

Houben, R. M. G. J., & Dodd, P. J. (2016). The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling.

PLoS Med, *13*(10), e1002152.
https://doi.org/10.1371/journal.pmed.1002152

Islami, F., Torre, L. A., & Jemal, A. (2015). Global trends of lung cancer mortality and smoking prevalence. *Translational Lung Cancer Research*, *4*(4), 327–338. https://doi.org/10.3978/j.issn.2218-6751.2015.08.04

Kiss, I. Z., Miller, J. C., & Simon, P. L. (2017). *Mathematics of epidemics on networks* (Vol. 46). Springer.

Lönnroth, K., Williams, B. G., Stadlin, S., Jaramillo, E., & Dye, C. (2008). Alcohol use as a risk factor for tuberculosis - a systematic review. *Bmc Public Health*, *8*(1), 289. https://doi.org/10.1186/1471-2458-8-289

Ma, Y., Horsburgh, C. R., White, L. F., & Jenkins, H. E. (2018). Quantifying TB transmission: a systematic review of reproduction number and serial interval estimates for tuberculosis. *Epidemiology and Infection*, *146*(12), 1478–1494. https://doi.org/10.1017/s0950268818001760

Meintjes, G., Schoeman, H., Morroni, C., Wilson, D., & Maartens, G. (2008). Patient and provider delay in tuberculosis suspects from communities with a high HIV prevalence in South Africa: a cross-sectional study. *BMC Infectious Diseases*, *8*(1), 72. https://doi.org/10.1186/1471-2334-8-72

Miller, J. C., & Ting, T. (2020). EoN (Epidemics on Networks): a fast, flexible Python package for simulation, analytic approximation, and analysis of epidemics on networks. *Journal of Open Source Software*, *4*(44), 1731. https://doi.org/10.21105/joss.01731

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., & Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*, *5*(3), e74. https://doi.org/10.1371/journal.pmed.0050074

Nadini, M., Sun, K., Ubaldi, E., Starnini, M., Rizzo, A., & Perra, N. (2018). Epidemic spreading in modular time-varying networks. *Scientific Reports*, *8*(1), 2352. https://doi.org/10.1038/s41598-018-20908-x

Newman, M. (2003). Mixing patterns in networks. *Physical Review E*, *67*(2). https://doi.org/10.1103/physreve.67.026126

Neyrolles, O., & Quintana-Murci, L. (2009). Sexual Inequality in Tuberculosis. *PLoS Med*, *6*(12), e1000199. https://doi.org/10.1371/journal.pmed.1000199

Nahid, P., Dorman, S. E., Alipanah, N., Barry, P. M., Brozek, J. L., Cattamanchi, A., ... & Higashi, J. M. (2016). Official American thoracic society/centers for disease control and prevention/infectious diseases society of America clinical practice guidelines: treatment of drug-susceptible tuberculosis. *Clinical Infectious Diseases*, *63*(7), e147-e195.

Nhamoyebonde, S., & Leslie, A. (2014). Biological Differences Between the Sexes and Susceptibility to Tuberculosis. *Journal of Infectious Diseases*, *209*(suppl 3), S100–S106. https://doi.org/10.1093/infdis/jiu147

Organization, W. H. (2007). *Addressing sex and gender in epidemic-prone infectious diseases*.

Perkins, S. E., Ferrari, M. F., & Hudson, P. J. (2008). The effects of social structure and sex-biased transmission on macroparasite infection. *Parasitology*, *135*(13), 1561–1569. https://doi.org/10.1017/s0031182008000449

Prevention, C. for D. C. and. (2010). *… of the Surgeon General. How tobacco smoke causes disease: the biology and behavioral basis for smoking-attributable disease: a report of the Surgeon …*.

Rhines, A. S. (2013). The role of sex differences in the prevalence and transmission of tuberculosis. *Tuberculosis*, *93*(1), 104–107. https://doi.org/10.1016/j.tube.2012.10.012

Rohani, P., Zhong, X., & King, A. A. (2010). Contact Network Structure Explains the Changing Epidemiology of Pertussis. *Science*, *330*(6006), 982–985. https://doi.org/10.1126/science.1194134

Sah, P., Leu, S. T., Cross, P. C., Hudson, P. J., & Bansal, S. (2017). Unraveling the disease consequences and mechanisms of modular structure in animal social networks. *Proceedings of the National Academy of Sciences*

of the United States of America, *114*(16), 4165–4170.
https://doi.org/10.1073/pnas.1613616114

Sah, P., Singh, L. O., Clauset, A., & Bansal, S. (2014). Exploring community structure in biological networks with random graphs. *BMC Bioinformatics*, *15*(220). https://doi.org/10.1186/1471-2105-15-220

Salathé, M., & Jones, J. H. (2010). Dynamics and Control of Diseases in Networks with Community Structure. *PLoS Computational Biology*, *6*(4), e1000736. https://doi.org/10.1371/journal.pcbi.1000736

Salim, M., Declercq, E., Deun, A. V., & Saki, K. (2004). Gender differences in tuberculosis: a prevalence survey done in Bangladesh. *Int. J. Tuberc. Lung Dis.*, *8*(8), 952–957.

Schurz, H., Salie, M., Tromp, G., Hoal, E. G., Kinnear, C. J., & Möller, M. (2019). The X chromosome and sex-specific effects in infectious disease susceptibility. *Human Genomics*, *13*(1), 1–12. https://doi.org/10.1186/s40246-018-0185-z

Sendagire, I., Van der Loeff, M. S., Mubiru, M., Konde-Lule, J., & Cobelens, F. (2010). Long delays and missed opportunities in diagnosing smear-positive pulmonary tuberculosis in Kampala, Uganda: a cross-sectional study. PLOS one, 5(12), e14459.

Waroux, O. le P. de, Cohuet, S., Ndazima, D., Kucharski, A. J., Juan-Giner, A., Flasche, S., Tumwesigye, E., Arinaitwe, R., Mwanga-Amumpaire, J., Boum, Y., Nackers, F., Checchi, F., Grais, R. F., & Edmunds, W. J. (2018). Characteristics of human encounters and social mixing patterns relevant to infectious diseases spread by close contact: a survey in Southwest Uganda. *BMC Infectious Diseases*, *18*(1), 172. https://doi.org/10.1186/s12879-018-3073-1

Watkins, R. E., & Plant, A. J. (2006). Does smoking explain sex differences in the global tuberculosis epidemic? *Epidemiology and Infection*, *134*(2), 333–339. https://doi.org/10.1017/s0950268805005042

WHO. (2018). *Global tuberculosis report*. World Health Organization. https://www.who.int/tb/publications/global_report/en/

Xu, Y., Cancino-Muñoz, I., Torres-Puente, M., Villamayor, L. M., Borrás, R., Borrás-Máñez, M., Bosque, M., Camarena, J. J., Colomer-Roig, E., Colomina, J., Escribano, I., Esparcia-Rodríguez, O., Gil-Brusola, A., Gimeno, C., Gimeno-Gascón, A., Gomila-Sard, B., González-Granda, D., Gonzalo-Jiménez, N., Guna-Serrano, M. R., … Comas, I. (2019). High-resolution mapping of tuberculosis transmission: Whole genome sequencing and phylogenetic modelling of a cohort from Valencia Region, Spain. *PLoS Med*, *16*(10), e1002961. https://doi.org/10.1371/journal.pmed.1002961

Table 3.1 Transitions, parameters, and values for assortativity model.

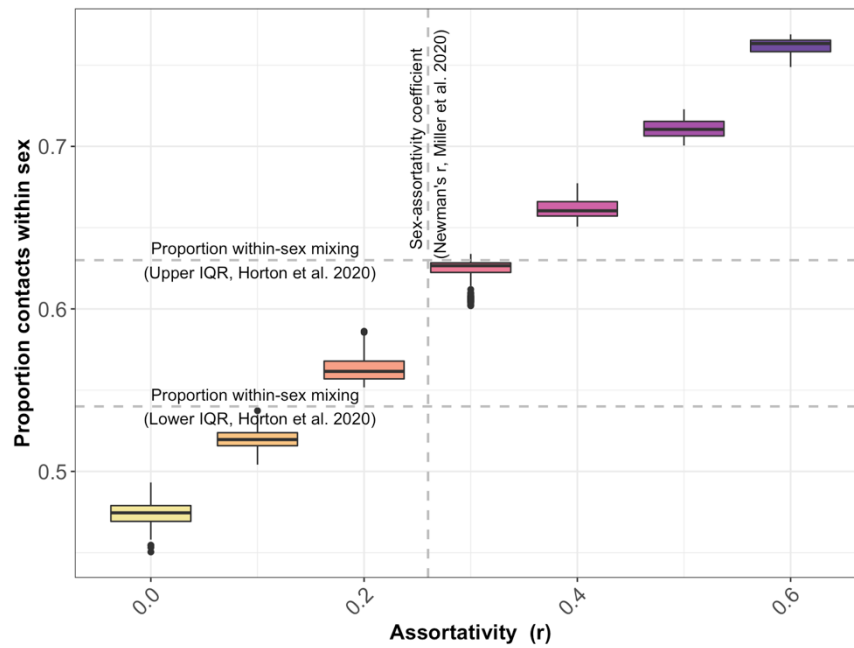| Transition	Definition	Parameter	Average values
$S \rightarrow L$	Transmission rate	τ	$0.04, 0.075, 0.1$
$L \rightarrow I$	Incubation rate	σ	$0.1, \infty$
$I \rightarrow R$	Recovery rate	γ	0.5
$R \rightarrow S$	Reversion rate	ϕ	$1 \cdot 10^{-5}, 0.1$

Figure 3.1 Relationship between assortativity and within-sex contacts.

The proportion of contacts within-sex tracks with Newman's r. For context, the range of proportional within-sex mixing from a meta-analysis by Horton et al. (2020) is shown in horizontal grey-dashed lines and the sex-assortativity coefficient from a social network in Uganda in a horizontal dashed line. Boxplots show calculated proportional within-sex mixing from 250 simulated networks at each assortativity level.
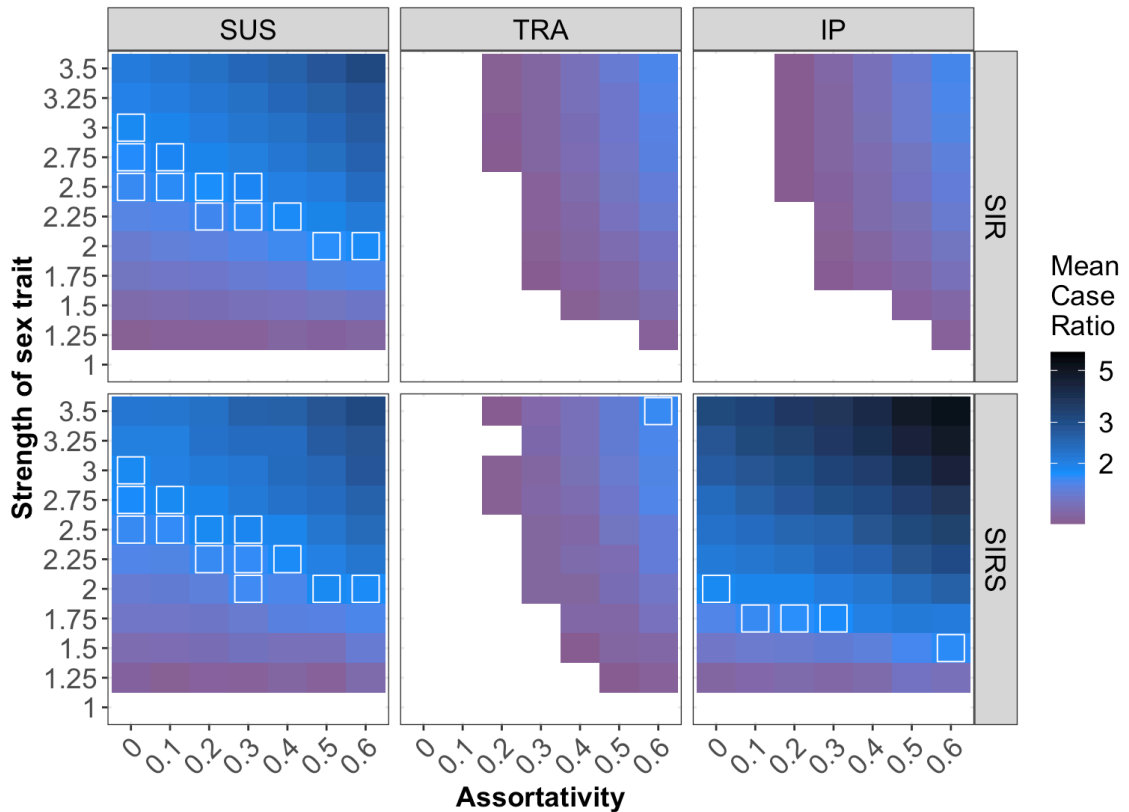
Figure 3.2 Relationship between assortativity, sex-traits, and male-bias.

M:F case ratio is influenced by sex-assortativity and sex-traits (columns). Sex-traits are susceptibility (SUS), transmissibility (TRA), and infectious period (IP). M:F case ratio is measured as the ratio of male to female recoveries (SIR) or infections at equilibrium (SIRS). Only parameter combinations leading to mean M:F case bias greater than 1.1 are colored and the white boxes show parameter combinations leading to a mean M:F case bias from 1.7 to 1.9. Sex-traits are incorporated by holding respective overall parameter rates constant but increasing the male parameter by the value on the y-axis relative to the female trait. Figure generated with 250 simulations of epidemics on Sah networks with $\tau = 0.04, R_0 \approx 1.5$.
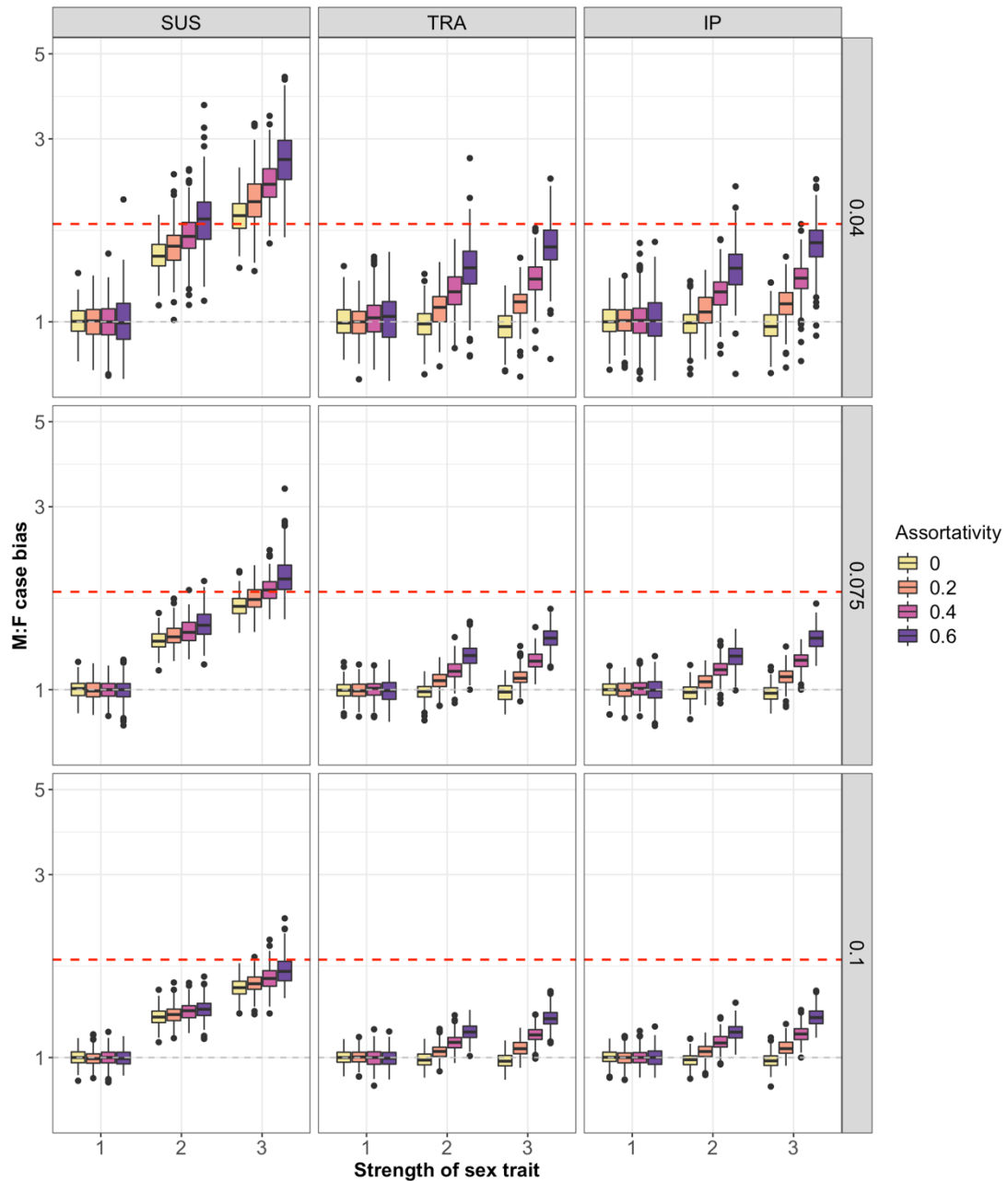
Figure 3.3 Relationship between assortativity, sex-traits, transmission rate, and male-bias.

Sex-assortativity increases M:F case bias more for pathogens with lower overall infection rates compared with higher overall infection rates (rows). The M:F case bias and sex-traits, are defined in the same way as in Figure 1. To improve figure clarity, only 3 levels of sex-trait strength are shown here. Figure generated with 250 SLIR simulations of epidemics on Sah networks.
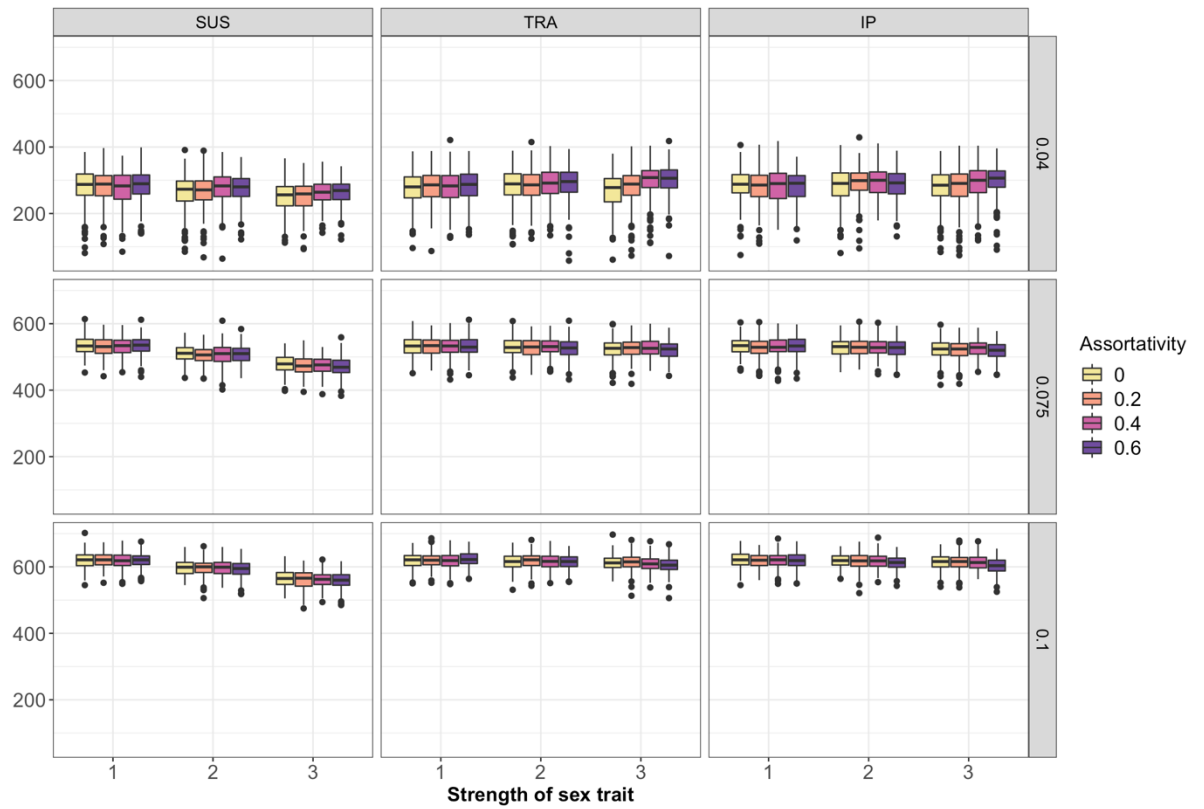
Figure 3.4 Effects of assortativity on epidemic dynamics.

Sex-assortativity and sex-traits (columns) generally have negligible effects on the final epidemic size compared with the effect of transmission rate (rows). To improve figure clarity, only 3 levels of sex-trait strength are shown here. Figure generated with 250 SLIR simulations of epidemics with varying transmission rates on Sah networks.

Chapter 4

THE EFFECTS OF CORE-PERIPHERY NETWORK STRUCTURE ON

DISEASE SPREAD IN ISOLATED POPULATIONS[3]

Abstract

Contact network epidemiology characterizes the effects of network structure on disease spread. Recent outbreaks of COVID-19 in nursing homes and prisons, where contact structures are far from homogenous, motivated us to explore the effect of core-periphery network structure on the outbreak size distribution. Core-periphery networks are distinguished by a densely-connected core and sparsely-connected periphery. We show that the outbreak size distribution transitions from a unimodal distribution centered at the origin to a bimodal distribution at lower transmission rates in core-periphery compared with homogenous networks. However, at higher transmission rates, core-periphery network structure decreases outbreak size. We also showed how peak size, peak time, and outbreak duration were affected in various ways by core-periphery network structure where the directions and magnitudes of changes depended on the transmission rate and the size of the core and periphery groups. Lastly, we explored how estimates of R_0, the basic reproductive number, may be affected by core-periphery contact structure and found that estimators that were naive of the underlying contact network structure underestimated (final size estimator) and overestimated (proportion extinct) R_0, with the true value in the middle. Understanding how epidemics are affected by core-periphery network structure could lead to improved interventions that use contact structure to slow spread or more accurate estimates of R_0 in structured populations.

Introduction

Contact networks, describing the organization of interactions among individuals in a population, have drastic effects on epidemic outcomes and the efficacy of interventions (Barclay et al., 2014; Cui et al., 2019; Ma et al., 2013; Mossong et al., 2008; Pastor-Satorras et al., 2015; Pastor-Satorras & Vespignani, 2001; Shirley & Rushton, 2005). Well-known examples that affect epidemics include variance in the degree distribution, clustering, and modularity. Large variation in the degree distribution, such as in scale-free networks (Barabasi & Albert, 1999), can affect whether epidemics are possible (Ma et al., 2013; Pastor-Satorras & Vespignani, 2001; Shirley & Rushton, 2005). Clustering, a distinguishing feature of small-world networks (Watts & Strogatz, 1998) generally decreases the epidemic final size (Volz et al., 2011). Modularity, or the propensity for nodes to mix with those in the same group (Newman, 2003) can affect when the outbreak peaks and ends (Lentz et al., 2012; Ma et al., 2013). The efficacy of interventions, such as vaccination herd immunity (Barclay et al., 2014; Ma et al., 2013), can also be influenced by network structure. For example, in scale-free networks, natural immunization confers higher herd immunity than random vaccination, for the same proportions vaccinated while the opposite is true for small-world networks (Ferrari et al., 2006). Network structure is sometimes more important for determining epidemic outcomes than interventions (Ma et al., 2013), underscoring the need to study various types of contact structure and disease containability.

A lesser-studied structure in network epidemiology is that of core-periphery (CP) networks which are defined by densely-connected cores and sparsely-connected periphery nodes (Borgatti & Everett, 2000; Csermely et al., 2013). Core-periphery networks are one of four theoretically defined meso-scale network types, which also include random, modular, and networks with a hole (Estrada, 2006). Core-periphery networks differ from modular and degree assortative networks because periphery nodes mix with core nodes in a non-group assortative, degree disassortative fashion. Interestingly, networks with degree distributions defined by a power law, $P(k) = 1/k^\beta$, have CP-like structure when $2 < \beta < 3$ (Chung & Lü, 2002), suggesting there could be similarity in epidemic outcomes, such as the outbreak size distribution, between power-law and CP networks. Others have classified outbreak size distributions as having one of two shapes, "J-shaped" and "U-shaped" (Kendall 1956; Nassell 1995). The term J-shape refers to a distribution of outbreak sizes that is monotonically decreasing with a mode at the origin while the term U-shape refers to a bimodal distribution. If CP networks have a diminished epidemic threshold, as power-law networks do, we may expect CP networks to have a U-shaped outbreak size distribution at lower transmission rates than homogenous networks. While others have shown the individual-level of infection is higher among core nodes in core-periphery like networks (Christley et al., 2005), there remains a gap in the network epidemiology literature about the effects of core-periphery structure on population-level epidemic dynamics.

We were motivated to study CP networks due to the spread of COVID-19 in contact networks with potentially similar structure. For example, on board the Diamond Princess Cruise Ship during its 3 week quarantine, crew workers were largely unconfined and passengers rarely left their cabins (Canadian Broadcasting Corporation, 2020). This outbreak resulted in nearly 700 passengers being infected, 9 deaths, and an estimated initial R_0 of 9.3 (Emery et al., 2020). It's still unclear whether passenger isolation prevented infections but highlights the need to understand outbreaks in structured, isolated populations. Nursing homes and prisons, which account for a significant proportion of cases and deaths in the United States, are instating similar isolation policies for residents during lockdown (The News and Observer, 2020; National Public Radio, 2020), who then form a sort of periphery. In both nursing homes and prisons, staff maintain their duties and may serve as the only daily interactions for residents. Together, we hypothesized, staff and resident interactions form a sort of CP contact network in these settings which have been hotspots for COVID-19.

The goal of this study was to understand how core-periphery contact networks influence disease spread. To this end, we simulated contact networks divided into densely-connected core nodes and sparsely-connected periphery nodes, described unique properties of this network structure, and compared epidemic dynamics on CP networks with homogeneous networks. Specifically, we examined the probability distributions of outbreak sizes, the peak size, peak

timing, and outbreak duration for outbreaks in core-periphery and homogenous

networks. In addition, we characterized how estimates of R_0 may be affected by

CP network structure to guide future intervention planning in these settings, such

as the vaccination threshold for herd-immunity. These results may additionally be

applicable to designing interventions that alter population contact structure such

as creation of "social bubbles" (Block et al., 2020) within workplaces.

Methods

Network simulations

In simulated contact networks, nodes represent individuals and edges

between them represent repeated interactions between nodes on which infection

can spread. To simulate core-periphery (CP) networks we modified the algorithm

presented in (Rombach et al., 2014). The original algorithm allowed for flexibility

in the size of the core and variability in the "coreness" of the core (i.e., how much

more connected it is relative to the periphery) but did not hold the network mean

degree constant as the coreness increased, a problem when evaluating the

effects of network structure on epidemics in isolation from network connectivity.

The original algorithm is defined as follows: $G(N, d, p, k)$ where N is network size,

dN are core (C) nodes and $(1 - dN)$ are periphery (P) nodes. Edges are

assigned independently with probabilities depending on whether the edge would

be C--C, P--P, or C--P. The edge probabilities of each type are given by $p \cdot k^2$ for

C-C edges and $p \cdot k$ for other edge types. In this original formulation, as k

increases, the number of edges and mean degree increases. In our modified

algorithm, the edge probabilities of each type are given by $p \cdot k$ for core-core

edges and p/k for other edge types. With this modified formulation,

homogeneous networks are given by $k = 1$ and CP networks have $k > 1$ (Figure

4.1). To understand disease spread in different populations, we varied the

network size ($N = 50, 100, 500, 1000, 5000$), the relative size of the core

population ($d = 0.25, 0.5, 0.75$) and the coreness ($k = 1, 2, 3, 4$).

To contextualize core-periphery networks within the literature about

network structure, we estimated the degree assortativity, group assortativity,

clustering, degree variation and relative degree of the core to the periphery.

Degree assortativity measures correlation in node degree among nodes that

share an edge (Newman, 2003). Group assortativity measures how often nodes

within the same group (i.e., core or periphery) interact (Newman, 2003).

Clustering is the proportion of triangles in a network (Wasserman & Faust, 1994).

Degree variation measures the variance in node degree of the network and the

relative degree of the core to the periphery is the mean degree of all core nodes

divided by the mean degree of periphery nodes. These statistics aid our

interpretation of any differences in epidemic dynamics found with changes in

coreness (k) and allow us to distinguish core-periphery networks from other

network types.

Disease model

To study disease spread in core-periphery networks, we used a

Susceptible-Latent-Infectious-Recovered (SLIR) model with infection transmitting

along an edge at rate τ, latent infections becoming infectious at rate σ ($1/\sigma$ is the incubation period), and recovery happening at rate γ ($1/\gamma$ is the infectious period). We assumed an incubation and infectious period of 6 days (Backer et al. 2020) and varied τ ($0.02, 0.04, 0.075, 0.01$) which we found resulted in a range of R_0 values from approximately 1 to 4 in homogeneously-mixing networks (calculations given below). For each simulation, we recorded the final outbreak size, peak size, peak timing, and outbreak duration. For each set of simulations, we recorded the proportion that faded out (resulted in 3 or fewer infections).

To understand how the ability to estimate R_0 is affected by core-periphery network structure, we compared three estimators of R_0 to the mean number of cases caused by the index case, $\bar{R}_\iota$ (Lloyd-Smith et al., 2005) which we considered to be the true value. First, we estimated R_0 assuming we had knowledge of the underlying contact structure with the formula $\frac{\tau <K^2 - K>}{\tau + \gamma <K>}$ where $< K >$ is the mean, $< K^2 - K >$ is the second moment (variance) of the degree distribution and τ, γ are the transmission and recovery rate (Kiss et al., 2017). However, because we are often naive of the underlying contact structure when planning interventions for different populations, we compared two additional estimators, one based on the final size of outbreaks (Ma & Earn, 2006) and one based on the proportion of outbreaks that went extinct (Bailey, 1975; Renshaw, 1991), which are both naive of the population contact structure. The formula for the final size estimator is $Z = 1 - e^{-R_0 Z}$, which can be solved numerically by

finding the root of $0 = 1 - e^{-R_0 Z} - Z$. Finally, the formula for estimating R_0 based on the proportion of outbreaks that take off is $\mathrm{R}_0 = \frac{1}{1-p}$ where p is the proportion that take off (operationally defined here as outbreaks that had more than 3 infections). To quantify the bias of each estimator $\widehat{R_0}$, we calculated $(\widehat{R_0} - \bar{R}_\iota)/\bar{R}_\iota$. We considered estimators to be underestimates of $\bar{R}_\iota$ if the value was less than 0 and overestimates if the value was greater than 0.

Each simulation was initialized by one infected individual selected at random from among all the individuals in the population and run until there were zero infected individuals remaining. We replicated each disease and network parameter combination 500 times. We implemented the model using a continuous-time, Gillespie algorithm with the Epidemics on Networks module (Miller & Ting, 2020) in Python Version 2. We used R (Team, 2020) Version 4.0.2 for data analysis and visualization. Scripts to fully reproduce results are available at https://github.com/CEIDatUGA/core-periphery-networks and description of the Gillespie algorithm used in this Chapter is given in Appendix II.

Results

Epidemics on Core-Periphery Networks

Relative to homogeneous networks, core-periphery networks had outbreaks with lower transmission rates (Figure 4.2) despite similar initial R_0, measured by the mean number of new infections caused by the index individual, $\bar{R}_\iota$ (Table 4.1).

The outbreak size distribution varied with core-periphery network structure (k) and the direction of change depended on the transmission rate, τ (Figure 4.3). In networks with 5000 nodes, when $\tau = 0.02$, the mean final size of outbreaks increased with k. For example, when $\tau = 0.02$, the mean final size of outbreaks increased from 1% to 19% when k increased from 1 to 4. In contrast, when $\tau > 0.02$, the final size of outbreaks decreased with k.

The final size of outbreaks (i.e., percentage of the population that got infected) also varied with the relative size of the core compared with the periphery (d) (Figure 4.4). Smaller, denser cores ($d = 0.25$) led to larger outbreaks than larger, more diffuse cores ($d = 0.75$) when $\tau = 0.02$. However, when $\tau = 0.1$, the mean final outbreak size was smaller when the relative sizes of core and periphery groups were unequal (i.e., when $d = 0.25$ or $d = 0.75$). The effects of d on the final size of outbreaks were similar across network sizes (Appendix III, Figure S3.1).

Besides the outbreak size distribution, core-periphery networks led to epidemics with smaller and earlier peaks compared with homogeneous networks except when transmission rates were low (Appendix III, Figure S3.2). In addition, epidemics on core-periphery networks were generally shorter in duration than on homogenous networks. For example, when $N = 5,000$ and $\tau = 0.1$, outbreak duration was 77 days ($k = 4$) compared with 95 days in homogenous networks. The opposite was true when transmission rates were low: when $\tau = 0.02$, core-periphery networks had outbreaks that lasted nearly 80% longer than

homogenous networks. The effects of k on the peak size, peak time, and outbreak duration were similar across network sizes (Appendix III, Figure S3.3). The size of the core (d) also affected trends in peak size and outbreak duration (Appendix III, Figure S3.4). With low transmission rates, $\tau = 0.02$, networks with smaller cores ($d = 0.25$) had outbreaks with larger peaks and epidemics that were longer lasting compared with outbreaks on networks with larger cores ($d = 0.75$).

Estimating R_0 in Core-Periphery Networks

With equal sized core and periphery groups, R_0 of outbreaks taking place on homogenous networks ($k = 1$) was generally well-estimated by the excess network degree variation estimator and the final size estimator (Figure 4.5, 4.6). The proportion extinct estimator tended to overestimate R_0 even in homogeneous networks. In contrast, estimates of R_0 based on the proportion extinct estimator were always biased high and did not increase monotonically with k. The estimator based on excess degree variation tended to overestimate R_0 and the final size estimator tended to underestimate R_0 as k increased (Figure 4.5, Figure 4.6). These patterns remained when core and periphery groups were different sizes.

Core-Periphery Network Structure

Simulated networks had stable network mean degrees, were connected (i.e., there were no isolated nodes), and had visual structural changes in how core and periphery were connected with increasing coreness, k. Among

networks with 5,000 nodes and evenly sized core and periphery groups ($d = 0.5$),

degree and group assortativity coefficients were approximately zero across

different values of k, indicating that mixing among nodes with different degrees

and nodes in different groups (i.e., core, periphery) was just as likely to mix as

nodes with similar degrees and in the same group (Appendix III, Figure S3.5 A,

B). As k increased, changes in the clustering coefficient were minimal, increasing

from a mean of 0.002 when $k = 1$ to a mean of 0.0035 when $k = 4$ (Appendix III,

Figure S3.5 C). The mean degree variance (Appendix III, Figure S3.5 D) and the

relative mean degree of core to periphery nodes (Appendix III, Figure S3.5 E)

both increased considerably as k increased.

Discussion

We explored the effects of core-periphery network structure on epidemics.

Our main results are that epidemics are possible in core-periphery networks at a

lower transmission rate than in homogeneous networks but as the transmission

rate increases, core-periphery network structure limits the epidemic size relative

to homogenous networks. We additionally showed that the peak size, peak time,

and outbreak duration were affected in different ways by core-periphery network

depending on the transmission rate and the size of the core and periphery

groups. Lastly, we explored how estimates of R_0 may be affected by core-

periphery contact structure and found that some estimators underestimated (final

size estimator, excess degree) and overestimated (proportion extinct) R_0, with

the true value in the middle. Understanding how epidemic outcomes are affected

by core-periphery network structure could inform interventions that use contact structure to slow spread and lead to more accurate estimates of R_0 in structured populations.

This study contributes to the broad body of work exploring how contact patterns can affect epidemic outcomes (Bansal et al., 2007; Keeling & Rohani, 2007; Meyers et al., 2005; Rohani et al., 2010). Our main findings involved the probability distribution of outbreak sizes in core-periphery relative to homogenous networks. In stochastic models of disease transmission, such as the one used here, the epidemic threshold occurs when the outbreak size distribution transitions from unimodal (i.e., "J-shaped") to bimodal (i.e., "U-shaped") (Kendall 1956; Nassell 1995). We showed that the transition from J-shape to U-shape occurs at a lower transmission rate in core-periphery networks relative to homogeneous networks. However, when $R_0 \geq 1.5$ core-periphery networks were associated with slightly smaller outbreaks. These findings can be understood in light of differences in degree variation. When networks have core groups, especially when they are small, degree variation increases and as degree variation approaches infinity, outbreaks of pathogens with low transmission rates become possible (Pastor-Satorras & Vespignani, 2001; Shirley & Rushton, 2005). When an infected individual is introduced into small, densely connected core groups, pathogens with lower transmission rates may initially persist in the population until spread becomes limited by the lack of connections to the periphery (i.e., "core-trapping"), as happens much faster for pathogens with

higher transmission rates. This is consistent with our findings about peak size and peak time: pathogens with lower transmission rates have delayed and larger peaks in addition to longer-lasting outbreaks in core-periphery networks compared with homogeneous networks suggesting temporary persistence within the dense core. Thus, our findings suggest core-periphery networks may facilitate initial disease spread by allowing temporary persistence in the core when epidemic dynamics are near the epidemic threshold but when $R_0 \gg 1$, core-periphery networks inhibit disease spread due to there being reduced chances for spread to periphery nodes (i.e., "core-trapping").

Due to these differences in the outbreak size distribution near the epidemic threshold in core-periphery networks and their effects on the final size of outbreaks, we showed how estimations of R_0 can be biased. Specifically, because core-periphery structure generally limits the size of outbreaks (except when $R_0 \approx 1$) the final size estimator was biased low and because the network invasibility increases in networks with large degree variation, the estimator based on the proportion of outbreaks that go extinct was biased high. Similar to the final size estimator, the excess degree estimator (Meyers et al., 2005) also tended to underestimate R_0 in core-periphery networks likely because the variance to mean degree ratio (i.e., $\frac{<k^2-k>}{<k>}$) increased while the final size decreased (when $R_0 \geq$ 1.5). Overall, these results suggest caution in interpreting estimates of R_0 in structured populations but that comparing multiple estimations may improve estimations and interventions based on R_0, such as the vaccine threshold.

While core-periphery networks have been less-well studied in the network epidemiology literature, they have been described elsewhere (Csermely et al., 2013; Holme, 2005; Rombach et al., 2014) and this study further characterizes CP structure for use in network modeling of infectious diseases. First, we modified an existing algorithm (Rombach et al., 2014) so that we could separate the effects of CP-structure on epidemic outcomes from changes in edge density and clustering. We then characterized networks generated by this algorithm, and in line with previous descriptions (Csermely et al., 2013), showed that core-periphery networks are distinguished from modular and degree-assorted networks because periphery nodes mix with core nodes in a non-group, degree-disassortative fashion, which we quantified here. We additionally showed that CP networks have near-zero clustering coefficients, which can decrease the final epidemic size (Volz et al., 2011), meaning this is not likely the reason that outbreaks are smaller in core-periphery networks when $R_0 > 1.5$. In contrast, both degree variance and mean degree of core relative to periphery nodes increased as coreness increased, which we believe, in addition to the lack of change in assortativity and clustering, are driving increased invasibility of CP networks near the epidemic threshold. The modified Rombach algorithm presented in this study is efficient for the networks at least up to 5,000 nodes and is suitable for future research in network epidemiology and a variety of other fields.

Core-periphery contact structure may be relevant to disease spread in settings that have exhibited COVID-19 outbreaks including cruise ships, nursing homes (Dool et al., 2008), and correctional facilities where passengers, residents, and incarcerated individuals may form a sort of periphery to the facility staff. The R_0 of COVID-19 is generally thought to be larger than 1.5 (e.g., Kucharski et al., 2020) and as a result, in these settings with frequent COVID-19 outbreaks, core-periphery contact structure may limit the final size of outbreaks. Additional decreases in core-core (i.e., staff-staff) interactions or creation of staff-resident "bubbles" (Block et al., 2020), may further reduce spread and these changes would only rely on altering the contact structure. In the event that a vaccine becomes available, contact structure within these settings will likely affect the optimal distribution strategy (Thedchanamoorthy et al., 2014). In nursing homes, vaccinating staff is an efficient way to reduce the spread of influenza among residents (Dool et al., 2008) and this is likely due to a core-periphery contact structure among staff and residents. This study highlights the advantages to understanding a population's contact structure and future research should seek to identify optimal interventions in core-periphery networks and test predictions about the final size with real data.

These simulations focused on the relatively idealized system of closed populations with static contact networks. Future work could aim to extend these findings to more realistic, dynamic, semi-closed populations (e.g., residents and staff in nursing homes) and to link ideas presented here with concrete

96

intervention applications (e.g., reduction of staff-staff interactions and creation of workplace "social bubbles" (Block et al., 2020). In addition, simulations assumed there was an equal chance of initial infection in core and periphery groups, when in reality, staff in nursing homes or prisons may be more likely to import infections than residents or incarcerated individuals. Finally, simulations assumed equal susceptibility, transmissibility, and infectious periods in core and periphery individuals as well as over the course of the epidemic. For many infectious diseases, age, sex, and interventions may affect these rates and interact with contact structure, like core-periphery networks, to further alter epidemic dynamics, a promising topic for future work.

Conclusions

Epidemic outcomes and estimation of key parameters are tied to populations' underlying contact networks. When contact patterns are structured into densely-connected core and less well-connected periphery groups, epidemics are possible at a lower transmission rate relative to homogenous networks but as R_0 increases, the final size of epidemics is smaller in core-periphery relative to homogenous networks, perhaps be due to a "core-trapping" mechanism. Overall, these findings contribute to the broad literature describing the effects on contact structure on epidemics and may shed light on the dynamics of COVID-19 in structured populations currently experiencing large outbreaks such as nursing homes and prisons.

Author Contributions

PM designed and implemented simulation study, analyzed data, wrote manuscript; JD designed simulation study and edited manuscript.

References

Backer, J. A., Klinkenberg, D., & Wallinga, J. (2020). Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance*, *25*(5), 2000062.

Bailey, N. (1975). *The mathematical theory of infectious diseases and its applications*. Hafner Press.

Bansal, S., Grenfell, B. T., & Meyers, L. A. (2007). When individual behaviour matters: homogeneous and network models in epidemiology. *Journal of The Royal Society Interface*, *4*(16), 879–891. https://doi.org/10.1098/rsif.2007.1100

Barabasi, A., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.

Barclay, V. C., Smieszek, T., He, J., Cao, G., Rainey, J. J., Gao, H., Uzicanin, A., & Salathé, M. (2014). Positive network assortativity of influenza vaccination at a high school: implications for outbreak risk and herd immunity. *PLoS ONE*, *9*(2), e87042. https://doi.org/10.1371/journal.pone.0087042

Block, P., Hoffman, M., Raabe, I. J., Dowd, J. B., Rahal, C., Kashyap, R., & Mills, M. C. (2020). Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nature Human Behavior*, 1–9. https://doi.org/10.1038/s41562-020-0898-6

Borgatti, S. P., & Everett, M. G. (2000). Models of core/periphery structures. *Social Networks*, *21*(4), 375–395. https://doi.org/10.1016/s0378-8733(99)00019-2

Christley, R. M., Pinchbeck, G. L., Bowers, R. G., Clancy, D., French, N. P., Bennett, R., & Turner, J. (2005). Infection in social networks: using network analysis to identify high-risk individuals. *American Journal of Epidemiology*, *162*(10), 1024–1031. https://doi.org/10.1093/aje/kwi308

Chung, F., & Lü, L. (2002). The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, *99*(25), 15879–15882. https://doi.org/10.1073/pnas.252631999

Csermely, P., London, A., Wu, L.-Y., & Uzzi, B. (2013). Structure and dynamics of core/periphery networks. *Journal of Complex Networks*, *1*(2), 93–123. https://doi.org/10.1093/comnet/cnt016

Cui, J., Zhang, Y., & Feng, Z. (2019). Influence of non-homogeneous mixing on final epidemic size in a meta-population model. *J. Biol. Dyn.*, *13*(sup1), 31–46. https://doi.org/10.1080/17513758.2018.1484186

Dool, C. van den, Bonten, M. J. M., Hak, E., Heijne, J. C. M., & Wallinga, J. (2008). The Effects of Influenza Vaccination of Health Care Workers in Nursing Homes: Insights from a Mathematical Model. *PLoS Medicine*, *5*(10), e200. https://doi.org/10.1371/journal.pmed.0050200

Emery, J. C., Russell, T. W., Liu, Y., Hellewell, J., Pearson, C. A., Group, C. C.-19 W., Atkins, K. E., Klepac, P., Endo, A., Jarvis, C. I., Davies, N. G., Rees, E. M., Meakin, S. R., Rosello, A., Zandvoort, K. van, Munday, J. D., Edmunds, W. J., Jombart, T., Auzenbergs, M., … Houben, R. M. (2020). The contribution of asymptomatic SARS-CoV-2 infections to transmission on the Diamond Princess cruise ship. *Elife*, *9*, 2081. https://doi.org/10.7554/elife.58699

Estrada, E. (2006). Spectral scaling and good expansion properties in complex networks. *Europhysics Letters*, *73*(4), 649–655. https://doi.org/10.1209/epl/i2005-10441-3

Ferrari, M. J., Bansal, S., Meyers, L. A., & Bjørnstad, O. N. (2006). Network frailty and the geometry of herd immunity. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1602), 2743–2748. https://doi.org/10.1098/rspb.2006.3636

Holme, P. (2005). Core-periphery organization of complex networks. *Physical Review E*, *72*(4), 046111. https://doi.org/10.1103/physreve.72.046111

Keeling, M. J., & Rohani, P. (2007). *Modeling Infectious Diseases in Humans and Animals*.

Kendall, D. G. (1956). Deterministic and stochastic epidemics in closed populations. *In Proc. 3rd Berkeley Symp. Math. Statist. Prob* (Vol. 4, pp. 149-165).

Kiss, I. Z., Miller, J. C., & Simon, P. L. (2017). *Mathematics of epidemics on networks* (Vol. 46). Springer.

Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., & Eggo, R. M. (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*, *20*(5), 553–558. https://doi.org/10.1016/s1473-3099(20)30144-4

Lentz, H. H. K., Selhorst, T., & Sokolov, I. M. (2012). Spread of infectious diseases in directed and modular metapopulation networks. *Physical Review E*, *85*(6), 066111. https://doi.org/10.1103/physreve.85.066111

Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., & Getz, W. M. (2005). Superspreading and the effect of individual variation on disease emergence. *Nature*, *438*(7066), 355–359. https://doi.org/10.1038/nature04153

Ma, J., Driessche, P. van den, & Willeboordse, F. H. (2013). The importance of contact network topology for the success of vaccination strategies. *Journal of Theoretical Biology*, *325*, 12–21. https://doi.org/10.1016/j.jtbi.2013.01.006

Ma, J., & Earn, D. J. D. (2006). Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bulletin of Mathematical Biology*, *68*(3), 679–702. https://doi.org/10.1007/s11538-005-9047-7

Meyers, L. A., Pourbohloul, B., Newman, M. E. J., Skowronski, D. M., & Brunham, R. C. (2005). Network theory and SARS: predicting outbreak diversity. *Journal of Theoretical Biology*, *232*(1), 71–81. https://doi.org/10.1016/j.jtbi.2004.07.026

Miller, J. C., & Ting, T. (2020). EoN (Epidemics on Networks): a fast, flexible Python package for simulation, analytic approximation, and analysis of epidemics on networks. *Journal of Open Source Software*, *4*(44), 1731. https://doi.org/10.21105/joss.01731

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., & Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*, *5*(3), e74. https://doi.org/10.1371/journal.pmed.0050074

Nåsell, I. (1995). The threshold concept in stochastic epidemic and endemic models. *Epidemic models: their structure and relation to data*, 5, 71.

Newman, M. (2003). Mixing patterns in networks. *Physical Review E*, *67*(2). https://doi.org/10.1103/physreve.67.026126

Pastor-Satorras, R., Castellano, C., Mieghem, P. V., & Vespignani, A. (2015). Epidemic processes in complex networks. *Reviews of Modern Physics*, *87*(3), 925–979. https://doi.org/10.1103/revmodphys.87.925

Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic dynamics and endemic states in complex networks. *Physical Review E*, *63*(6 Pt 2), 066117. https://doi.org/10.1103/physreve.63.066117

Renshaw, E. (1991). *Modelling Biological Populations in Space and Time*. Cambridge University Press.

Rohani, P., Zhong, X., & King, A. A. (2010). Contact Network Structure Explains the Changing Epidemiology of Pertussis. *Science*, *330*(6006), 982–985. https://doi.org/10.1126/science.1194134

Rombach, M. P., Porter, M. A., Fowler, J. H., & Mucha, P. J. (2014). Core-Periphery Structure in Networks. *SIAM Journal on Applied Mathematics*, *74*(1), 167–190. https://doi.org/10.1137/120881683

Sah, P., Singh, L. O., Clauset, A., & Bansal, S. (2014). Exploring community structure in biological networks with random graphs. *BMC Bioinformatics*, *15*(220). https://doi.org/10.1186/1471-2105-15-220

Shirley, M. D. F., & Rushton, S. P. (2005). The impacts of network topology on disease spread. *Ecological Complexity*, *2*(3), 287–299. https://doi.org/10.1016/j.ecocom.2005.04.005

Team, R. C. (2020). *R: A language and environment for statistical computing*. https://www.R-project.org/.

Thedchanamoorthy, G., Piraveenan, M., Uddin, S., & Senanayake, U. (2014). Influence of vaccination strategies and topology on the herd immunity of complex networks. *Social Network Analysis and Mining*, *4*(213), 1–16. https://doi.org/10.1007/s13278-014-0213-5.pdf

Volz, E. M., Miller, J. C., Galvani, A., & Meyers, L. A. (2011). Effects of Heterogeneous and Clustered Contact Patterns on Infectious Disease Dynamics. *PLoS Computational Biology*, *7*(6), e1002042. https://doi.org/10.1371/journal.pcbi.1002042

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications* (1st ed., Vol. 24). Cambridge University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Tables and Figures

Table 4.1 Reproductive ratio in core-periphery networks

k	$\tau = 0.02$	$\tau = 0.04$	$\tau = 0.075$	$\tau = 0.1$
1 (Homogeneous)	1 (0.1)	1.9 (0.1)	3.2 (0.1)	3.6 (0.1)
2	1 (0.1)	2 (0.1)	3 (0.2)	3.7 (0.2)
3	1 (0.1)	1.9 (0.1)	3.2 (0.2)	3.6 (0.2)
4	1 (0.1)	1.9 (0.1)	2.9 (0.2)	4 (0.3)

Average number of new infections caused by the index infected individual ($\bar{R}_i$) with varying coreness, k, and transmission rates, τ. Shown for networks with 5000 nodes, equal sized core and periphery groups, and 500 simulations.
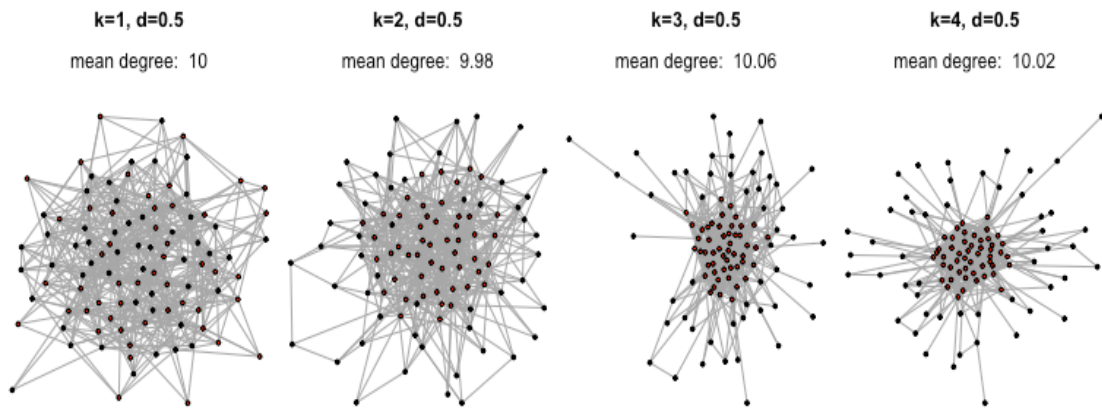
Figure 4.1 Core-periphery network examples.

Examples of core-periphery networks with increasing coreness (k). Networks shown have 100 nodes and 50% are in the core (d=0.5). Node color shows whether the node is "core" (red) or "periphery" (black).
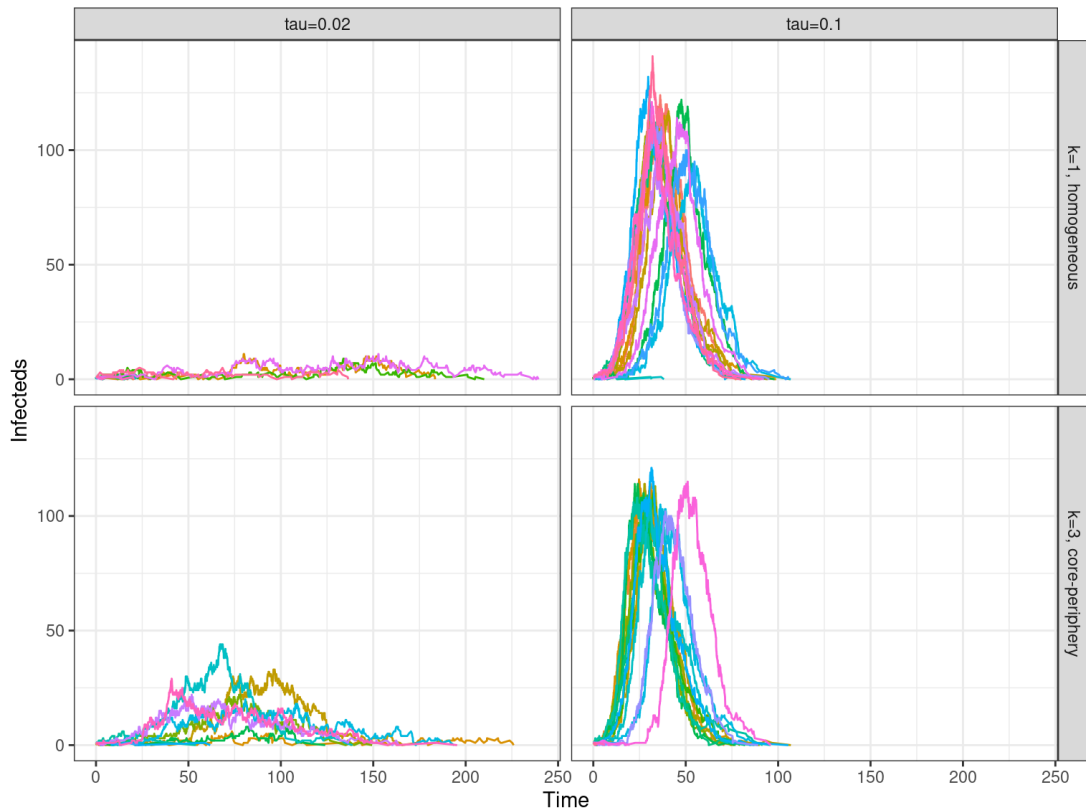
Figure 4.2 Simulated outbreaks in core-periphery and homogeneous networks.

Outbreaks are larger in core-periphery (shown for k=3) than homogeneous (k=1) networks when transmission rate is low, but not when transmission rates are high. Columns are faceted by transmission rate (τ). Results are shown for epidemics on networks with 500 nodes and 50% of nodes in the core (d=0. 5). Lines and colors show results from 20 simulation replicates.
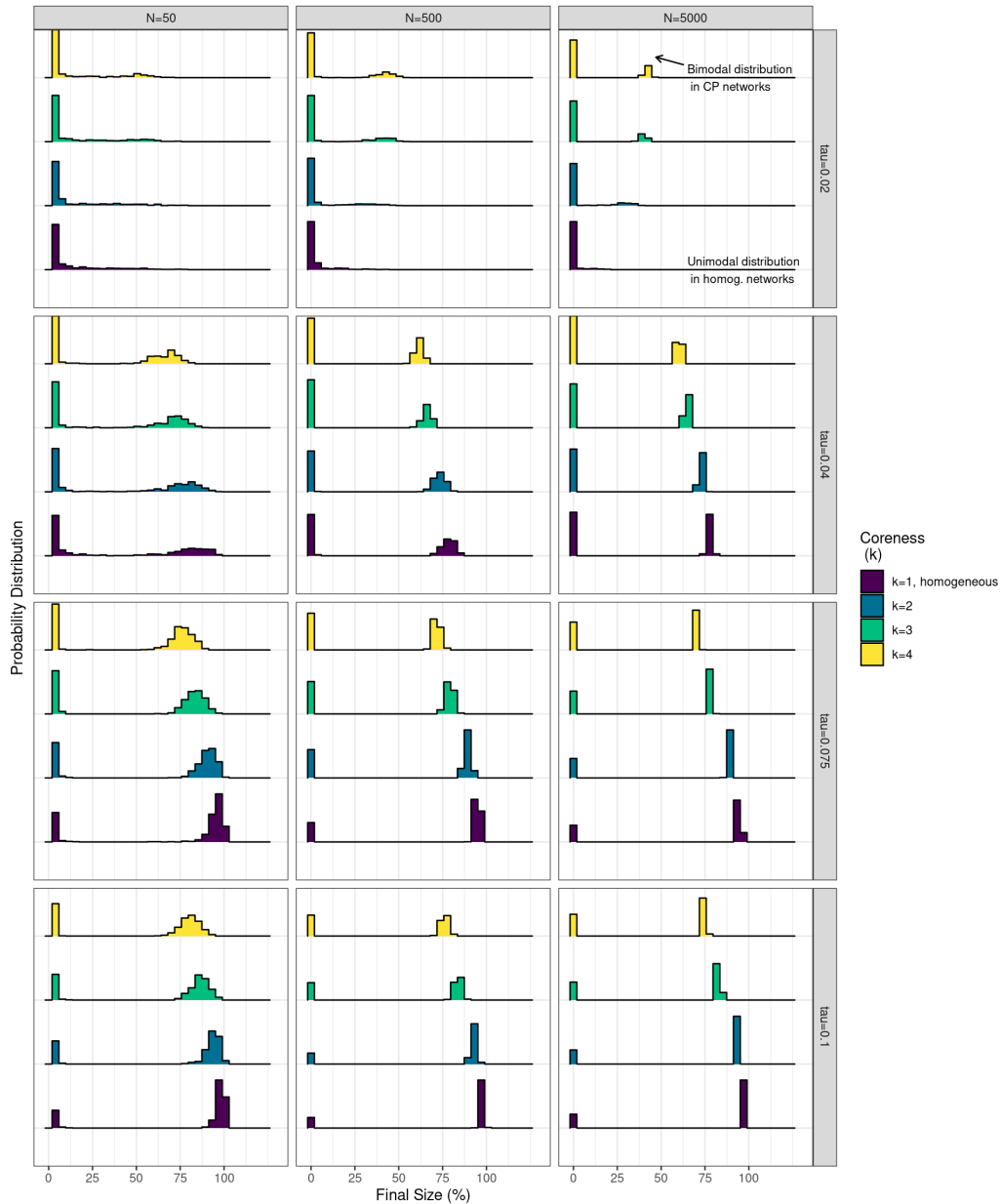
Figure 4.3 Effects of core-periphery networks on final size.

Core-periphery network structure increases the final size of outbreaks when transmission rates are low. Coreness (k) is shown in colors and the final size (percent of the population infected at the end of the simulation) is shown along the x-axis. Columns are faceted by network size and rows are faceted by transmission rate (τ). Histograms are scaled to the max count in each network size, transmission rate panel. Results shown for evenly sized core and periphery groups (d=0.5) and 500 simulations of each parameter combination.
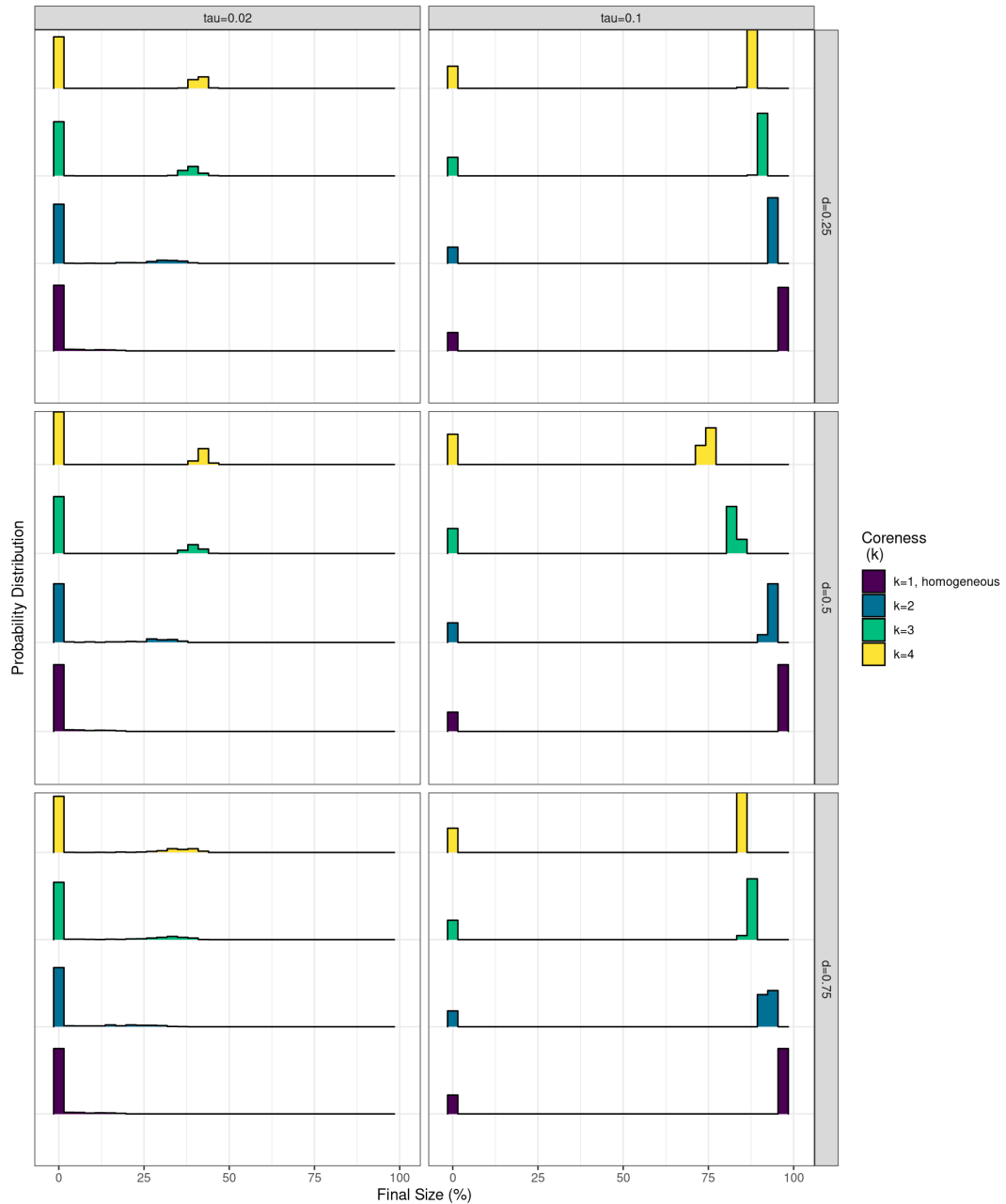
Figure 4.4 *Effects of core size on final size.*

The size of core relative to the periphery (d) affects the final size of simulated outbreaks. Coreness (k) is shown in the color of density distribution and the final size (percent of the population infected at the end of the simulation) is shown along the x-axis. Columns are faceted by transmission rate (τ) and rows are faceted by the relative size of the core (d). Histograms are scaled to the max count in each transmission rate, core size panel. Plot shows results for networks with 5000 nodes and 500 simulations of each parameter combination.
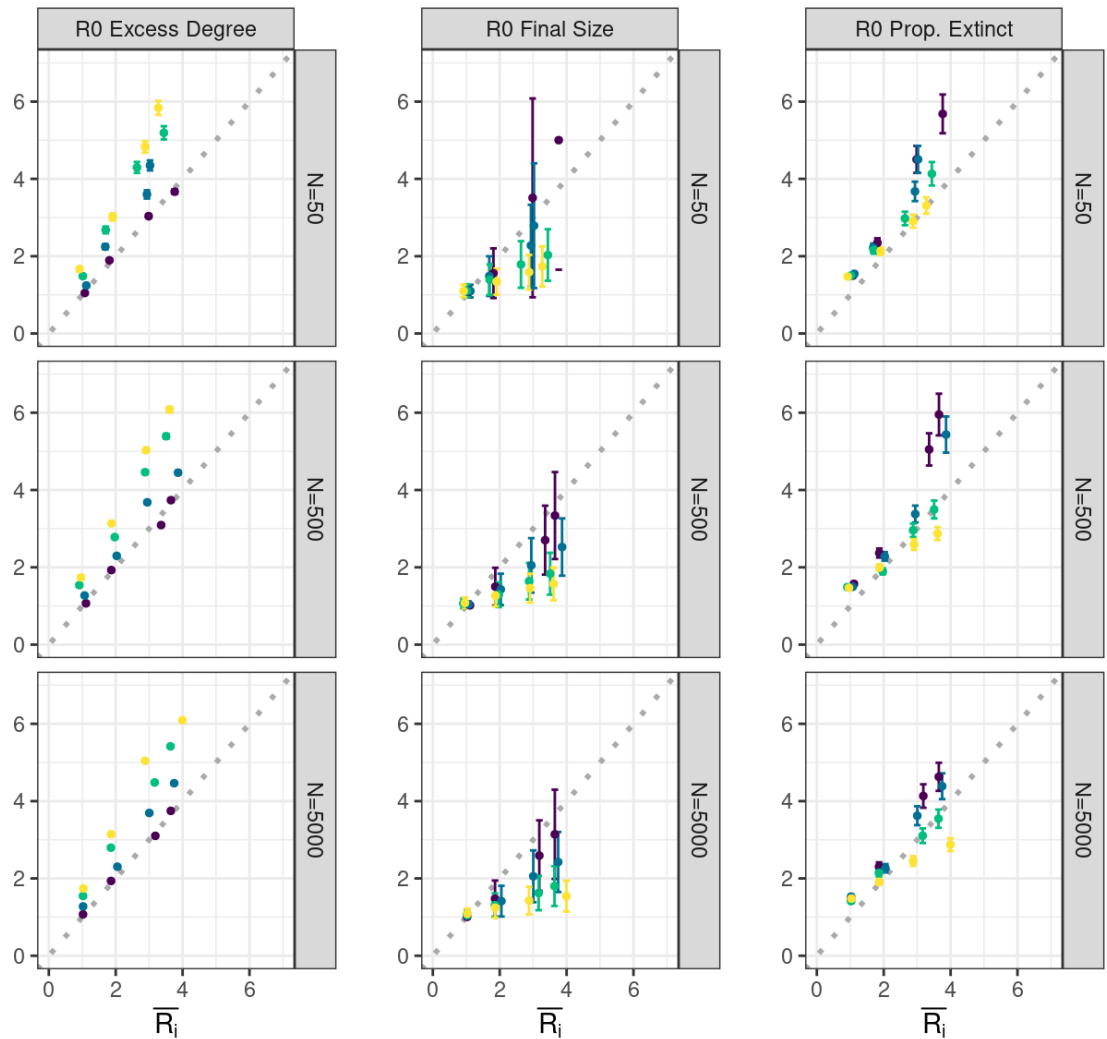
Figure 4.5 Relationship between the mean number of infections caused by the index case and estimators of the reproductive number.

$\widehat{R_0}$ estimators based on the excess-degree, final-size, and proportion-extinct estimators have varying relationships with the mean number of infections caused by the index case, $\bar{R}_\iota$. Coreness is shown in colors (same as figure 4), rows are faceted by network size, points and error bars show mean of 500 simulations and standard error of the mean. Plot shows results for networks with evenly sized core and periphery groups (d=0.5).
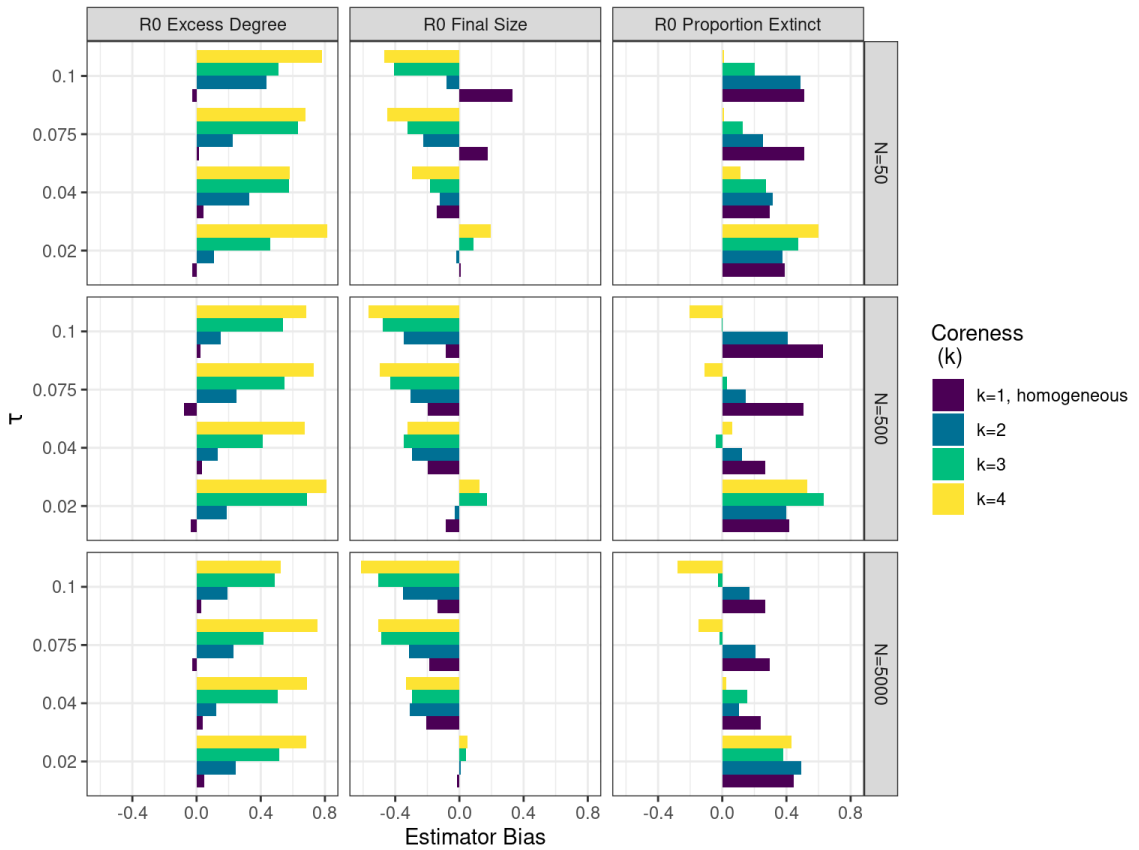
Figure 4.6 Bias of reproductive number estimators for outbreaks on core-periphery contact networks.

Bias of $\widehat{R_0}$ estimators calculated as $(\widehat{R_0} - \bar{R}_\iota)/\bar{R}_\iota$ where $\bar{R}_\iota$ is the mean number of infections caused by the index case for each set of parameters. Values less than zero indicate the estimator underestimated $\bar{R}_\iota$ and values above zero indicate overestimated $\bar{R}_\iota$. Coreness is shown in colors and τ on the y-axis. Rows are faceted by network size. Plot shows results for networks with evenly sized core and periphery groups (d=0.5) and mean bias calculated from 500 simulations.

Chapter 5

CHARACTERISTICS OF COVID-19 OUTBREAKS IN CARE, CORRECTIONAL,

AND FOOD PROCESSING FACILITIES IN THE UNITED STATES[4]

Abstract

COVID-19 has disproportionately affected populations living and working in care, correctional, and food processing facilities in the United States but we lack a broader understanding of the extent and variation in outbreak sizes within these populations. The distribution of outbreak sizes for COVID-19 should include frequent, large outbreaks given the reproductive ratio is estimated to be larger than one unless control strategies have successfully mitigated spread. Because these populations have highly structured, non-homogenous contact strategies, we modeled the spread of COVID-19 in core-periphery networks and characterized the outbreak size distribution under a number of scenarios. We then analyzed the outbreak size distribution in United States care, correctional, and food processing facilities. We show a mismatch between simulated outbreak distributions, without interventions, and actual outbreak distributions in these populations. We suggest that most large outbreaks were prevented by the interventions taken by these facilities. To better understand how the mitigation of COVID-19 may have been achieved, we study the outbreak size distribution in care facilities over time and compare the size of outbreaks in facilities with access to two specific interventions. These results provide preliminary, but promising, support that the interventions taken have reduced spread in vulnerable populations.

Introduction

In the United States, COVID-19 cases and deaths have been concentrated in settings such as nursing homes, prisons, and food processing plants due to high population densities, pressure to work while sick, high prevalence of pre-existing conditions, and limited access to healthcare (D'Adamo et al., 2020; McMichael et al., 2020; Waltenburg et al., 2020). For example, in late February an outbreak in a long-term care facility in Washington led to 81 resident cases, 34 staff cases, and 23 deaths (McMichael et al., 2020). Frequent outbreaks have similarly been reported in correctional and food processing (Waltenburg et al., 2020) facilities but we lack a broader understanding of the proportion of individuals infected, or the outbreak sizes, within these settings and this is an identified gap in our understanding of spread in these vulnerable populations (Wallace et al., 2020).

The distribution of outbreak sizes can be "J-Shaped" or "U-Shaped" (Kendall, 1956; Nāsell, 1995). The term "J-shape" refers to a distribution of outbreak sizes that is monotonically decreasing with a mode at the origin while the term "U-shape" refers to a bimodal distribution. Theoretically, the transition between "J-Shaped" and "U-Shaped" occurs when the reproductive ratio is above one" (Kendall, 1956; Nāsell, 1995) but contact network structure can affect where the transition occurs (Pastor-Satorras & Vespignani, 2001; Shirley & Rushton, 2005). We expected care, correctional, and food processing facilities would all have highly structured contact patterns with large variation in contact rates

113

between staff, residents, and workers at food processing plants being a notable

feature (Dool et al., 2008), especially in lockdown settings due to COVID-19

(Grabowski & Mor, 2020). Though contact structure is important for determining

the threshold for large outbreaks to occur, and thus the shape of the outbreak

size distribution, the reproductive ratio for COVID-19 has generally been

estimated to be much larger than one (e.g., Kucharski et al., 2020). In this

situation, we would expect "U-Shaped" outbreak size distributions unless disease

control measures effectively mitigated transmission.

Here, we modeled the unmitigated spread of COVID-19 in structured

contact networks with a range of transmission rates and initial conditions to

understand the effects on the distribution of outbreak sizes. We then compiled

publicly available data and characterized the outbreak size distribution in United

States care, correctional, and food processing facilities. Our main result is to

show the mismatch between simulated outbreak distributions, without

interventions, and actual outbreak distributions in these populations. In addition,

we used descriptive analyses to study changes in outbreak sizes over time,

space, and with access to two interventions. Combined, this study provides

preliminary, but promising, support that the interventions taken have reduced

spread in vulnerable populations.

Methods

Simulated Outbreak Size Distribution

To model the spread of COVID-19 in core-periphery networks, we used a Susceptible- Latent- Infectious- Recovered model. We assumed a mean infectious period of 3.5 days $(1/\gamma)$ which was informed by estimates of the time from symptom onset to hospitalization in residents at a Washington skilled nursing facility (Arentz et al., 2020). We also assumed a 3.5 day incubation period $(1/\sigma)$ when individuals were not yet infectious (Li et al., 2020).

To study how the shape and mean of the outbreak size distribution changes across scenarios, we varied the reproductive number and initial conditions. We varied the reproductive number by setting the transmission rate (τ) such that the basic reproductive number ranged from 0.8 (for facilities that may have controlled the outbreak) to 3.5 (Gatto et al., 2020). The reproductive number was calculated based on the network structure as $\tau < K^2 - K >/(\tau + \gamma) < K >$ where $< K >$ represents the mean degree and $< K^2 - K >$ represents the degree variance (Kiss et al., 2017). To determine the effect of the number of initially infectious individuals on the shape and mean of the outbreak size distribution, we sparked each simulation by one, three, or five individuals at random. To calculate the outbreak size, we ran each simulation until there were zero infected individuals. We simulated the model with each transmission rate and initial condition combination 1,000 times. We implemented the model using a continuous-time, Gillespie algorithm in the Epidemics on Networks module (Miller

& Ting, 2020) in Python Version 2 (further details on the algorithm are given in Appendix II).

To model the structure of core-periphery networks, we used the algorithm presented in Chapter 4 of this Dissertation. Specifically, we assumed 25% of nodes were in the core and 75% were in the periphery. We set the parameter that controls how densely connected the core is relative to the periphery, k, to 4 representing highly structured networks that we propose may be present in some nursing homes and prisons. All networks were initialized with a mean degree of 10 and 500 nodes (Appendix IV, Figure S4.1 for how this compares to population size in care, correctional, and food processing facilities).

Outbreak Size Distribution in Care, Correctional, and Food processing Facilities

We gathered data about completed (i.e., no reported ongoing transmission) outbreaks in populations associated with three contexts: care facilities, correctional facilities, and food processing facilities to compare with simulated data. For care facilities, we used the CDC's National Healthcare Safety Network (NHSN) system data. NHSN data were downloaded on November 17, the range of reporting dates was from May 24 to November 11. The majority of facilities reported each week and we used the last reporting date (November 11) for total confirmed cases per facility. We excluded rows of data that did not pass the data quality assurance test (as recommended by NHSN). To meet inclusion criteria about outbreaks not being ongoing, we also excluded facilities that (1) reported 3 or more confirmed cases in the week leading up to the last reporting

date, (2) facilities that reported more resident cases than occupied resident beds as these likely reflect facilities with high turnover, (3) facilities that reported 0 total resident cases as we were specifically interested in facilities that had an introduced case of COVID-19 among the resident population, and (4) facilities that had fewer than 10 residents to improve comparisons across facilities of varying sizes.

We characterized trends in care facility outbreaks over time, space, and with access to two interventions (access to resident testing and asymptomatic surveillance of residents the week before the first case was confirmed in the facility). We compared outbreak sizes over time by extracting the date (week since the first reporting date, May 24) that the facility reported their first resident case and used Spearman's correlation coefficient to measure the strength of the relationship between outbreak size and the date of first resident case across facilities. We additionally compared spatial trends in the outbreak size distributions by US states and regions (defined with Census Bureau-designated major regions). We also examined the relationship between the average timing of outbreaks (mean week of first resident case) and the average outbreak size across states using Pearson's correlation coefficient. In addition, we described trends in the proportion of facilities that "flattened" the COVID-19 outbreak curve over time (across month of first resident case reported) by calculating the proportion of facilities that ever reported more than 5 confirmed COVID-19 cases among residents and staff in a single week. Finally, we extracted information

from the NHSN data about facility access to interventions (Appendix IV, Figure

S4.2). We chose to focus on two interventions that were widely reported: whether

the facility had in-house testing and whether the facility reported doing

surveillance testing of asymptomatic residents. Because these intervention

variables were only available in reporting weeks starting on August 16,

intervention analyses only included facilities that reported their first case later

than August 16. For both interventions, facilities were recorded as having access

to the intervention if they reported access to the intervention the week prior to

their first resident case reported. We compared the outbreak sizes of facilities

that reported access to interventions compared with those that did not using t-

tests.

      For correctional facilities, we used UCLA's Behind Bars Dataset, which

was downloaded on September 8 and September 14, 2020, when approximately

186,000 people had been infected and at least 1,120 inmates and correctional

officers had died according to reporting from the New York Times. We excluded

facilities from the September 14 dataset if there were 3 or more cases reported

compared with the September 8 dataset. As we were specifically interested in

transmission among facilities with residents and staff, we also excluded facilities

with zero estimated incarcerated individuals (e.g., administrative buildings).

Lastly, we excluded correctional facilities with more cases than estimated daily

population sizes which likely represent high turnover and do not meet our

definition of closed populations. Because we were interested in the proportion of

the population in each facility that was infected, we supplemented Behind Bars dataset with resident and staff population sizes using data we gathered from each state department of corrections website.

Finally, for food processing facilities, we extracted clusters from the New York Times Cluster Tracker, which lists reported COVID-19 clusters of all types linked to at least 50 cases. To specifically extract the outbreaks in food processing facilities, we searched for the following keywords: "MEAT", "PORK", "BEEF", "POULTRY", "SEAFOOD", "FARM", "FISH", "DAIRY", "JBS", "POTATOES", "FOOD", "SAUSAGE", "PACKING", "TURKEY", "BACON", "APPLE", "PROCESSING", "PACKAGING", and "FRUIT". We downloaded these data on September 8 and September 14, 2020. We excluded facilities from the September 14 dataset that had 3 or more cases than the September 8 dataset. At this time, the most recent CDC report about COVID-19 in food processing facilities was released on July 10, 2020 and estimated that among 23 states reporting COVID-19 outbreaks in meat and poultry processing facilities, 16,233 cases in 239 facilities occurred, including 86 deaths (Waltenburg et al., 2020). Similar to our analysis for correctional facilities, we supplemented data on outbreaks at food processing facilities with information about employees working at each facility from each company's websites.

We defined the final outbreak size as the number of reported cases divided by population size (Appendix IV, Figure S4.1). Because there was varying information for each facility type, we calculated the final size for residents

119

(measured by number of occupied beds) in care facilities, for incarcerated individuals (measured by prison capacity) in correctional facilities, and for workers (measured by number of staff) in food plants.

All data analyses and simulations were performed in R version 4.0.0 and Python Version 2. Scripts to reproduce results are available at https://github.com/CEIDatUGA/miller-covid-clusters.

Results

The theoretical shape of the outbreak size distribution as demonstrated by simulations was qualitatively affected by the basic reproductive number and was only "J-shaped" if the basic reproductive number was less than or close to 1. Otherwise it was "U-shaped" (Figure 5.1). Initial conditions affected the height of the outbreak size distribution at smaller outbreak sizes, especially when the reproductive number was close to one. For example, when the reproductive ratio was one, 67% of outbreaks had fewer than 10 infected individuals when there was one index case compared with only 25% of outbreaks when there were 3 initially infected individuals. The mean outbreak size was affected by both the reproductive number and the initial conditions. When the reproductive ratio was 1.2, the mean outbreak size was 0.05 when there was one introduced case, 0.10 when there were 3 introduced cases, and 0.12 when there were 5 introduced cases.

In United States care (n=8,863) and correctional (n=274) facilities, the outbreak size distributions were "J-shaped" (Figure 5.2). Among care and

correctional facilities reporting at least 1 resident case, the mean outbreak size was 0.30 and 0.13, respectively. In food processing facilities (n=80), which all lacked data on outbreaks with fewer than 50 cases, the outbreak size distribution was "J-shaped" in the tail of the distribution for which we had data. The mean outbreak size among workers at food processing facilities was 0.23.

In care facilities, the shape and mean of the outbreak size distribution varied depending on the month when the facility reported the first case among residents (Figure 5.2D). If the facility reported the first case from January through May 2020 (n=3,636), the mean size was 0.362 and 91% of outbreaks had 3 or more resident cases compared with a mean size of 0.13 and only 43% of outbreaks with 3 or more resident cases in facilities reporting their first case in October 2020 (n=354). There was a significant correlation between week of first case and outbreak size across all facilities (Spearman's correlation coefficient, -0.32, $P < 0.0001$). When excluding the first reporting week of May 24, the correlation between the date of first case and outbreak size across all facilities decreased, but remained high (Spearman's correlation coefficient, -0.26, $P < 0.0001$).

The outbreak size distribution also varied by state with Massachusetts (mean outbreak size: 0.43, n=227), Connecticut (mean outbreak size: 0.40, n=145), and South Carolina (mean outbreak size: 0.40, n=136) having the largest average outbreak sizes. In contrast, Hawaii (mean outbreak size: 0.09, n=16), Oregon (mean outbreak size: 0.14, n=47), and Wyoming (mean outbreak size:

0.15, n=9) reported the smallest outbreak sizes among residents. By region, facilities in the West and Midwest reported the smallest outbreaks while facilities in the South and Northeast reported the largest outbreaks. In general, facilities in states that had large average outbreak sizes also reported earlier outbreaks (Figure 5.4, Spearman's correlation coefficient between average week of first resident case by state and mean outbreak size by state was -0.54, P<0.0001).

The later the facility reported their first resident case, the less likely it was to report large weekly increases in resident cases (Figure 5.5). In facilities that reported their first resident case from January to May, 84% of facilities reported 5 or more cases per week at least once. In contrast, this percentage dropped to 68% in July and to 10% in November. These trends suggest that facilities that had later outbreaks successfully "flattened the curve" likely due to a combination of increased vigilance and improved access to interventions which I investigate next.

In the 2,021 care facilities that reported their first resident case after August 16 (when intervention data were available), the mean outbreak size was 6% (95% CI: 2-8%) lower in facilities that reported residents having access to testing the week before the first resident case was reported (mean outbreak size: 0.17, n=1,050) compared with those that did not report access to testing (mean outbreak size: 0.23, n=504) and this was a significant difference (t-test, $t = -3.9$, $P = 0.0001$). The final size was also 5% (95% CI: 2-7%) lower among care facilities that reported surveillance testing of asymptomatic residents the week

before the first case was reported (mean outbreak size: 0.17, n=671) compared with facilities that did not (mean outbreak size: 0.23, n=1,091) which was also a significant difference (t-test, $t = -3.5, P = 0.0004$). In addition, there were 6,347 fewer COVID-19 cases (and 917 fewer deaths) among residents in facilities that reported doing asymptomatic surveillance testing of residents the week leading up to the first confirmed case compared with facilities that did not. Finally, both interventions were reported with increased frequency over time (Appendix IV, Figure S4.2) with less than 10% of facilities reporting "point of care testing" within the facility in August compared with 87% of facilities by November. Similarly, surveillance testing of asymptomatic residents increased among facilities from approximately 30% of facilities in August to nearly 50% of facilities in November.

Discussion

Outbreak size distributions are indicative of underlying rates of transmission in populations. Without interventions, we showed how the outbreak size distribution in structured contact networks resemble one of two shapes, "J-Shaped" and "U-Shaped", as (Kendall, 1956) and others have previously shown for well-mixed populations. As predicted by theory, the distributions were only "J-Shaped" when the basic reproductive ratio was close to one, indicating outbreaks are expected to go extinct before causing large epidemics in the population (Nāsell, 1995). The fact that we found COVID-19 outbreak size distributions in US care, correctional, and food processing facilities had "J-Shaped" distributions,

despite COVID-19 having an estimated reproductive ratio much larger than one (Gatto et al., 2020), is evidence that control measures have mitigated spread.

In United States care facilities, we explored temporal and spatial trends in the outbreak size distribution. There was a significant negative relationship between outbreak size and when each facility reported their first resident case, but most of the variation in outbreak sizes that was explained by the date of first resident case was due large outbreaks reported prior to the first reporting week on May 24. This indicates a discontinuity in how outbreak sizes were changing over time, with facilities that had early outbreaks (beginning prior to May 24) reporting much larger outbreaks on average than those that began after May 24. In addition, there was wide variation in outbreak sizes across states with states in the Northeast and South reporting the largest outbreak sizes in the country while states in the West and Midwest reported the smallest outbreaks. Some variation in outbreak sizes across states is likely due to the average timing of spread in each state, and there was a negative correlation between the average timing of the first case report date and mean outbreak size at the state level. This suggests a variable representing community transmission will help explain variation in outbreak sizes at the facility-level. In addition, there was a large increase over time in the number of facilities that flattened their outbreak curves. Together, these patterns reflect expected trends in outbreak sizes if care facilities had increased outbreak awareness and vigilance in control strategies over time.

Exactly which interventions reduced spread in these facilities is an important question for control in these vulnerable populations. In care facilities, we studied the relationship between outbreak size and two interventions: whether residents had access to testing in the facility and whether asymptomatic surveillance testing of residents was performed the week leading up to the first confirmed resident case in the facility. Because intervention variables were largely reported only after August 16, we only used the subset of facility data that reported their first confirmed resident case after this date. However even within this limited time frame, the average outbreak size was smaller among facilities that reported either of these two interventions the week leading up to the first resident case compared with facilities that did not report these interventions. Finally, consistent with our hypothesis that temporal declines in outbreak size were due to increased outbreak awareness and control vigilance, both of these interventions were reported with increased frequency over time.

This analysis has several limitations. In the simulation study, we assumed heterogeneous contact patterns structured into core (staff) and periphery (resident) groups (Dool et al., 2008; Grabowski & Mor, 2020). Detailed case studies in these settings, particularly ones describing contact patterns over time and in food processing facilities, would be valuable for improved understanding of spread. In addition, the transmission model we used is simple: it does not account for presymptomatic transmission or age-specific susceptibility. However, the main focus of the simulations was to show the outbreak size distributions in

125

structured populations under non-intervention scenarios and this result is unlikely to change with more detailed transmission models. The datasets were also limited to self-reported (care) and publicly-available (correctional, food) data. Food processing data in particular were sparsely available, highlighting the importance of a centralized surveillance system, similar to that of NHSN, to understanding outbreaks in these settings. Another dataset limitation was that reporting of some intervention variables to the NHSN began August 16 and as a result we reported differences in mean outbreak sizes with intervention access in facilities that reported their first case after this date. Other variables, such as access to N95 masks and resident access to testing were available prior to August 16 but there was little variation across facilities (generally facilities reported having both interventions), making it hard to determine their effects. Finally, the analysis of care facility data was descriptive and its aim was to provide support for our hypothesis that the wide deviation from expected outbreak size distributions could be due to increased awareness and uptake of control strategies. Detailed time series analyses and regressions accounting for facility outbreak timing, levels of community transmission (to inform the number of COVID-19 introductions to each facilities), and interventions may yield a more precise understanding of how COVID-19 can be controlled in these high-transmission settings.

Conclusions

Outbreak size distributions are indicative of underlying rates of transmission in populations. COVID-19 has led to frequent and deadly outbreaks in care, correctional, and food processing facilities in the US. However, this study suggests that improvement and uptake of control strategies over time mitigated spread and prevented the majority of large outbreaks from occurring. Though promising, the global pandemic is far from over and control measures will likely need to be in place in these populations for the foreseeable future.

Author Contributions

PM designed and implemented simulation study, oversaw data collection, analyzed data, wrote manuscript. CK, DV, TB, JK, VL, CR, OT, and JW assisted with data collection and edited manuscript. JD designed simulation study and edited manuscript.

References

Arentz, M., Yim, E., Klaff, L., Lokhandwala, S., Riedo, F. X., Chong, M., & Lee,
M. (2020). Characteristics and Outcomes of 21 Critically Ill Patients With
COVID-19 in Washington State. *JAMA*, *323*(16), 1612–1614.
https://doi.org/10.1001/jama.2020.4326

D'Adamo, H., Yoshikawa, T., & Ouslander, J. G. (2020). Coronavirus Disease
2019 in Geriatrics and Long-Term Care: The ABCDs of COVID-19.
Journal of the American Geriatrics Society, *68*(5), 912–917.
https://doi.org/10.1111/jgs.16445

Dool, C. van den, Bonten, M. J. M., Hak, E., Heijne, J. C. M., & Wallinga, J.
(2008). The Effects of Influenza Vaccination of Health Care Workers in
Nursing Homes: Insights from a Mathematical Model. *PLoS Medicine*,
5(10), e200. https://doi.org/10.1371/journal.pmed.0050200

Gatto, M., Bertuzzo, E., Mari, L., Miccoli, S., Carraro, L., Casagrandi, R., &
Rinaldo, A. (2020). Spread and dynamics of the COVID-19 epidemic in
Italy: Effects of emergency containment measures. *Proceedings of the
National Academy of Sciences*, *117*(19), 10484–10491.
https://doi.org/10.1073/pnas.2004978117

Grabowski, D. C., & Mor, V. (2020). Nursing Home Care in Crisis in the Wake of
COVID-19. *JAMA*, *324*(1), 23–24. https://doi.org/10.1001/jama.2020.8524

Kendall, D. G. (1956). *Deterministic and Stochastic Epidemics in Closed
Populations* (https://projecteuclid.org/euclid.bsmsp/1200510207). 149–
166. https://projecteuclid.org/euclid.bsmsp/1200502553

Kiss, I. Z., Miller, J. C., & Simon, P. L. (2017). *Mathematics of epidemics on
networks* (Vol. 46). Springer.

Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., &
Eggo, R. M. (2020). Early dynamics of transmission and control of COVID-
19: a mathematical modelling study. *The Lancet Infectious Diseases*,
20(5), 553–558. https://doi.org/10.1016/s1473-3099(20)30144-4

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., & Shaman, J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, *368*(6490), 489–493. https://doi.org/10.1126/science.abb3221

McMichael, T. M., Clark, S., Pogosjans, S., Kay, M., Lewis, J., Baer, A., Kawakami, V., Lukoff, M. D., Ferro, J., Brostrom-Smith, C., Riedo, F. X., Russell, D., Hiatt, B., Montgomery, P., Rao, A. K., Currie, D. W., Chow, E. J., Tobolowsky, F., Bardossy, A. C., … Harney, J. (2020). COVID-19 in a Long-Term Care Facility — King County, Washington, February 27–March 9, 2020. *MMWR. Morbidity and Mortality Weekly Report*, *69*(12), 339–342. https://doi.org/10.15585/mmwr.mm6912e1

Miller, J. C., & Ting, T. (2020). EoN (Epidemics on Networks): a fast, flexible Python package for simulation, analytic approximation, and analysis of epidemics on networks. *Journal of Open Source Software*, *4*(44), 1731. https://doi.org/10.21105/joss.01731

Nāsell, I. (1995). Epidemic models: Their structure and relation to data. In D. Mollison (Ed.), *The threshold concept in stochastic epidemic and endemic models*. Cambridge University Press. https://doi.org/10.1007/bf02459495

Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic dynamics and endemic states in complex networks. *Physical Review E*, *63*(6 Pt 2), 066117. https://doi.org/10.1103/physreve.63.066117

Shirley, M. D. F., & Rushton, S. P. (2005). The impacts of network topology on disease spread. *Ecological Complexity*, *2*(3), 287–299. https://doi.org/10.1016/j.ecocom.2005.04.005

Wallace, M., Hagan, L., Curran, K. G., Williams, S. P., Handanagic, S., Bjork, A., Davidson, S. L., Lawrence, R. T., McLaughlin, J., Butterfield, M., James, A. E., Patil, N., Lucas, K., Hutchinson, J., Sosa, L., Jara, A., Griffin, P., Simonson, S., Brown, C. M., … Marlow, M. (2020). COVID-19 in Correctional and Detention Facilities — United States, February–April 2020. *MMWR. Morbidity and Mortality Weekly Report*, *69*(19), 587–590. https://doi.org/10.15585/mmwr.mm6919e1

Waltenburg, M. A., Victoroff, T., Rose, C. E., Butterfield, M., Jervis, R. H., Fedak, K. M., Gabel, J. A., Feldpausch, A., Dunne, E. M., Austin, C., Ahmed, F. S., Tubach, S., Rhea, C., Krueger, A., Crum, D. A., Vostok, J., Moore, M.

J., Turabelidze, G., Stover, D., … Team, C.-19 R. (2020). COVID-19 Among Workers in Meat and Poultry Processing Facilities - United States, April-May 2020. *MMWR. Morbidity and Mortality Weekly Report*, *69*(27), 887–892. https://doi.org/10.15585/mmwr.mm6927e2
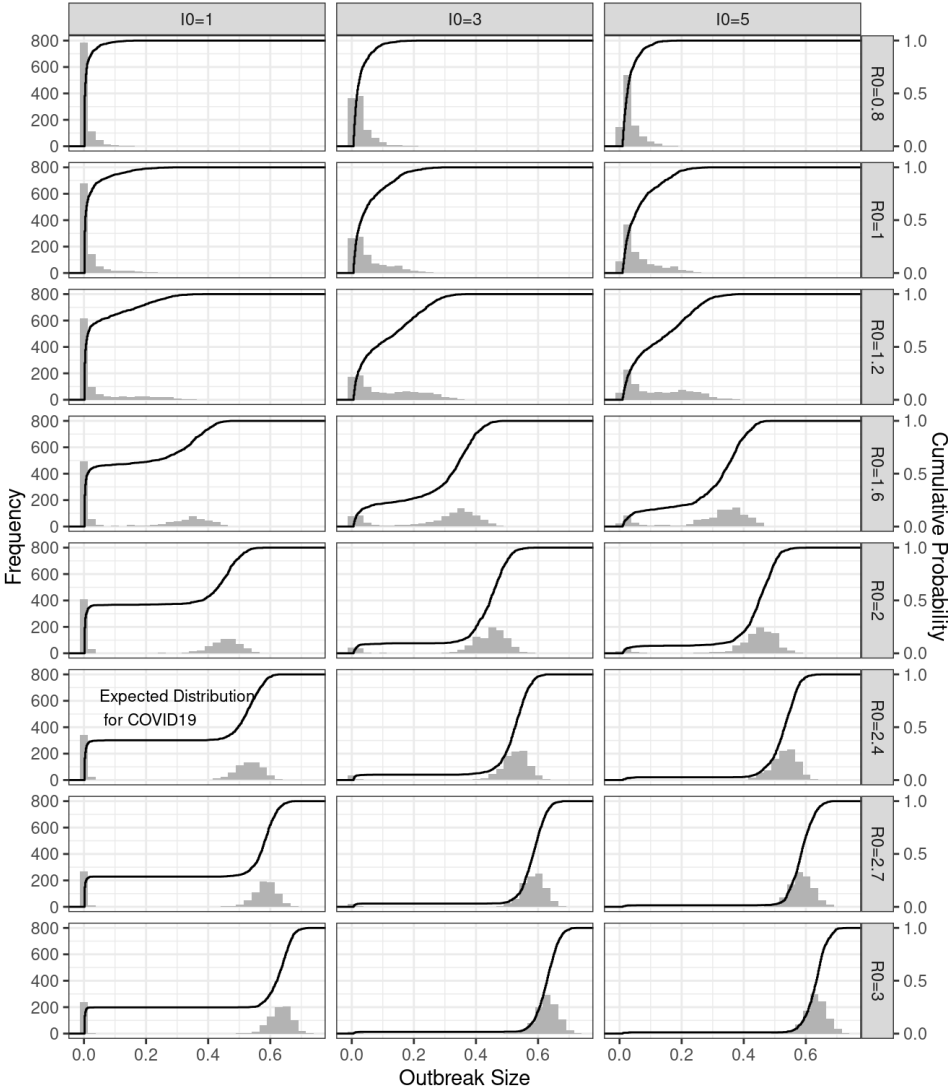
Figure 5.1 Simulated outbreak size distribution.

Simulated outbreak size distribution with varying reproductive ratios (rows) and numbers of initially infectious cases (columns) show "J-shaped" curves are only found when the reproductive ratio is approximately one. Each plot shows the probability (histograms) and cumulative probability distribution (lines) of outbreak sizes.
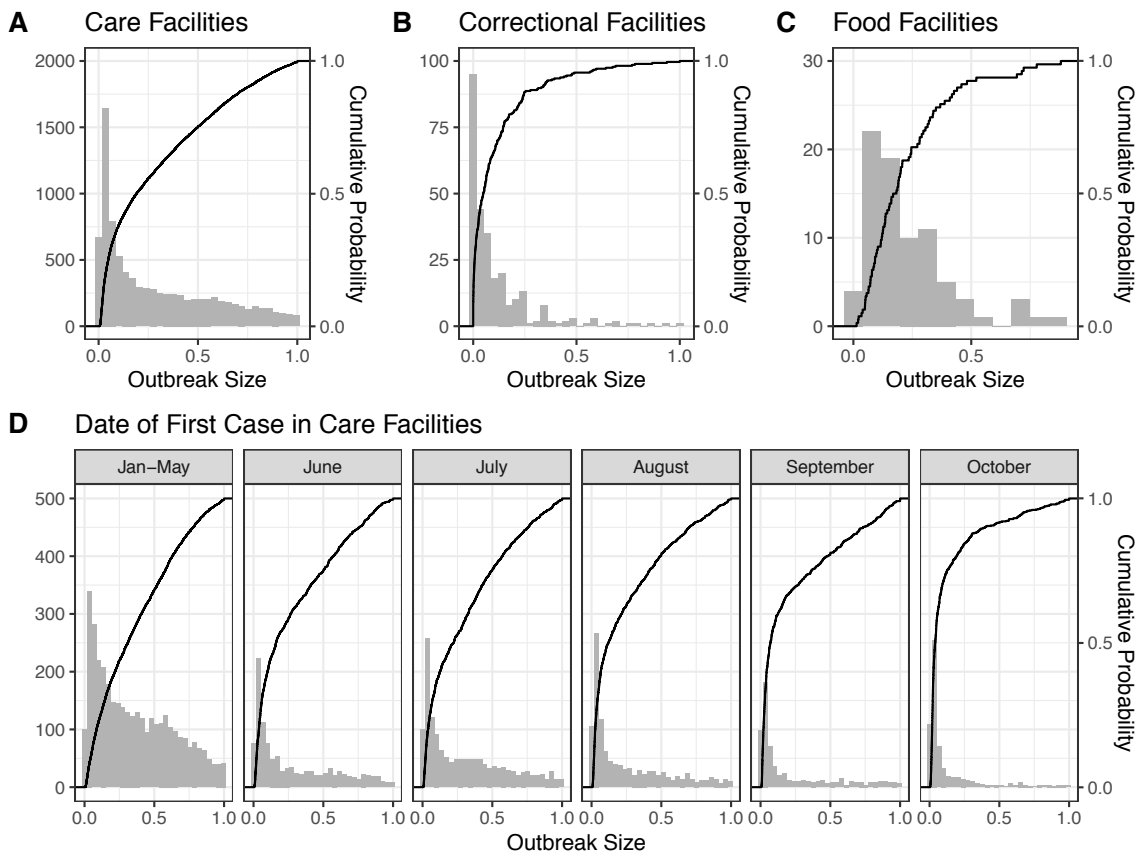
Figure 5.2 COVID-19 Outbreak Size Distributions in the United States.

COVID-19 outbreak size distributions in United States care (A), correctional (B), and food processing (C) facilities are "J-Shaped". In care facilities (D), facilities reporting their first case in later months (labels) had outbreak size distributions that were more concentrated around small outbreaks. Each plot shows the frequency (histograms) and cumulative probability distribution (lines) of outbreak sizes.
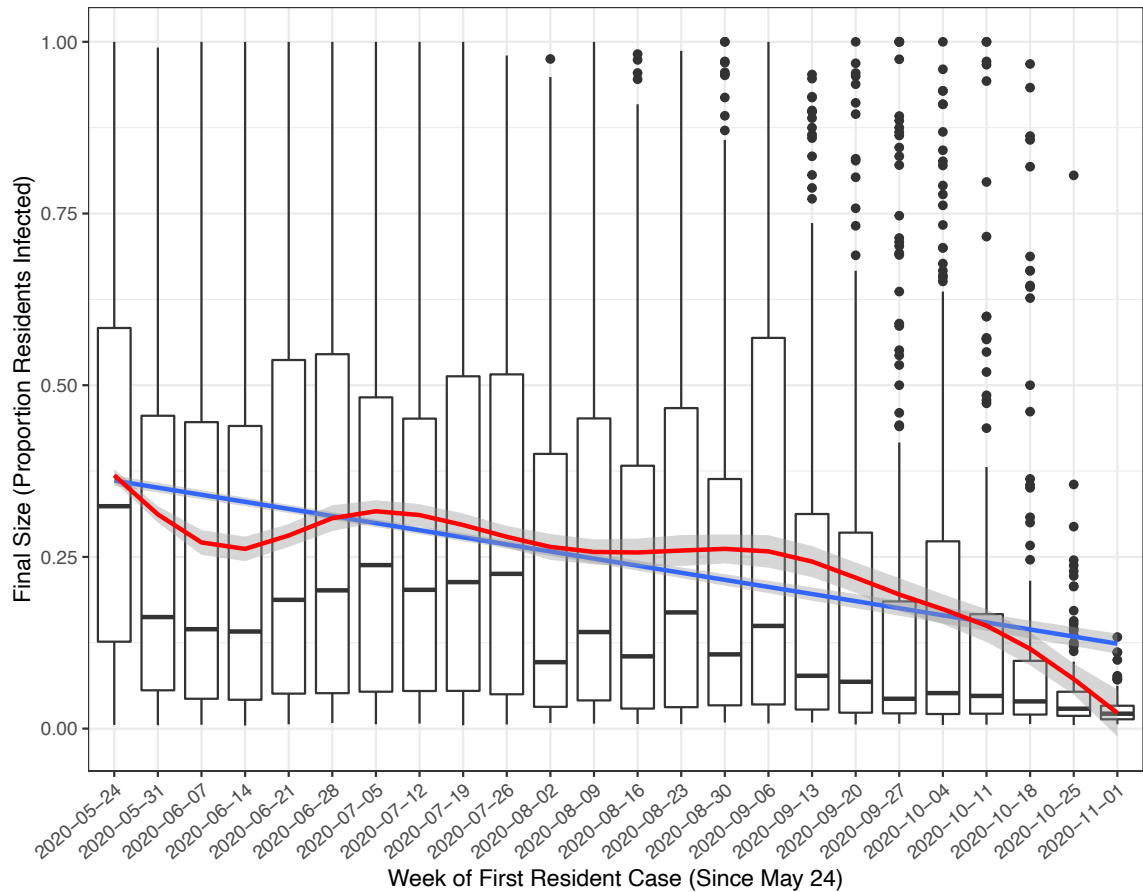
Figure 5.3 Decreasing trend in care facility COVID-19 outbreak sizes over time.

In the United States, COVID-19 outbreak sizes among residents (y-axis) in care facilities decreases over time (measured by week of first resident case reported shown on the x-axis). Boxplots represent the distribution of outbreak sizes in care facilities that reported their first resident case that week. Lines represent linear (blue) and general additive (red) model fits.
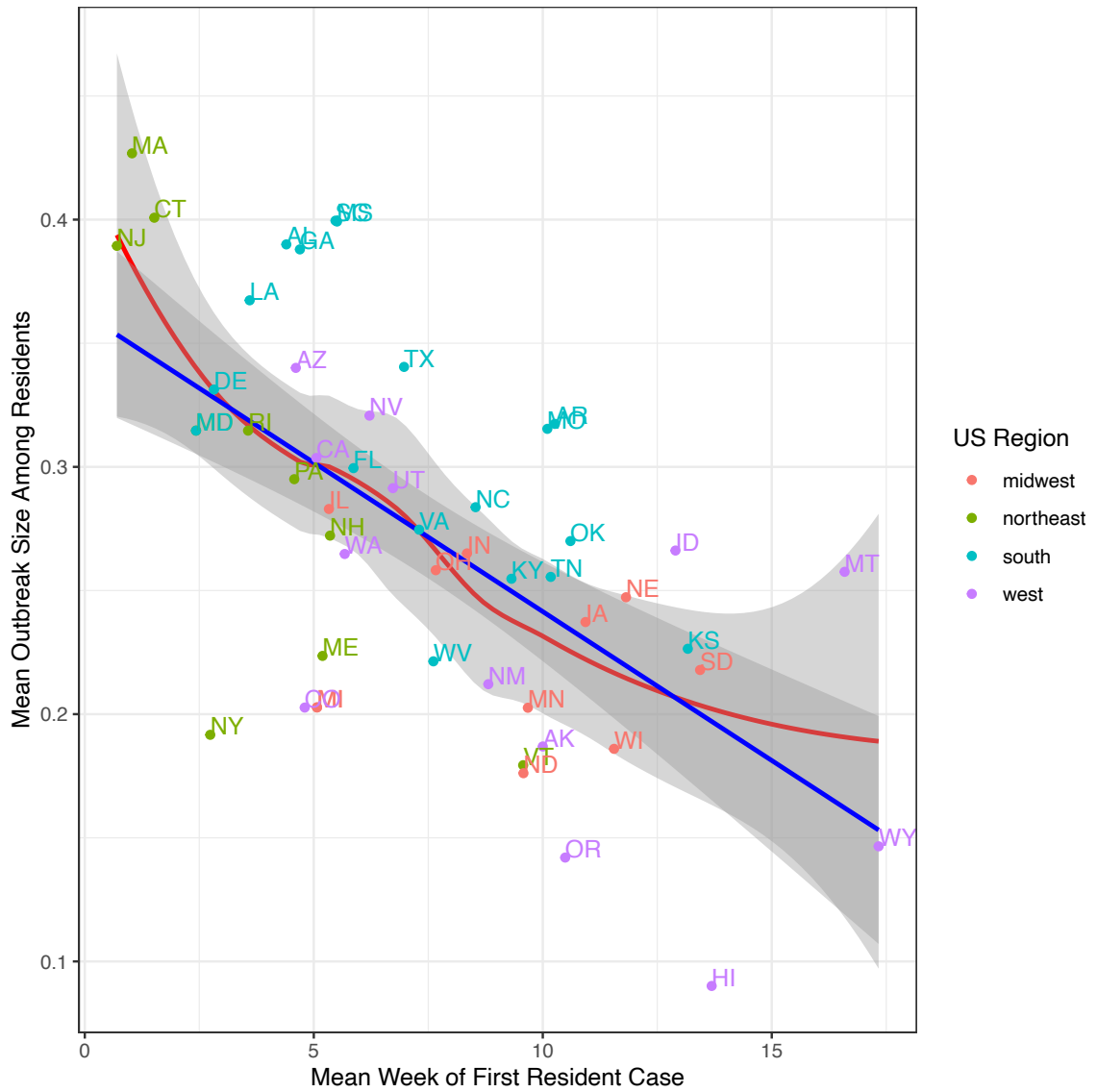
Figure 5.4 Spatial and temporal trends in care facility COVID-19 outbreak sizes.

In the United States, the mean outbreak size by state (y-axis) decreases over time (measured by mean week of first resident case reported shown on the x-axis). Lines represent linear (blue) and general additive (red) trends.
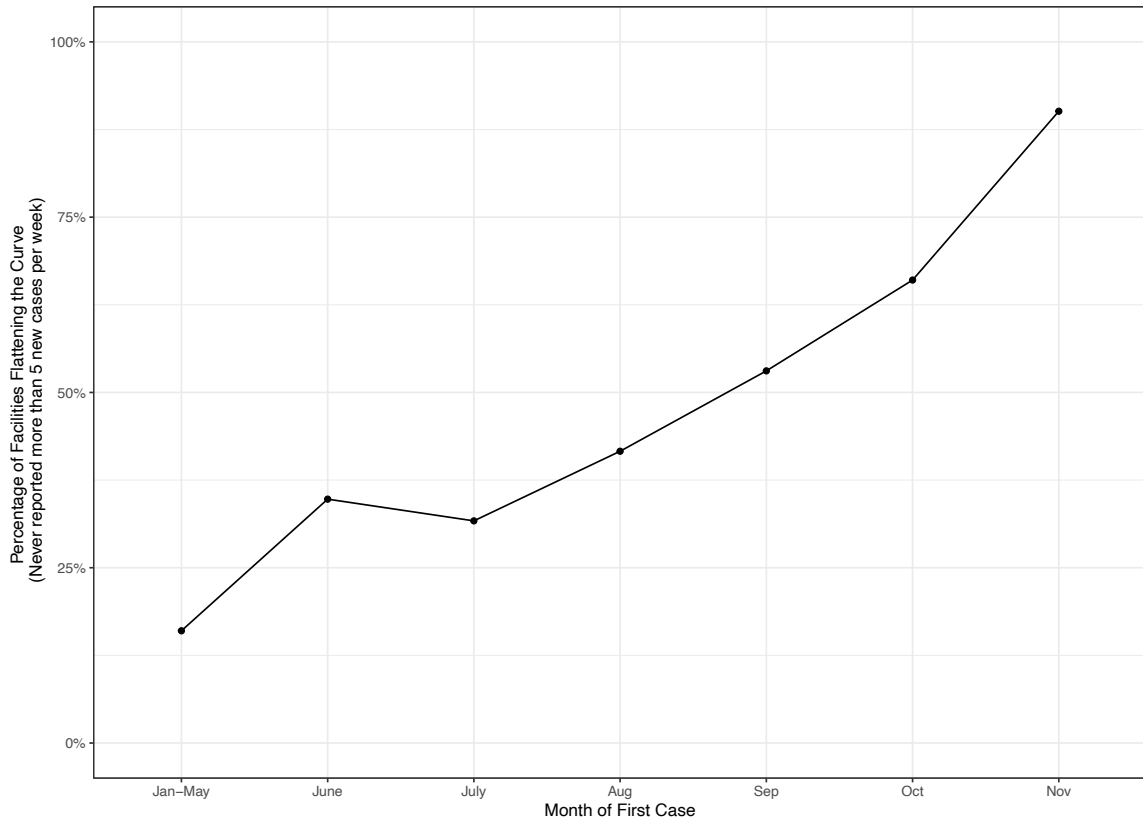
Figure 5.5 Increasing trend in care facilities flattening the outbreak curve.

In the United States, the percentage of care facilities that never reported more than 5 weekly cases (y-axis) increased over time (measured by month of first resident case reported shown on the x-axis) strongly suggesting that increased vigilance and interventions "flattened the curve" in these vulnerable populations.

Chapter 6

CONCLUSIONS

Executive Summary

Traditional, homogenous-mixing models of infectious diseases have had a

long history of successes (Anderson & May, 1991) and have continued to provide

valuable insights during our current global pandemic of COVID19 (Kissler et al.,

2020). Extensions of homogeneous mixing models that divide populations into

subcompartments based on age, risk-level, or space have also proven to be

extremely useful in understanding heterogeneity in infectious diseases incidence

(Keeling & Rohani, 2007). In particular, age-assortative contact patterns are key

to modeling respiratory transmitted infections such as measles, pertussis, and

tuberculosis (Arregui et al., 2018; Mossong et al., 2008; Rohani et al., 2010;

Schenzle, 1984). The overarching theme of this dissertation is to further

understand and quantify how social mixing patterns affect infectious disease

dynamics.

In the first chapter, I introduced social contact networks, which describe

the organization of interactions between individuals in a population, and a

number of theoretical predictions that stem from their structure. An important idea

in contact network epidemiology literature is that more central individuals will be

infected faster and more often than less central individuals (Christley et al.,

2005). In the second chapter, I studied whether recently diagnosed tuberculosis

cases were more central in their social networks than community controls.

Contrary to expectation and theory, this study found few differences in the

centrality between cases and controls. In addition, Chapter 2 revealed stark

differences in assortative mixing patterns by sex and made me wonder whether

this assortative structure may contribute to excess burden of tuberculosis among

men, the topic of Chapter 3. In this chapter, I designed a simulation study to

disentangle the effects of sex-assorted mixing and biological differences in

infection rates by sex on male-bias of tuberculosis. I showed that realistic levels

of sex-assortativity cannot generate male-bias in tuberculosis alone but

assortativity can amplify the consequences of biological sex-differences in

infection rates and lead to higher levels of male-bias compared with random

mixing networks.

Chapter 1 additionally identified a gap in the network epidemiology

literature surrounding the effects of core-periphery network structure on

epidemics which is the topic of Chapter 4. The main finding from this chapter is

that core-periphery network structure can lower the epidemic threshold for large

outbreaks to occur due to the temporary persistence of infections in the core

group. In the current pandemic of COVID19, I speculated that some high

transmission settings, such as nursing homes and correctional facilities, may

have core-periphery structure and therefore the outbreak size distribution in

these settings should match simulated distributions, unless interventions have

substantially mitigated transmission. However, I identified a mismatch between

simulations, without interventions, and real data suggesting preliminary, but

promising, evidence that interventions taken inside these facilities have stopped

large outbreaks from occurring frequently. Altogether, this dissertation uses a

multi-scale approach to understanding how social structure and contact networks

affect epidemics.

Intellectual Significance of Findings

This dissertation advances our understanding of how and when contact

patterns affect disease spread and persistence, and thus the scenarios when

traditional models may be sufficient or limited in describing disease dynamics.

First, I showed how infection status was not associated with social network

centrality in a population with endemic levels of TB. This result contradicts a main

idea in contact network epidemiology (Christley et al., 2005), and suggests the

same theory used to predict infection in newly invading infectious diseases may

not apply to diseases that are endemic. Second, I not only found strong levels of

sex-assorted contact patterns but also that TB cases were more likely to report

mixing with males than control individuals, providing further support for the idea

that males are contributing to excess transmission (Dodd et al., 2016). Third, I

described conditions that increase the ability of sex-assortative contact patterns

to amplify sex-differences in infectious diseases, testing the hypothesis

presented in a recent review (Horton et al., 2020). Fourth, I filled a gap in the

network epidemiology literature surrounding the effects of core-periphery network structure on infectious disease dynamics at the population-level, and showed how the presence of a densely connected core group can diminish the epidemic threshold required for large outbreaks to occur, related to the phenomenon in scale-free networks (Chung & Lü, 2002; Pastor-Satorras & Vespignani, 2001). Core-periphery network structure may be present in a number of emerging COVID19 hotspots and be helpful in explaining epidemiology in these settings. Lastly, I found evidence that the interventions taken in COVID19 hotspots significantly reduced outbreak sizes.

Practical implications of Findings

This dissertation also has a number of practical implications. The second chapter suggests targeted treatment or contact tracing based on network centrality may be unhelpful for endemic infectious diseases. However, case finding could be optimized by focusing on male-contacts of recently diagnosed male cases. I also illustrated how different sampling designs affect inferences about underlying network structures. Results from the third chapter beg for basic research aiming to quantify proposed sex-differences in infection rates which could lead to improved treatments and predictions of disease burden, not just for tuberculosis but for many infectious diseases that are sex-biased (Guerra-Silveira & Abad-Franch, 2013). The fourth chapter suggests that pathogens with transmission rates that are close to the epidemic threshold may benefit from contact networks that are structured into densely connected core groups and

sparsely connected periphery nodes, such as in correctional facilities and nursing

homes, when and that interventions aiming to disrupt the core connectivity in this

situation may be effective (e.g., by producing work "social bubbles" (Block et al.,

2020)). In addition, I showed how estimation of key parameters may be affected

by network structure and how having multiple estimators may allow for improved

estimation. Finally, the last chapter provides early evidence that certain

interventions have mitigated spread in care facilities in the United States which

may be helpful in maintaining vigilance in these populations as the pandemic is

far from over.

Implications for respiratory transmitted infections

Tuberculosis and COVID-19 have the shared feature of being respiratory-

transmitted infections but differ in important ways (Table 6.1). As a result of their

shared transmission modes, and because both currently lack a vaccine that

protects against contagious forms of disease, prevention largely depends on

reducing contact with infectious individuals through testing, case isolation, and

contact tracing. (More drastic prevention measures such as lock-downs have

also been taken for COVID-19). In addition, tuberculosis and COVID-19 both

result in mostly subclinical infections. However, the nature and time scale of

subclinical infection are different for these diseases. Asymptomatic, pauci-

symptomatic, and pre-symptomatic individuals testing positive for COVID-19 can

be contagious and generally have undetectable viral loads for weeks. In contrast,

it is traditionally thought that individuals with latent tuberculosis infection are not

contagious (though newer studies question this traditional dogma (Drain et al.,

2018)) and the duration of latent infection in tuberculosis is extremely variable

(lasting months to decades). A final similarity between tuberculosis and COVID-

19 is the importance of transmission within closed settings (e.g., households,

health care settings, bars, worship settings, and workplaces).

What general lessons have we learned about respiratory transmitted

diseases, variable time scales of infection, and contact patterns that this

dissertation addresses? This dissertation highlighted the nuances of social

structure and contact networks and their relative importance for respiratory

transmitted infectious diseases potentially depending on the time scale of

infection. While social structure (e.g., age- and sex-assortativity) caused by

social roles leads to stable patterns in the types of individuals one interacts with,

contact networks describe specific interactions between individuals which can be

more fluid across time. This dissertation suggests the long time-scale of

tuberculosis may cause incidence patterns to be more affected by the relatively

stable social structures than by the contact structure of recent interactions. In

contrast, for COVID-19 with a much shorter time scale of infection, recent contact

networks are much more likely to be important for understanding transmission,

especially in small populations living or working in closed settings such as

nursing homes, correctional facilities, or food-processing plants. Thus,

researchers aiming to identify the effect of contact networks on tuberculosis may

need to collect multiple years of data. In contrast, for COVID-19, more subtle

patterns in assortative mixing may matter less than recent contact networks for modeling transmission patterns.

Future research questions

  The questions and answers posed in this dissertation have led to more questions, as often happens in research. First, can we predict infection based on node position in contact networks for endemic infectious diseases? Second, when, why, and how should we incorporate both age and sex assortative patterns in models of infectious spread? Third, what are underlying contact patterns in emerging COVID19 hotspots, and are they representative of core-periphery networks? Additionally, can we use hypothesis testing to identify which idealized network structures best represent real-world networks? Fourth, how vulnerable are care, correctional, and food processing facilities to future COVID19 outbreaks? Lastly, what are the spatial and demographic factors that led to large outbreaks in COVID19 hotspots?

References

Anderson, R. M., & May, R. M. (1991). *Infectious Diseases of Humans*.

Arregui, S., Iglesias, M. J., Samper, S., Marinova, D., Martin, C., Sanz, J., & Moreno, Y. (2018). Data-driven model for the assessment ofMycobacterium tuberculosistransmission in evolving demographic structures. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(14), E3238–E3245. https://doi.org/10.1073/pnas.1720606115

Block, P., Hoffman, M., Raabe, I. J., Dowd, J. B., Rahal, C., Kashyap, R., & Mills, M. C. (2020). Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nature Human Behavior*, 1–9. https://doi.org/10.1038/s41562-020-0898-6

Christley, R. M., Pinchbeck, G. L., Bowers, R. G., Clancy, D., French, N. P., Bennett, R., & Turner, J. (2005). Infection in social networks: using network analysis to identify high-risk individuals. *American Journal of Epidemiology*, *162*(10), 1024–1031. https://doi.org/10.1093/aje/kwi308

Chung, F., & Lü, L. (2002). The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, *99*(25), 15879–15882. https://doi.org/10.1073/pnas.252631999

Dodd, P. J., Looker, C., Plumb, I. D., Bond, V., Schaap, A., Shanaube, K., Muyoyeta, M., Vynnycky, E., Godfrey-Faussett, P., Corbett, E. L., Beyers, N., Ayles, H., & White, R. G. (2016). Age- and Sex-Specific Social Contact Patterns and Incidence of Mycobacterium tuberculosis Infection. *American Journal of Epidemiology*, *183*(2), 156–166. https://doi.org/10.1093/aje/kwv160

Drain, P. K., Bajema, K. L., Dowdy, D., Dheda, K., Naidoo, K., Schumacher, S. G., ... & Sherman, D. R. (2018). Incipient and subclinical tuberculosis: a clinical review of early stages and progression of infection. *Clinical microbiology reviews*, *31*(4).

Guerra-Silveira, F., & Abad-Franch, F. (2013). Sex bias in infectious disease epidemiology: patterns and processes. *PLoS ONE, 8*(4), e62390. https://doi.org/10.1371/journal.pone.0062390

Horton, K. C., Hoey, A. L., Béraud, G., Corbett, E. L., & White, R. G. (2020). Systematic Review and Meta-Analysis of Sex Differences in Social Contact Patterns and Implications for Tuberculosis Transmission and Control. *Emerging Infectious Diseases*, *26*(5), 910–919. https://doi.org/10.3201/eid2605.190574

Keeling, M. J., & Rohani, P. (2007). *Modeling Infectious Diseases in Humans and Animals*.

Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H., & Lipsitch, M. (2020). Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*, *368*(6493), 860–868. https://doi.org/10.1126/science.abb5793

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., & Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*, *5*(3), e74. https://doi.org/10.1371/journal.pmed.0050074

Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic dynamics and endemic states in complex networks. *Physical Review E*, *63*(6 Pt 2), 066117. https://doi.org/10.1103/physreve.63.066117

Rohani, P., Zhong, X., & King, A. A. (2010). Contact Network Structure Explains the Changing Epidemiology of Pertussis. *Science*, *330*(6006), 982–985. https://doi.org/10.1126/science.1194134

Schenzle, D. (1984). An age-structured model of pre- and post-vaccination measles transmission. *IMA Journal of Mathematics Applied in Medicine and Biology*, *1*(2), 169–191. https://doi.org/10.1093/imammb/1.2.169

Tables and Figures

Table 6.1. Major similarities and differences in the epidemiology of tuberculosis and COVID-19.

	Tuberculosis		COVID-19
Transmission mode	Respiratory (droplet, airborne)		
Global Prevalence	10 million infections, 1.6 million deaths (2019)	Most infections subclinical	~50 million infections, 1.2 million deaths (2020)
Time scale	Months-Decades		Days-Weeks
Reproductive ratio	5-15 (0.5-1.5)		2-4
Diagnosis	Non-Specific Tests		Limited at first
Treatment	Low treatment adherence		No "cure"
Prevention	Low vaccine efficacy		No vaccine
Control	Passive case-finding, contact tracing		
Risk groups	Co-infection, smokers, males		Older age, inability to social distance

APPENDIX I

Supplement: Association between male-bias in tuberculosis cases and social
network structure in Kampala, Uganda

Sensitivity analyses

To examine potential biases on measures of node position and estimates
of assortativity in second-level egocentric sampling, we performed two sensitivity
analyses. First, we assessed the reliability of node position estimates from
egocentric samples. To account for uncertainty in structure and size of the "true"
social network in Kampala, we analyzed two types of networks proposed to
resemble real-world social networks (small-world, SW, and scale-free, SF)
across five different network sizes (N $= 5 \cdot 10^4, 7.5 \cdot 10^4, 1 \cdot 10^5, 1.25 \cdot 10^5, 1.5 \cdot$
10^5). We generated small-world networks with the Watts-Strogatz algorithm [1],
starting connectivity of 5, and edge-rewiring probability of 0.05. We generated
scale-free networks with the Barabasi-Albert algorithm [2] with linear attachment
and 5 edges added in each step. Each combination of network type and size
were replicated 15 times and sampled in three different ways: ego-only, first-
level, and second-level egocentric sampling. We then calculated the correlation
between true node centrality and estimated node centrality for the ego nodes
(n=240).

To understand how the Kampala network relates to synthetic networks, we calculated the clustering coefficient (the probability that neighbors of a node are themselves connected [3]) and fit to a power-law degree distribution because small-world networks are characterized by high clustering coefficients and scale-free networks by a power-law degree distributions [4]. The power law degree distribution was fit to the Kampala data and synthetic networks using the method of maximum likelihood and the goodness-of-fit tests [5]. We additionally evaluated whether other distributions fit the data equally using the Vuong test statistic (approach detailed in [5]; implemented in the R package, poweRlaw [6]).

In a second sensitivity analysis on estimates of assortativity in egocentric samples, we generated 15 replicates of synthetic networks across a spectrum of network assortativity (r$\approx 0, 0.2, 0.4, 0.6, 0.8$) and two different network sizes ($N = 1 \cdot 10^3, 1 \cdot 10^4$). We generated networks with varying assortativity by first assigning each node to one of two groups (i.e., sex). Then, we set the number of within- and between-group edges per node from Poisson distributions (similar to algorithm in [7]). The mean degree in synthetic networks was 10. For example, in one combination, each node had an average of 5 edges within-group and 5 edges between-group resulting in an overall network assortativity value of approximately 0. Following network simulations, we sampled each network using second-level egocentric sampling type with 240 randomly egos. Finally, we calculated the correlation between sampled assortativity and true assortativity of

the underlying network to understand the effects of egocentric sampling on estimates of network assortativity.

Observed Kampala network fit to power-law degree distribution

The degree distribution of the Kampala network could plausibly follow a power-law (Kampala data fell within bootstraps from the best-fit power-law distribution, p=0.25). However, the log-normal distribution could also not be ruled out for the tail of the degree distribution (likelihood-ratio-test showed that neither power-law nor log-normal was preferred, p=0.56). Recommended approach from Clauset et al. (2009) [5].
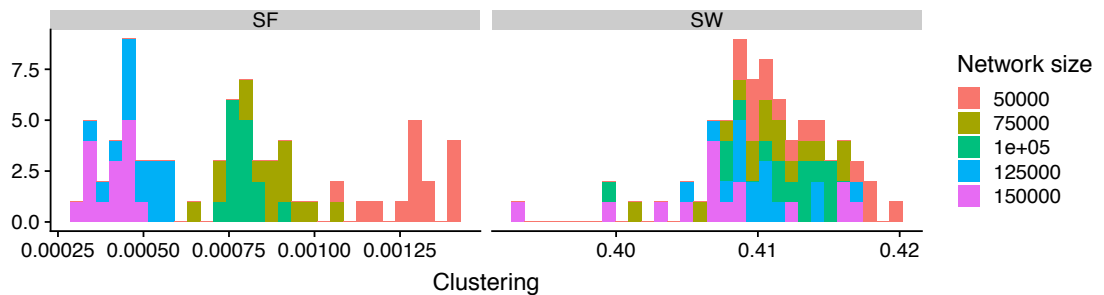
Figure S.1.1 Clustering coefficients in simulated networks.

Clustering coefficients from simulated networks of varying sizes that were sampled using a second-level egocentric design. We used these distributions to understand whether the underlying network in Kampala more closely resembles scale-free (SF) or small-world (SW) structure. The clustering coefficient of the Kampala network was 0.1, which was in between estimates from scale-free and small-world networks.
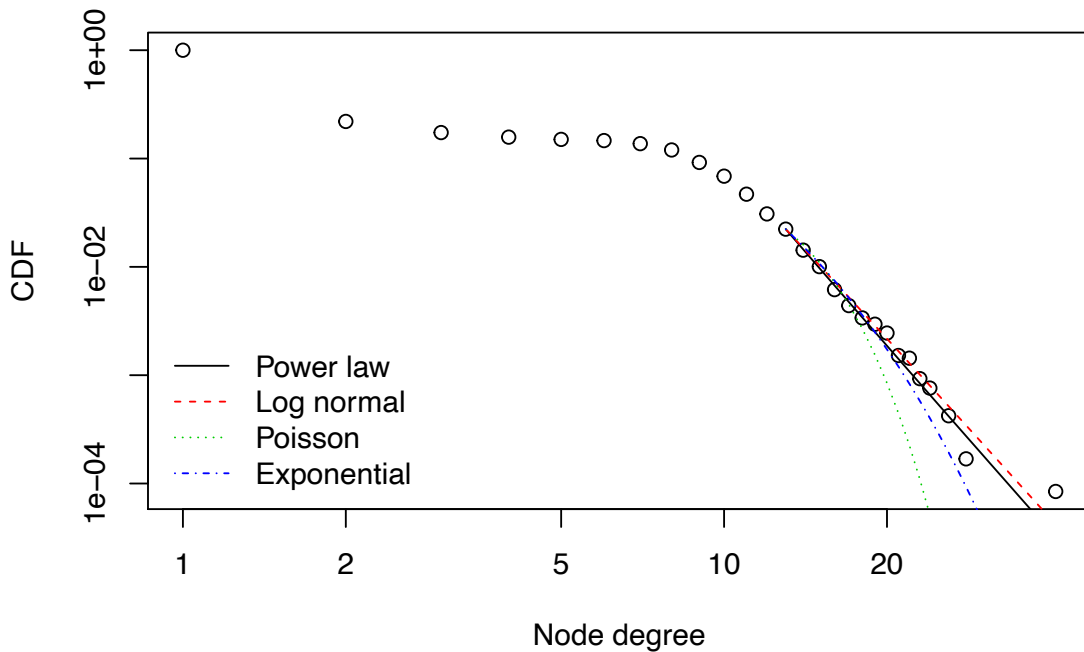
Figure S.1.2 Cumulative distribution function of Kampala network.

The cumulative distribution function of node degree in the Kampala network can be explained by power-law and log-normal distributions but not exponential or Poisson distributions. The tail of the distribution was estimated to begin at 13 and values larger than 13 were used to fit the distributions (n=264). A fit to a power-law distribution is an indication of scale-free networks but log-normal could not be ruled out using the methods detailed in Clauset et al. (2009) [5].
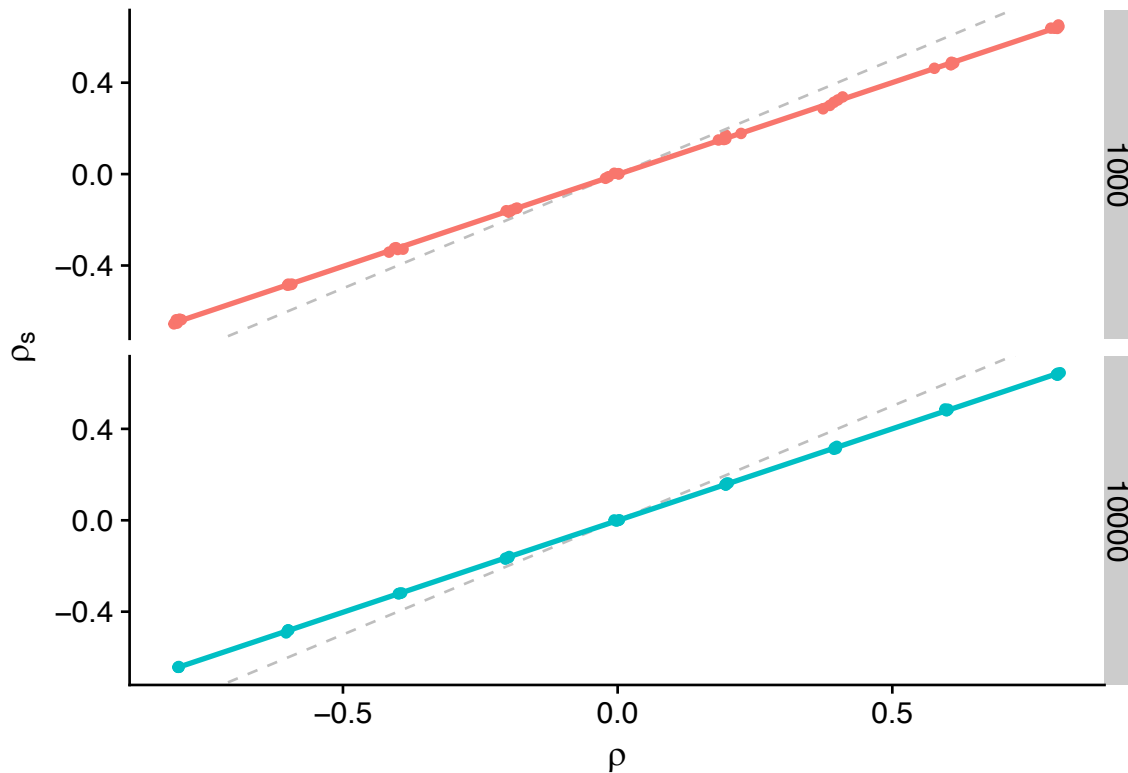
Figure S.1.3 Correlation between underlying assortativity coefficients and sampled assortativity coefficients in simulated networks.

Estimated assortativity (ρ_s) is highly correlated with underlying network assortativity (ρ) in second-level egocentric samples from networks of size 1,000 and 10,000. When there is preferential assortative mixing within group (i.e., positive assortativity coefficients), estimated assortativity in second-level samples tend to underestimate the true amount of assortativity. The dashed grey line indicates the 1-1 line. Assortativity calculated with Newman's assortativity coefficient [8].

Table S.1.1 Social Network Estimates.

Social Network Estimates for Index Individuals stratified by Relationship Status and Sex. Values indicate the number of individuals (proportion) or mean (± standard errors) for each variable

	Single		In a relationship		Sig.
	Female n=44	**Male** n=93	**Female** n=34	**Male** n=76	
Node position					
Degree	10.6 (±0.58)	10.4 (±0.41)	9.6 (±0.48)	10.8 (±0.53)	
Closeness	0.078 (±0.002)	0.075 (±0.002	0.074 (±0.003)	0.078 (±0.001	
Betweenness	0.012 (±0.003)	0.014 (±0.003	0.010 (±0.002	0.019 (±0.002	
Distance to case	3.5 (±0.4)	3.2 (±0.2)	3.9 (±0.3)	3.2 (±0.2)	
Mixing variables					
Proportion of all contacts with adult men	0.27 (±0.03)	0.49 (±0.03)	0.29 (±0.04)	0.46 (±0.03)	∗
Proportion of all contacts with adult women	0.42 (±0.03)	0.38 (±0.02)	0.43 (±0.04)	0.36 (±0.02)	∗
[1]Proportion of all contacts with children	0.31 (±0.03)	0.13 (±0.02)	0.28 (±0.03)	0.18 (±0.02)	∗
[2]Proportion of all contacts occurring within HH	0.33 (±0.04)	0.25 (±0.02)	0.33 (±0.03)	0.30 (±0.02)	∗
Proportion of HH contacts occurring with children	0.34 (±0.03)	0.28 (±0.02)	0.32 (±0.03)	0.28 (±0.02)	∗

∗ Significant difference ($p < 0.05$) between means by index sex (male, female)
◊ Significant difference ($p < 0.05$) between means by relationship status (single, monogamous/polygamous)

References

1.   Watts DJ, Strogatz SH. Collective dynamics of "small-world" networks. Nature. Nature Publishing Group; 1998;393: 440–442. doi:10.1038/30918

2.   Barabasi A, Albert R. Emergence of scaling in random networks. Science. 1999;286: 509–512.

3.   Wasserman S, Faust K. Social network analysis: Methods and Applications. Cambridge University Press. 1994;8.

4.   Kiss IZ, Miller JC, Springer PSC. Mathematics of epidemics on networks. Springer. 2017.

5.   Clauset A, Shalizi CR, Newman MEJ. Power-Law Distributions in Empirical Data. SIAM Review. Society for Industrial and Applied Mathematics; 2009;51: 661–703. doi:10.1137/070710111

6.   Gillespie CS. Fitting Heavy Tailed Distributions: The poweRlawPackage. Journal of Statistical Software. 2015;64: 1–16. doi:10.18637/jss.v064.i02

7.   Girvan M, Newman MEJ. Community structure in social and biological networks. PNAS. National Academy of Sciences; 2002;99: 7821–7826. doi:10.1073/pnas.122653799

8.   Newman ME. Mixing patterns in networks. Physical review E. 2003 Feb 27;67(2):026126

APPENDIX II

Supplement: Effects of assortative mixing and sex-traits on male-bias in

tuberculosis: A modelling study

Description of the rewiring algorithm

Rewired networks were initialized as small-world or scale-free networks

because these networks resemble patterns of clustering and high-degree

network hubs found in real-world social networks. Small world networks were

initialized as a ring with edges to 5 nearest neighbors being rewired with

probability 0.05 according to the Watts Strogatz algorithm (Watts and Strogatz

1998). Scale free networks were initialized with linear preferential attachment and

5 edges were added each step according to the Barabasi Albert algorithm

(Barabasi and Albert 1999).

Small world and scale free networks were then rewired to generate sex-assorted

networks using Algorithm 1.

Algorithm 1. Sex-assorted network rewiring algorithm.

1. Simulate a network, G, of size N with nodes notated as $v_{1,..N}$
2. Randomly assign sex to each vertex v_i
3. Randomly rewire 20% of edges occurring between-sex
4. Check that the network is still a single component, if not, reject rewiring and
   return to step 3
5. Check for multiple edges or self-edges, and randomly rewire those edges

6. Continue process until the desired level of assortativity was reached within a small range of error ($\epsilon = 0.05$)

Description of how disease was simulated on networks

Here, I provide a description of the Gillespie algorithm used to simulate stochastic models of disease spread used in Chapters 3-5.

The implementation of the Gillespie algorithm with the "Epidemics On Networks" Python module (Miller & Ting, 2020) calculates the time when the next event occurs and then determines what that event will be. The iterative steps for the algorithm, which were simulated using the Gillespie_simple_contagion function, are found in Kiss, Miller & Simon (2017) and listed below for a *SIR model*.

Algorithm 2. Iterative steps for the Gillespie algorithm used to simulate disease spread with a SIR model on networks.

Input: Network G (with susceptible, infected and recovered nodes), per-edge transmission rate τ, and recovery rate γ.

1. Count the number of individuals in each compartment, S, I, and R
2. Calculate the total rate (T) of all possible events that may occur. For an SIR model, T is the sum of the rates each infected node recovers, $\gamma \times R$ and the sum of the rate each susceptible neighbor is infected. The rate each susceptible node is infected is $\tau \times number\ of\ infected\ neighbors$.
3. Calculate the waiting time, w, until the next event occurs based on an exponential distribution with mean equal to the total rate of all possible events, $w \sim \exp(T)$. Update time step by adding w to current time.
4. Determine what the next event will be. Nodes will recover based on a random draw from a uniform distribution $r \sim \text{unif}(0, T)$. If $r < \gamma \times R/T$, a random infected node is chosen to recover. Add this node to list R and remove from list I. Else, a susceptible node will get infected with

probability proportional to $\tau \times number\ of\ infected\ neighbors$ and it will be removed from list S and added to list I.

5. Repeat with updated node statuses until there are no more infected nodes.

Further modifications made for the models in Chapters 3-5 are as follows: for models with a latent period (SLIR and SLIRS) the total rate of change also included events based on disease progression to active infection (L→I transition). For models with events based on reversion to susceptibility (i.e., SIRS and SLIRS models), the total rate also includes the R→S transition.

In Chapter 3, the following modifications were also made. Node sex was modeled in addition to disease status ($S_f, S_m, I_f, I_m, R_f, R_m$). In the models varying sex-specific susceptibility and transmissibility, rates of infection by an infected (source) node to a susceptible (target) neighbor depended on the sex of the source and target nodes. Thus, additional infection rates were calculated in Step 2 of Algorithm 2. Specifically, these are the sum of the rate of male nodes infecting male nodes, female nodes infecting female nodes, female nodes infecting male nodes and male nodes infecting female nodes. In the model varying sex-specific infectious periods, the rate of recovery depends on the infected node's sex and so the sum of the recovery rates in Step 2 of Algorithm 2 is the sum of male and female infected nodes recovering. How sex-specific infection and recovery rates were calculated is discussed in the main text (Chapter 3, *Disease Model*). The algorithm in Chapter 3 was terminated once

there were no longer any infected individuals (SIR, SLIR) or after 300 time steps
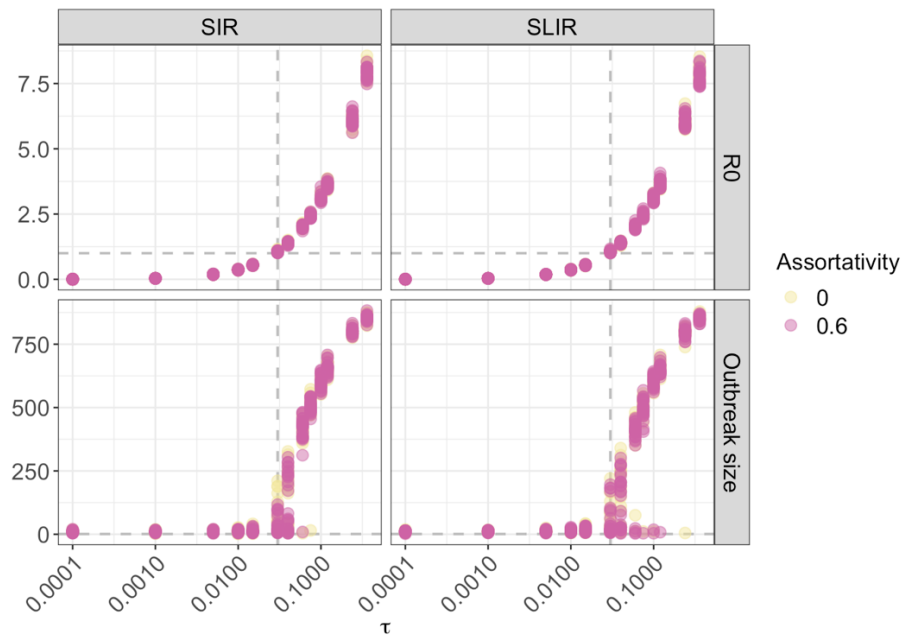
(SIRS and SLIRS).

Figure S.2.1 Reproductive ratio and outbreak size across transmission rates.

Comparison between analytically calculated R_0 (Kiss, Miller, Simon 2017) and simulated epidemic size on non-assorted and assorted networks. Horizontal grey line shows where $R_0 = 1$ and where ending epidemic size $> I_0$. Vertical grey line approximately where $R_0 = 1$. Results are consistent for SIR and SLIR models. No sex-trait heterogeneity is included in these simulated data. Other parameters: $\gamma = 0.5$, $I_0 = 10$. For the SLIR model, $\sigma = 0.25$.
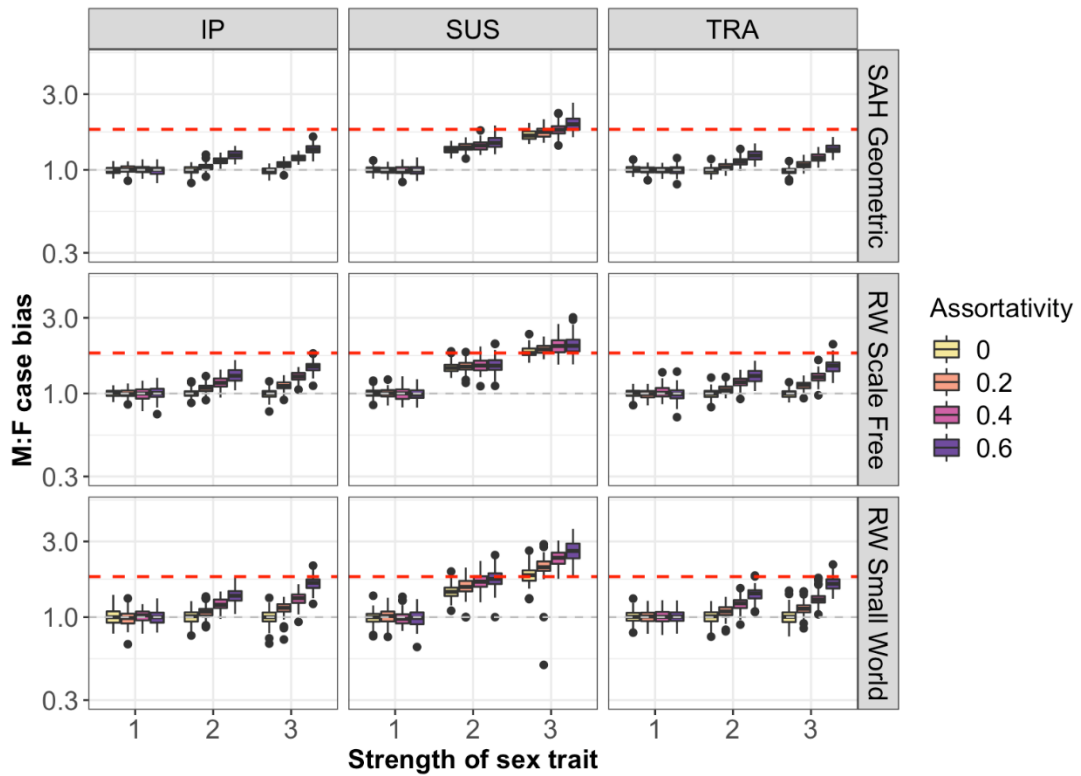
Figure S.2.2 Sensitivity of male-bias to network algorithm.

Assortativity had similar effects on male-bias in Sah and rewired networks despite rewired networks being variable in other network statistics as assortativity is increased (shown in Figure S1). The M:F case bias, sex-traits, and relative male:female values of sex-traits are the same as in Figure 1. Figure generated with 250 SLIR simulations of epidemics with $\tau = 0.075$ and $R_0 \approx 2.5$.
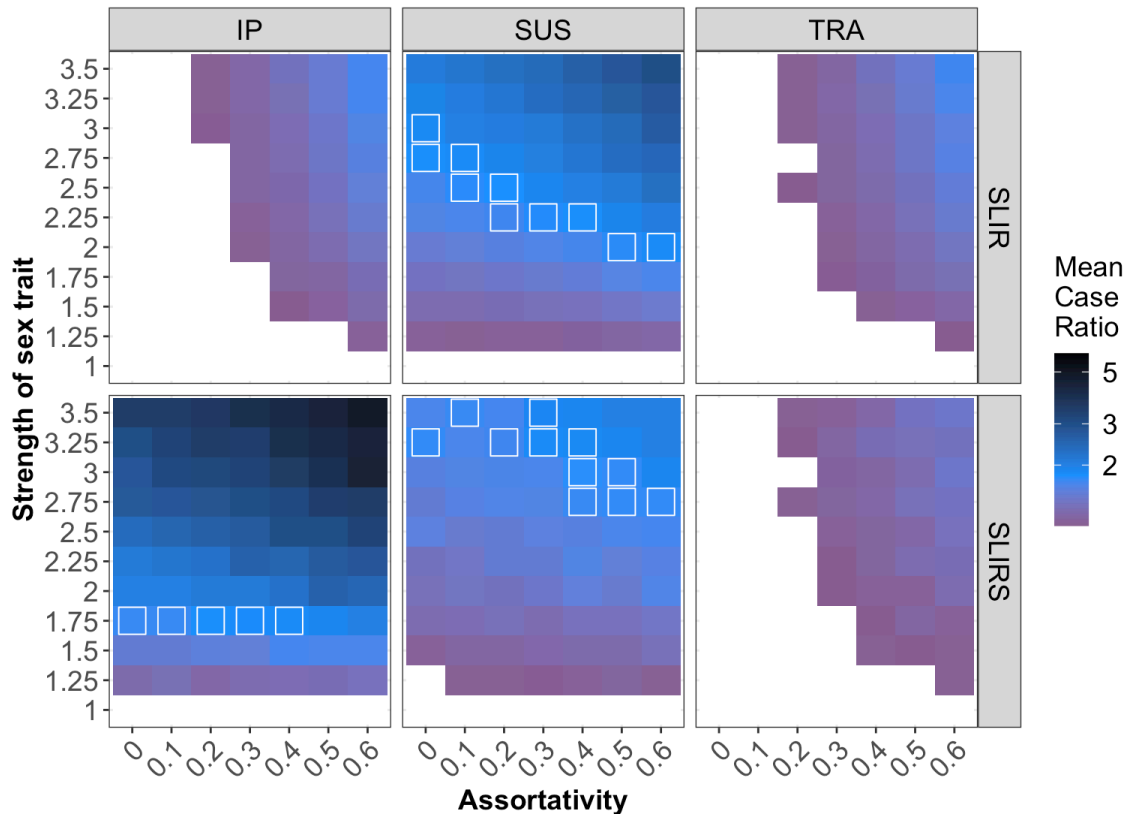
Figure S.2.3 Sensitivity of male-bias to models with latent period.

Same as Figure 2.1 in main paper but shown for SLIR and SLIRS. Interaction effects of sex-assortativity and sex-traits on M:F case bias (shown in color). Sex-traits (vertical columns) are infectious period (IP), susceptibility (SUS), and transmissibility (TRA). M:F case bias is measured as the ratio of male to female recoveries (SLIR) or infections at equilibrium (SLIRS). Only parameter combinations leading to mean M:F case bias greater than 1.1 are colored (white boxes show mean M:F case bias from 1.7 to 1.9). Sex-traits are incorporated by holding respective overall parameter rates constant but increasing the male parameter by the value on the x-axis relative to the female trait. Figure generated with 250 simulations of epidemics on Sah networks with $\tau = 0.075, R_0 \approx 2.5$.
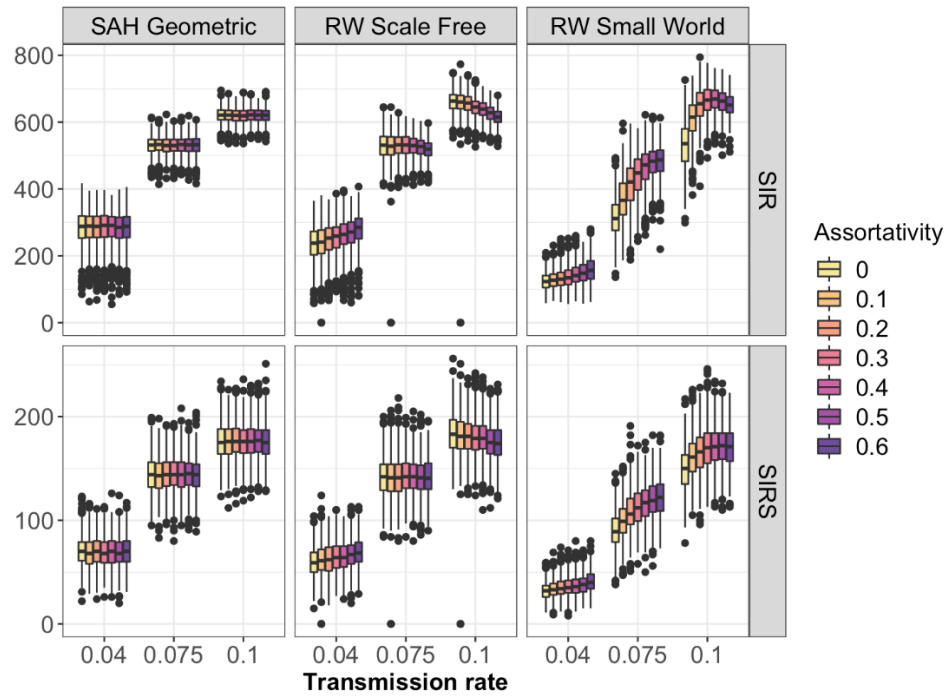
Figure S.2.4 Effect of assortativity on outbreak size and prevalence.

The effect of assortativity on the final size or equilibrium prevalence of SIR and SIRS epidemics depends on underlying network type. Results are shown for outbreaks with no differences in male and female sex-traits. Figure generated with 250 simulations of epidemics.
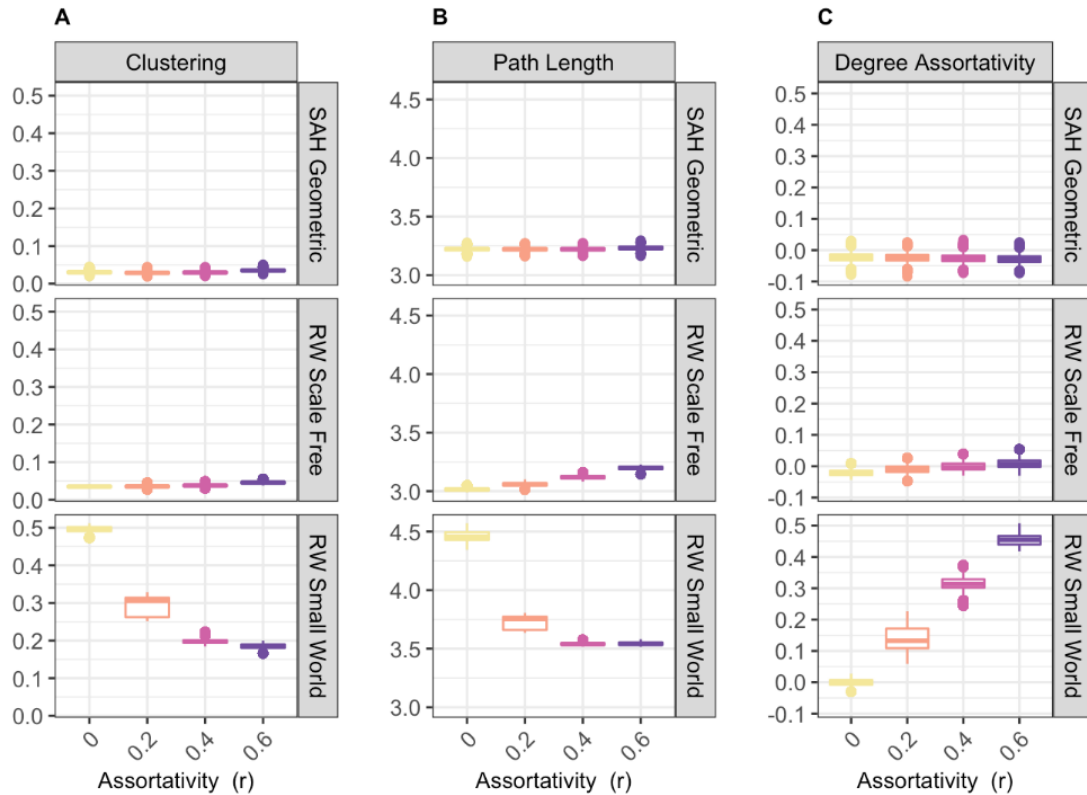
Figure S.2.5 Network statistics across assortativity algorithms.

We simulated TB spread on networks with varying assortativity generated by rewiring edges or with an algorithm developed by Sah et al. (2017). As assortativity increased, we measured changes in clustering (A), average path length (B), and degree assortativity (C) which can impact epidemic dynamics. The rewired small-world networks were especially vulnerable to disruptions in key network structures. Sah network structures were not affected by increases in assortativity.

References

Barabasi, AL, and R Albert. 1999. Emergence of Scaling in Random Networks. *Science* 286 (5439): 509–12.

Kiss, I. Z., Miller, J. C., & Simon, P. L. 2017. Mathematics of epidemics on networks (Vol. 46). Springer.

Miller, J. C., & Ting, T. 2020. EoN (Epidemics on Networks): a fast, flexible Python package for simulation, analytic approximation, and analysis of epidemics on networks. *Journal of Open Source Software*, *4*(44), 1731. https://doi.org/10.21105/joss.01731

Watts, D J, and S H Strogatz. 1998. Collective Dynamics of 'Small-World' Networks. *Nature* 393 (6684). Nature Publishing Group: 440–42. doi:10.1038/30918.

APPENDIX III

Supplement: The Effects of Core-Periphery Network Structure on Disease

Spread in Isolated Populations
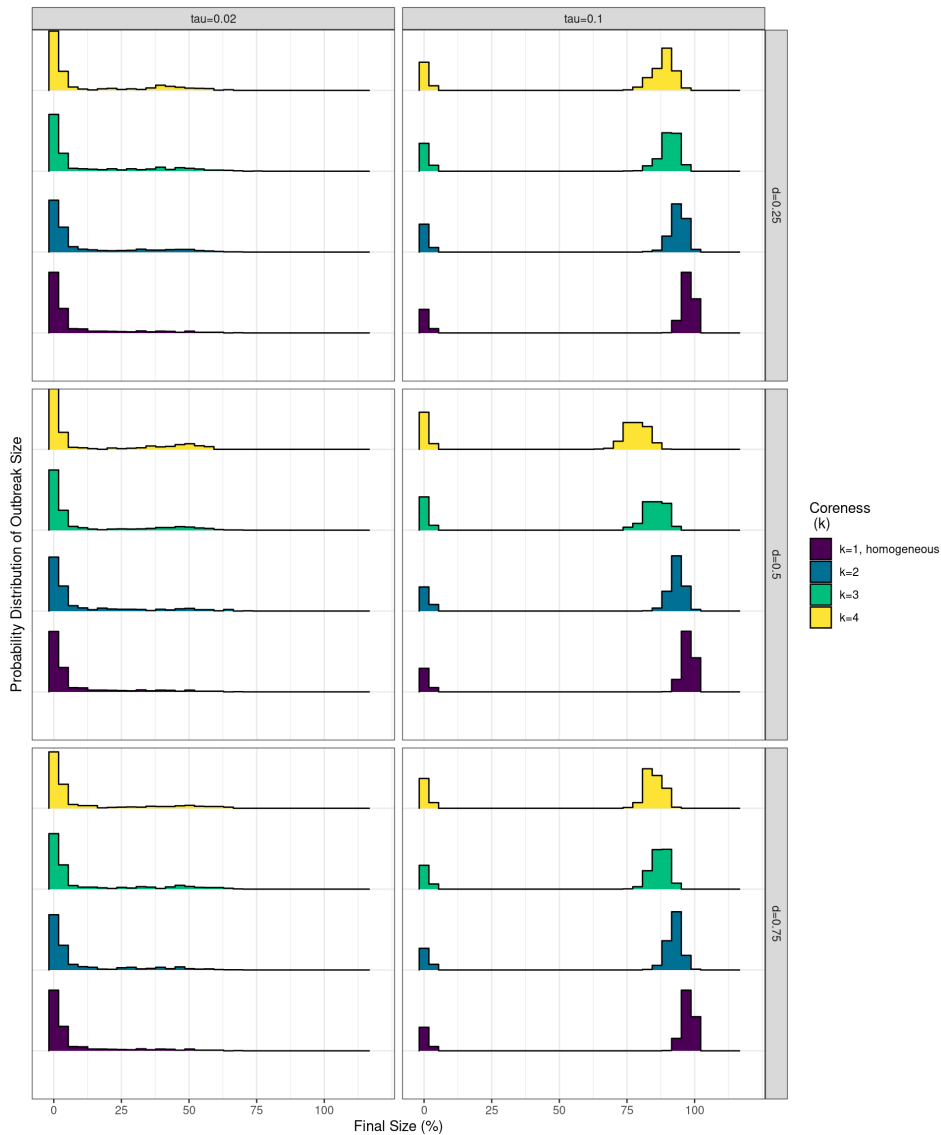
Figure S.3.1 Effects of core size on outbreak size.

The size of core relative to the periphery (d) affects the final size of simulated outbreaks. Coreness (k) is shown in the color of histogram distribution and the final size (percent of the population infected at the end of the simulation) is shown along the x-axis. Columns are faceted by transmission rate (τ) and rows are faceted by the relative size of the core (d). Plot shows results for networks with 100 nodes and 500 simulations of each parameter combination.
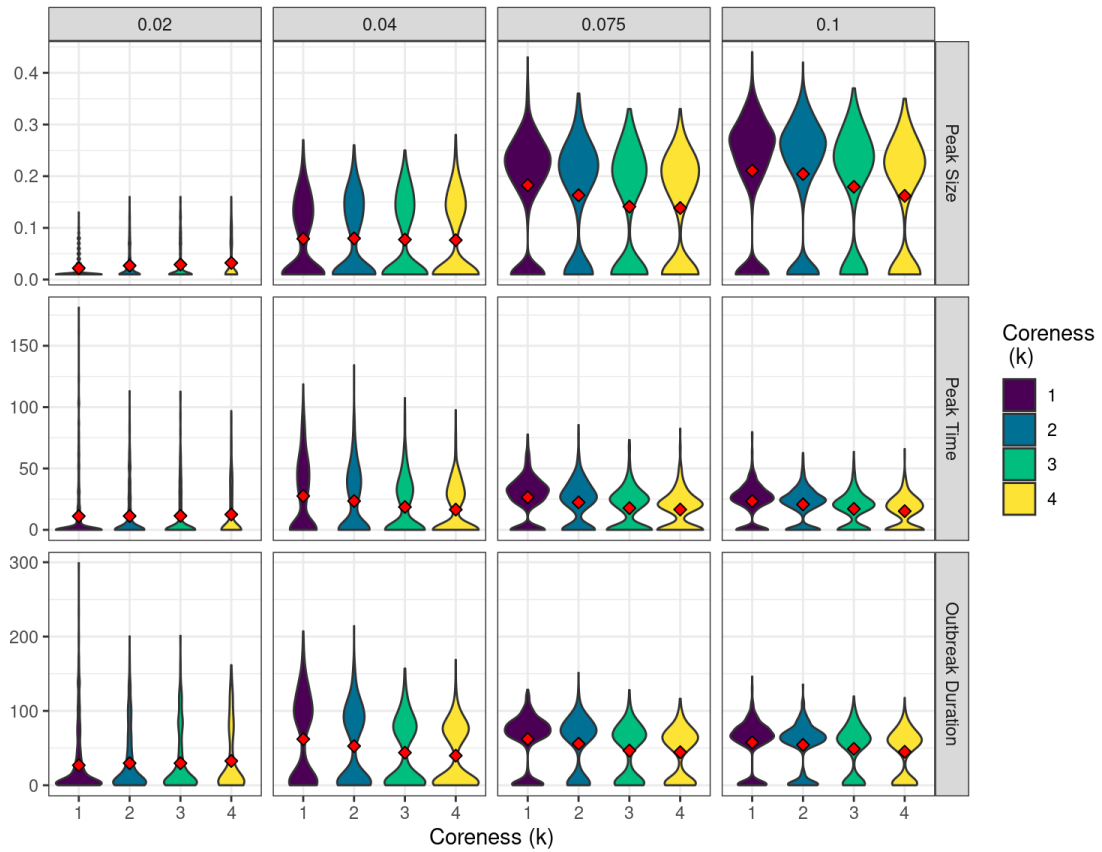
164
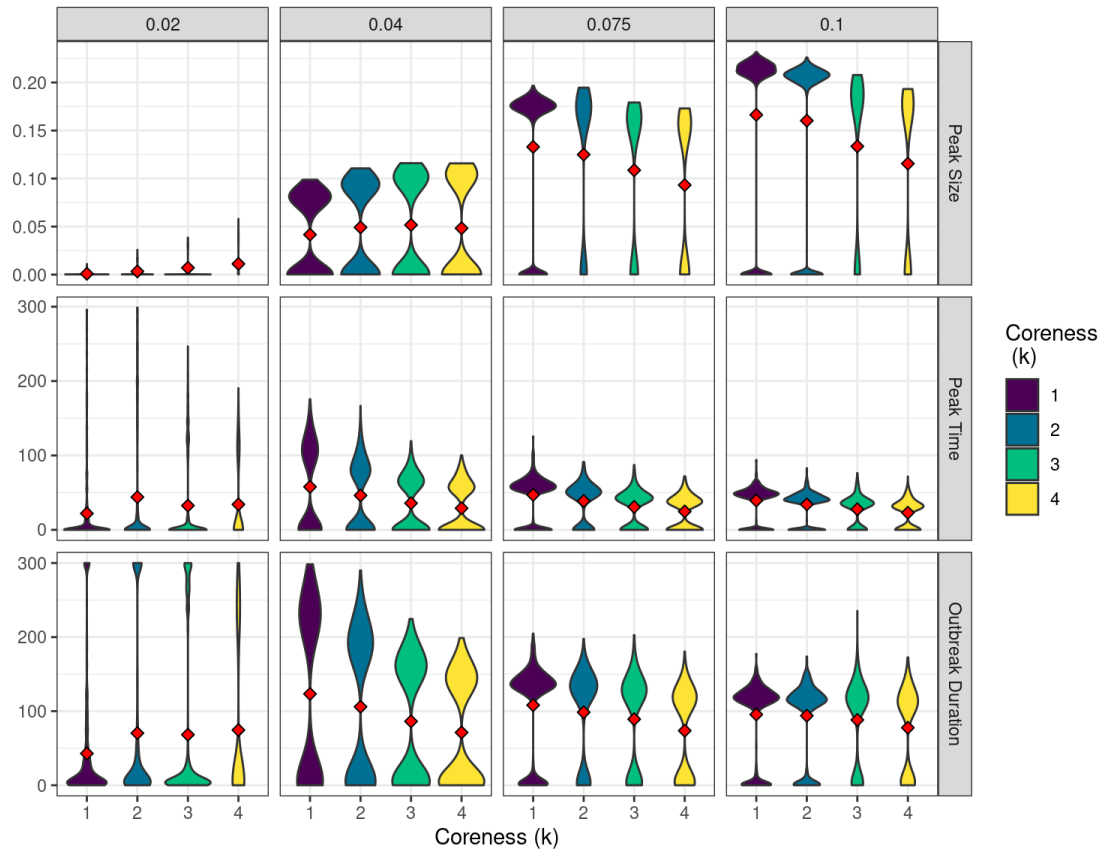
Figure S.3.2 Effects of coreness on epidemic dynamics (N=5000).

Relative changes in the peak size, peak time, and outbreak duration as k increased were larger when the transmission rate was lower compared with higher transmission rates. Coreness (k) is shown in the color of violin and the epidemic statistics are shown along the y-axis. Columns are faceted by transmission rate (τ) and rows are faceted by epidemic statistic. Plot shows results for networks with 5,000 nodes and 500 simulations of each parameter combination.
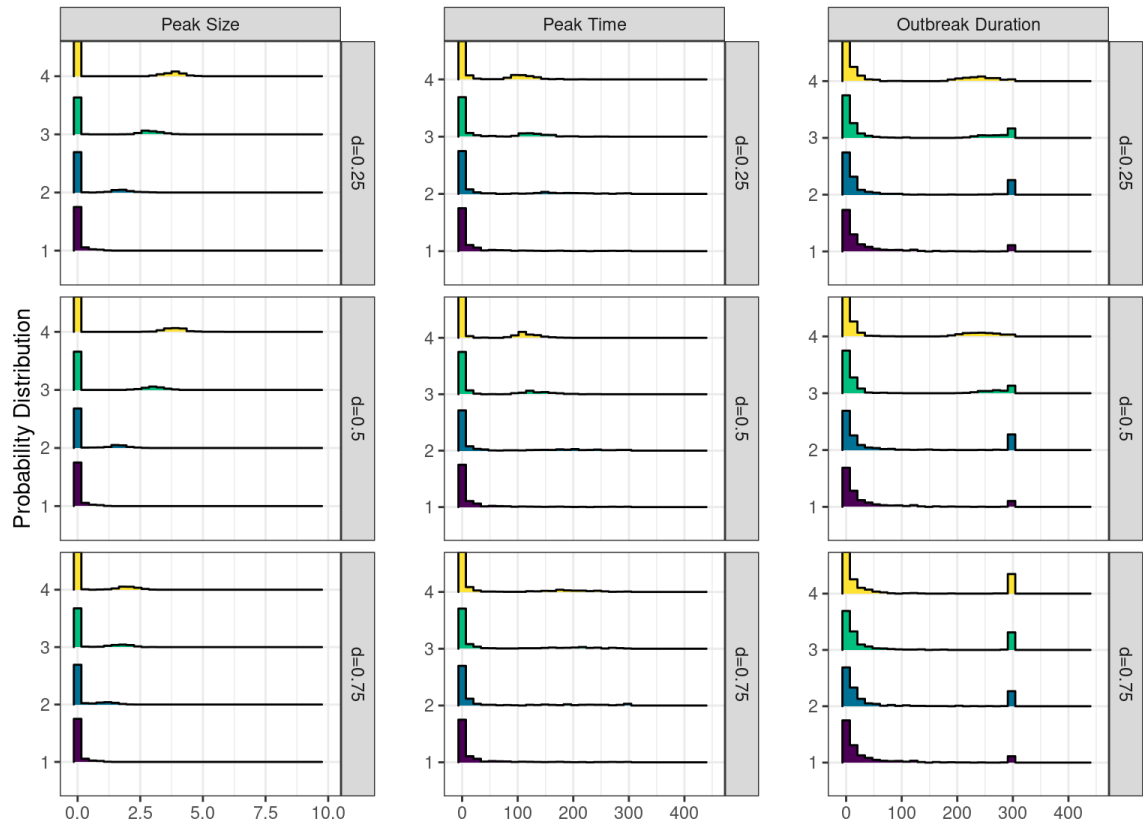
Figure S.3.3 Effects of coreness on epidemic dynamics (N=100)

Relative changes in the peak size, peak time, and outbreak duration as k increased were larger when the transmission rate was lower compared with higher transmission rates. Coreness (k) is shown in the color of violin and the epidemic statistics are shown along the y-axis. Columns are faceted by transmission rate (τ) and rows are faceted by epidemic statistic. Plot shows results for networks with 100 nodes and 500 simulations of each parameter combination.

Figure S.3.4 Effects of core size on epidemic dynamics.

The size of the core (d) affected trends in peak size, peak time, and outbreak duration. Coreness (k) is shown in the color of probability distribution and columns show epidemic statistics. Rows are faceted by the size of the core relative to periphery. Plot shows results for networks with 5,000 nodes, $\tau = 0.02$ and 500 simulations of each parameter combination.
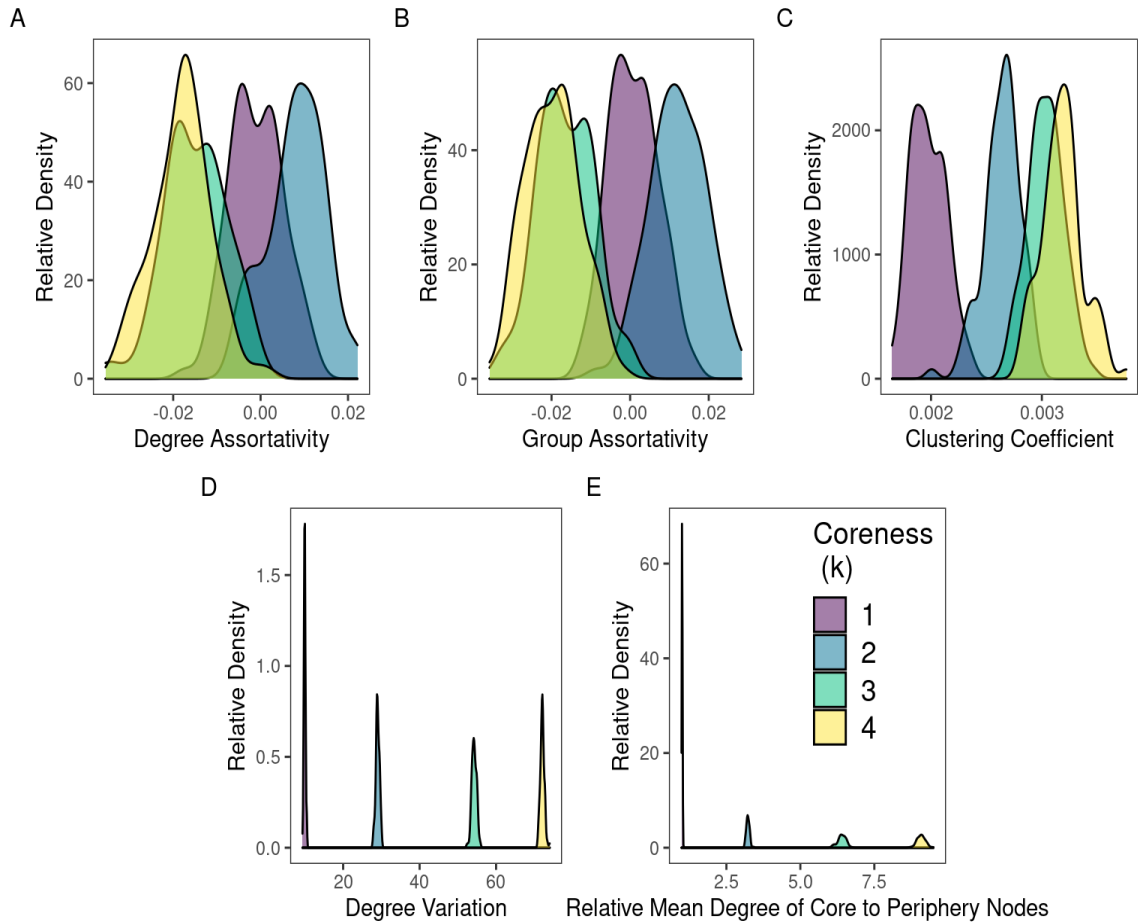
Figure S.3.5 Characteristics of core-periphery networks.

Characteristics of core-periphery networks. Core-periphery networks are neither degree-assorted (A) nor group-assorted (B) and changes in clustering are minimal (C) as coreness, k, increases. Distinguishing features of core-periphery networks are large variation in node degree (D) and high relative mean degree of core and periphery nodes (E). Coreness (k) is shown in color. Plot shows the density distribution of estimated statistics from networks with 5000 nodes and evenly sized core and periphery groups (d=0.5).

APPENDIX IV

Supplement: Characteristics of COVID-19 Outbreaks in Care, Correctional, and

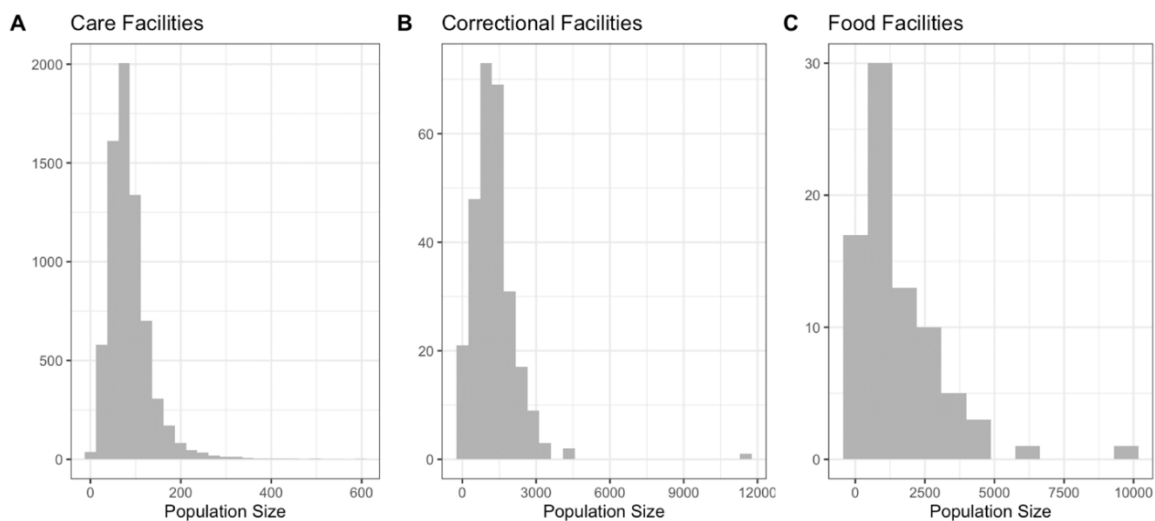Food Processing Facilities in the United States

Figure S.4.1 Population size in facilities with COVID19 outbreaks.

Population size variables in care (A), correctional (B), and food processing (C)
plants. Population size in care, correctional, and food facilities was based on
residents, incarcerated individuals, and employees per facility and estimated
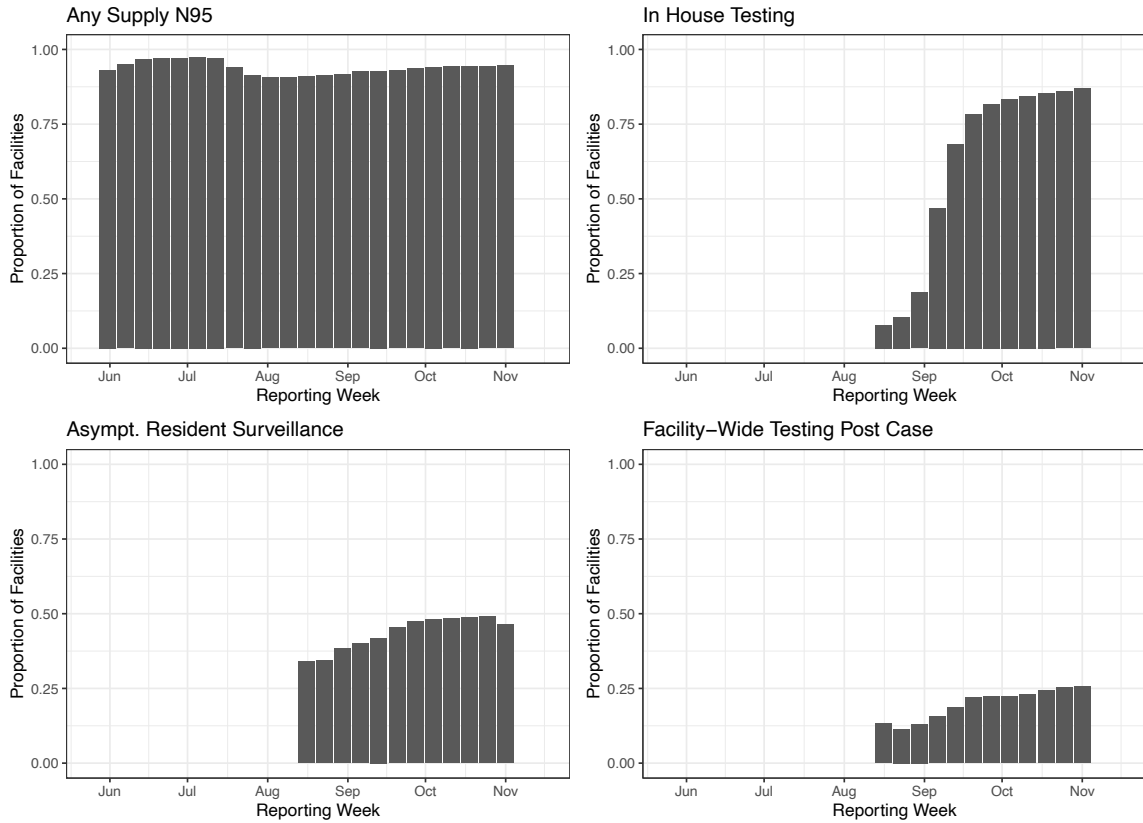from CDC NHSN or publicly available information.

Figure S.4.2 Proportion of care facilities reporting access to four interventions over time.

The proportion of care facilities that reported any current supply of N95 masks, access to in-house testing, surveillance testing asymptomatic residents, and testing residents facility-wide after a case is found varies over time. In particular, most facilities consistently reported having a supply of N95 masks and other intervention variables were only reported after August 16, 2020.