

INTEGRATION OF SOCIAL NETWORK ANALYSIS AND MOLECULAR  
EPIDEMIOLOGY: AN INVESTIGATIVE ANALYSIS OF A MANITOBA  
NETWORK IN THE STUDY OF *CHLAMYDIA TRACHOMATIS*  
TRANSMISSION

By Teresa Cabral

A thesis

Submitted to the Faculty of Graduate Studies in  
Partial Fulfillment of the requirements  
for the Degree of

MASTER OF SCIENCE

Department of Medical Microbiology  
Faculty of Medicine  
University of Manitoba  
Winnipeg, Manitoba



National Library  
of Canada

Acquisitions and  
Bibliographic Services

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque nationale  
du Canada

Acquisitions et  
services bibliographiques

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file Votre référence*

*Our file Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-76905-4

Canada

**THE UNIVERSITY OF MANITOBA  
FACULTY OF GRADUATE STUDIES  
\*\*\*\*\*  
COPYRIGHT PERMISSION PAGE**

**Integration of Social Network Analysis and Molecular Epidemiology: An Investigative  
Analysis of a Manitoba Network in the Study of *Chlamydia trachomatis* Transmission**

**BY**

**Teresa Cabral**

**A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University  
of Manitoba in partial fulfillment of the requirements of the degree  
of**

**MASTER OF SCIENCE**

**TERESA CABRAL ©2002**

**Permission has been granted to the Library of The University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to University Microfilm Inc. to publish an abstract of this thesis/practicum.**

**The author reserves other publication rights, and neither this thesis/practicum nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.**

## ACKNOWLEDGEMENTS

I wish to thank my advisor, Dr. John Wylie, for his continual help and guidance in this project. I would also like to thank my thesis advisory committee, Dr. Amin Kabani, Dr. Magdy Dawood and Dr. Lawrence Elliot for their support, assistance and encouragement in this research.

I also wish to thank the people at Cadham Provincial Lab for their assistance in the lab, as well as those individuals for their emotional support throughout the years. You know who you are.

Last and definitely not least, I would like to thank my significant other and my family whose continual support, patience and understanding helped me to pursue my life-long goals.

This thesis would not have been possible without the data from the Communicable Disease Control Unit of Manitoba Health, and studentship funding from the University of Manitoba, and the Manitoba Health Research Council. I also would like to thank the physicians and public health nurses throughout Manitoba who provided the information on which these data are based.

## ABSTRACT

*Chlamydia trachomatis* is the most frequently occurring bacterial sexually transmitted disease in Canada and worldwide. The burden and cost of the disease is primarily due to the sequelae which include pelvic inflammatory disease, infertility and ectopic pregnancy. Preventive programs have been implemented and have shown some success in reducing the number of cases over the last several decades. Unfortunately, the decrease in disease prevalence is no longer evident. In order to reduce and possibly eliminate chlamydia in Canada, and worldwide, new control strategies need to be implemented.

Recently, an emerging concept for studying STD transmission is that of social network analysis. The use of social network analysis has the potential to further our understanding of STD epidemics and contribute to the development of more effective targeted control strategies. The main premise of this analysis is to study the epidemic at a population level as opposed to focusing on individuals alone. In addition, the incorporation of molecular epidemiology tests the ability of constructing networks from contact tracing data.

In this research, we analyzed , sexual networks formed in Manitoba over a six month period. Routinely collected case/contact information was gathered by public health nurses and used to construct the sexual network. Specimens from infected individuals were typed by DNA sequence analysis to assess the accuracy of transmission events identified through contact tracing information. In addition, the characteristics of the chlamydia - infected population and the sexual partnering behaviors of those individuals was

determined. The results of these analyses will assist in the development of effective prevention programs to reduce chlamydial infections in Manitoba.

## TABLE OF CONTENTS

<b>Chapter 1</b>	<b>Introduction</b>	<b>Page Number</b>
1.1	Burden of disease	1
1.2	Epidemiology of <i>Chlamydia trachomatis</i>	5
1.3	Beyond the individual : core groups and the reproductive number	14
1.3.1	Core groups in Manitoba	21
1.4	Social Networks and STD epidemiology	22
1.4.1	Social network concepts	27
1.4.2	Structure of networks	31
1.4.3	Temporal changes in networks	35
1.4.4	Mixing patterns in networks	37
1.5	Molecular epidemiology and network analysis	41
1.6	Research Goals	44
 <b>Chapter 2</b>		
<b>Methods and Materials</b>		
2.1	STD control in Manitoba and data source for sexual network construction	45
2.2	Data management	46
2.3	Manitoba Sexual Network construction	47
2.4	Specimen collection and DNA extraction	47
2.5	PCR amplification	48
2.6	Nucleotide sequencing analysis	50
2.7	Statistical analysis	51

### **Chapter 3 Concordancy between Molecular and social data**

3.1	Introduction	53
3.2	Overview of the Manitoba sexual network	54
3.3	Genotypic diversity of chlamydia in Manitoba	55
3.4	Concordancy between epidemiologic and molecular data	59
3.5	Geographic patterns	63
3.6	Discussion	66

### **Chapter 4 Mixing patterns and characteristics of chlamydia infected cases present in Manitoba**

4.1	Introduction	73
4.2	Mixing patterns among study participants and their sex partners	75
4.2.1	Stratification of the data	76
4.2.2	Mixing patterns across age groups	76
4.2.3	Mixing patterns across sexual activity classes	80
4.3	Characterization of the chlamydia infected population in Manitoba	83
4.3.1	Case characteristics based on partner number	84
4.3.2	Case characteristics based on partner location	87
4.3.3	Case characteristics based on transmission success	90
4.3.4	Characteristics of cases present within certain network sizes	93
4.3.5	Characteristics of cases based on network measures of closeness and betweenness	94
4.4	Discussion	97



**Chapter 5      Demographic analysis of molecular genotypes and geographic distribution of sexual network connections in Manitoba**

5.1	Introduction	109
5.2	Overall geographic distribution of sexual partnerships	110
5.3	Geographic connections within co-infected networks	113
5.4	Geographic distribution of partnerships within different network sizes	116
5.5	Analysis of molecular genotypes identified in the province	116
5.6	Geographic distribution of chlamydia molecular genotypes	122
5.7	Predicting positivity from component size vs. 3-plex number	130
5.8	Discussion	131
	<b>Summary</b>	<b>140</b>
	<b>References</b>	<b>146</b>

## LIST OF TABLES

2.1	Summary of PCR and sequencing primers	49
3.1	Summary of number of components of size $n$ within the dataset	56
3.2	Summary of identified proportions of chlamydia genotypes	57
3.3	Summary of variant genotype nucleotide sequence	58
3.4	Univariate analysis of variables associated with discordant networks	64
4.1	Age mixing matrices between cases and their contacts	79
4.2	Partner number mixing matrices between cases and their contacts	82
4.3	Characterization of chlamydia infected cases based partner number	85
4.4	Characterization of chlamydia infected cases based on partner location	88
4.5	Characterization of chlamydia infected cases based in transmission success	92
4.6	Average network size for cases based on age, gender and risk area	94
4.7	Average closeness for cases within networks based on age, gender and risk area	96
4.8	Average betweenness for cases within networks based on age, gender and risk area	96
5.1	Demographic analysis of rare and common chlamydia genotypes	121

## LIST OF FIGURES

1.1	Chlamydia incidence levels from 1991 to 1998 within Manitoba and Canada	5
1.2	Network consisting of six individuals	24
1.3	Egocentric or personal network	25
1.4	A social network	26
1.5	Structural features of networks: linear, radial and branched	29
1.6	Network consisting of a bridge individual	29
1.7	Network measures: K core, clique, and k plexes	30
3.1	Percent of concordant components with respect to chlamydia genotype per network size	61
3.2	Distribution of chlamydia genotypes within the two largest components	62
4.1	Chlamydia incidence levels in Manitoba by RHA	77
5.1	Provincial geographic sexual connections	111
5.2	Sexual partner mixing between individuals in Manitoba and Winnipeg	114
5.3	Provincial geographic sexual connections of co-infected networks	115
5.4	Geographic sexual connections for networks of size 2	117
5.5	Geographic sexual connections for networks of size 3	117
5.6	Geographic sexual connections for networks of size 4	117
5.7	Geographic sexual connections for networks of size 5	118
5.8	Geographic sexual connections for networks of size 6	118
5.9	Geographic sexual connections for networks of size 7,8,9 and 10	118
5.10	Geographic sexual connections for networks of size 11,12,13,14,18 and 19	119
5.11	Geographic sexual connections for networks of size 39,41 and 82	119
5.12	Provincial geographic distribution of common chlamydia genotypes	123

5.13	Provincial geographic distribution of rare chlamydia genotypes	124
5.14	Provincial geographic distribution of chlamydia variant genotypes	125
5.15	Winnipeg geographic distribution of chlamydia genotypes	127
5.16	Winnipeg geographic distribution of chlamydia variant genotypes	128
5.17	Linear regression of the number of 3-plexes against positivity	132
5.18	Linear regression of component size against positivity	133

## LIST OF ABBREVIATIONS

- $\beta$ - Transmission probability
- C – Sex partner change rate
- CSW – Commercial Sex Worker
- D - Duration of infectivity
- DNA – Deoxyribonucleic Acid
- FSW – Female Sex Worker
- GC – *Neisseria gonorrhoea*
- IDU – Injecting Drug User
- KW - Kruscal-Wallis
- MCDC – Manitoba Communicable Disease Control
- MOMP – Major Outer Membrane Protein
- PCR – Polymerase Chain Reaction
- PID – Pelvic Inflammatory Disease
- $R_0$  – Reproductive Number
- RHA – Regional Health Authority
- STD – Sexually Transmitted Disease
- VD – Variable Domain

# CHAPTER 1

## INTRODUCTION

### 1.1 Burden of Disease

The genus *Chlamydia* consists of four distinct species: *Chlamydia trachomatis*, *Chlamydia psittaci*, *Chlamydia pneumoniae*, and *Chlamydia pecorum*. All of these chlamydia species are obligate intracellular pathogens that infect humans and in the case of *Chlamydia pecorum* and *Chlamydia psittaci*, can infect birds as well. *C. trachomatis* serotypes A, B and C cause trachoma, the most common cause of infectious blindness worldwide. The Lymphogranuloma venereum (LGV) biovar, is an ulcerative sexually transmitted disease (STD) found primarily in developing countries. *C. trachomatis* serovars D-K are sexually transmitted and cause ocular and genital infections in both males and females. Through sexual transmission, these organisms infect the columnar epithelial cells of the urethra in men and the endocervix in women, causing inflammation and epithelial ulceration and scarring. This organism has also become the most prevalent STD in both the United States and Canada, surpassing infection rates of *Neisseria gonorrhoea* (GC).

The cost of treatment for *Chlamydia trachomatis* in Canada has been estimated to range between \$89 million and \$123 million annually (1990 dollars) (Goeree *et al.*, 1993). In the United States, estimates in 1990 were as high as \$2.18 billion with 4 million new cases occurring annually. Ten million cases per year are believed to occur in Europe and 89 million cases worldwide (Guaschino *et al.*, 2000). There is a large number of

undiagnosed cases, largely a result of the substantial number of asymptomatic infections. In the U.S., from 1984 to 1997, reported rates of chlamydia increased from 3.2 to 207.0 cases per 100,000 population. This trend primarily reflects increased screening and identification of asymptomatic infections and improved reporting capacity rather than a true increase in disease incidence.

*C. trachomatis* infections are the most common, preventable bacterial disease, surpassing *N. gonorrhoea*. However, approximately 75% of infected females and 50% of infected males are asymptomatic and are clinically silent making it difficult to identify and treat infected persons, for the prevention of long term sequelae and the interruption of transmission chains.

Clinical symptoms include urethritis, epididymitis, and proctitis in men. In women, cervicitis, acute urethral syndrome, bartholinitis and salpingitis are seen to occur. Conjunctivitis and disseminated infections can appear in both males and females (Stamm and Holmes, 1990). A large portion of infected individuals are asymptomatic, however, if symptoms do occur, they can appear within one to three weeks after exposure. Men and women can experience discharge or pain during urination. Up to 40 % of women with untreated chlamydia will develop pelvic inflammatory disease (PID) and of these, 20% will become infertile, 18% will experience chronic pelvic pain and 9 % will have a life threatening tubal pregnancy (Rice and Schachter,1991). PID results in scarring of the fallopian tubes which may inhibit the process of fertilization. In men, chlamydial

infections may lead to epididymitis, pain and swelling in the scrotal area. Left untreated this condition can cause male infertility.

The cause for great concern regarding chlamydial infections is centered around the long term sequelae that are suffered, in particular by females. In women, chlamydia infections play a role in the development of pelvic inflammatory disease (PID) or salpingitis, involuntary infertility and ectopic pregnancy (Rice and Schachter, 1991; Coste *et al.*, 1994). Women who have positive IgG antibodies to *C. trachomatis* are at a five times higher risk of ectopic pregnancy than women who do not have IgG antibodies (Rice and Schachter, 1991; Coste *et al.*, 1994). Severe complications can arise from maternal chlamydial infections. Associations between preterm birth, premature rupture of membranes, intrauterine growth retardation and presence of chlamydia have been found (Claman *et al.*, 1995; Cohen *et al.*, 1990; Rettig, 1998). In addition, a study looking at commercial sex workers (CSW) in Nairobi found that those infected with chlamydia had a fourfold increased risk for HIV seroconversion in comparison to CSWs who did not have chlamydial infections (Plummer *et al.*, 1991). The high cost of treatment and burden of illness is primarily due to the high incidence and prevalence of these organisms, as well as the high numbers of asymptomatic cases who do not get treatment before sequelae develop.

In an attempt to reduce the cost and repercussions of infection, programs and policy guidelines have been developed. Infections due to chlamydia have become nationally notifiable in Canada since 1990 which allows for the gathering of statistics, sex partner



information, and the monitoring and management of infections within specific populations. In addition, mass and selective screening programs to detect asymptomatic patients with disease have also been sanctioned (Bureau of Communicable Disease Epidemiology, 1989).

Following the implementation of a screening program in 1987 in Manitoba, there was a clear downward trend in chlamydia incidence levels (Figure 1.1). However, as in other parts of Canada, in 1999, Manitoba's most commonly reported STD continues to be *C. trachomatis*, with an incidence rate of 362.2 per 100,000 and 153.4 per 100,000 for females and males, respectively (Health Canada, 2000). Compared to previous years, the rate of infection for females in 1999 has slightly decreased from that of 1998 but remain above the levels present in 1996 and 1997. For males there was a slight increase from that in 1998, however, rates were still above 1996 and 1997 levels (Health Canada, 2000). In 1999, 2,995 cases of chlamydia were reported with 2,122 (70.9%) being females between 15-24 years of age. The higher rates among females is largely due to females being screened at the time of their annual pap tests and their attendance at prenatal and family planning clinics. Males are generally tested less often, typically only when they are identified as having been exposed to an infected partner or when symptomatic.

Unfortunately, from 1997 to the present, there has been a steady increase in infection rates following a brief period of relatively steady rates. This upward trend suggests that current prevention efforts may have now reached their maximal effect for preventing

infections and new methodologies need to be implemented to regain the downward trends in rates.

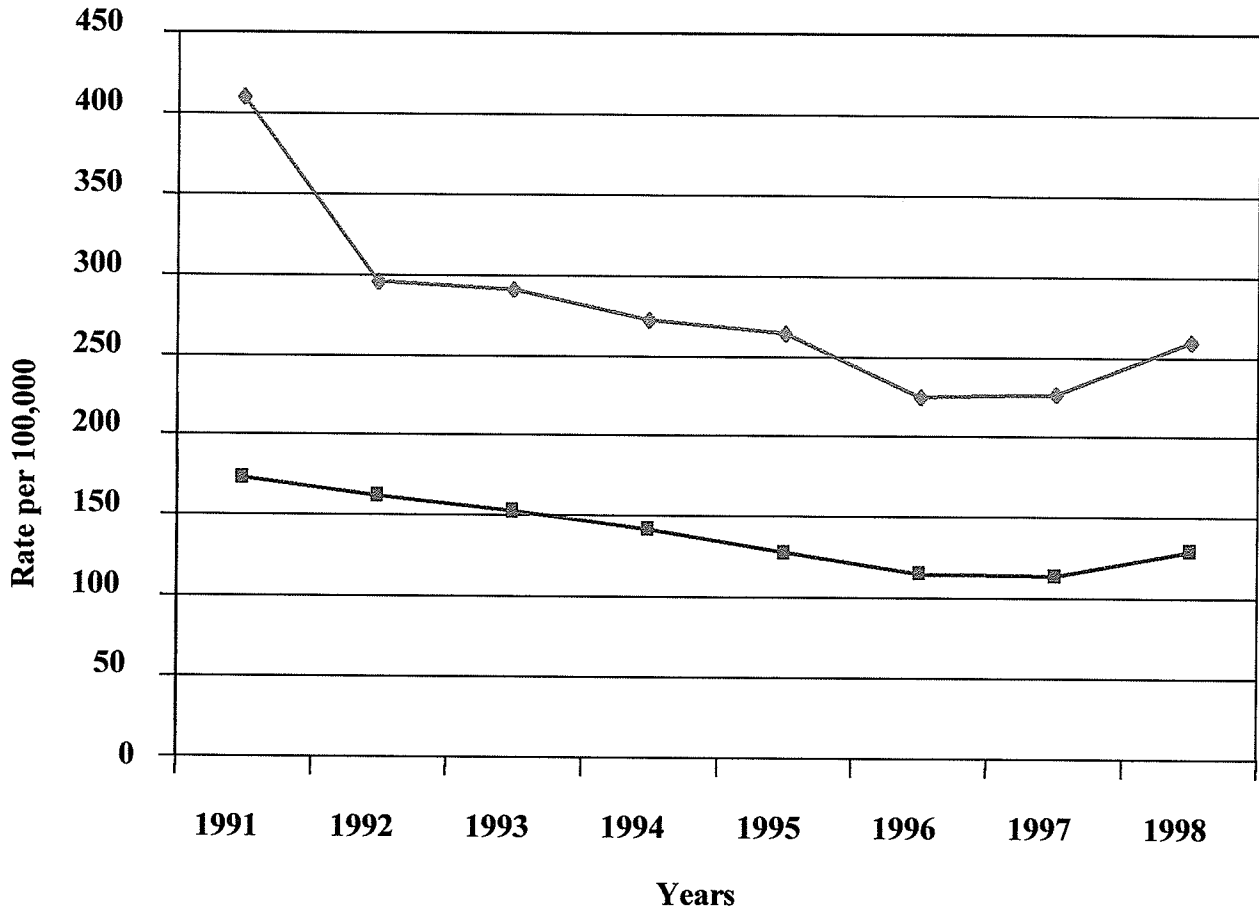


Figure 1.1: Chlamydia incidence levels from 1991 to 1998 within Manitoba —◆— and Canada —■—. Figure supplied by Debbie Nowicki, CDC Unit, Manitoba Health.

## 1.2 Epidemiology of *Chlamydia trachomatis*

Epidemiology investigates the distribution and determinants of disease. Over the past half century, epidemiologic investigation into STDs have included clinical observations of

microbial etiology, descriptions of the prevalence and distribution of disease, and the identification of the risk factors associated with disease (Aral and Holmes, 1999). This section will highlight some of the general findings for bacterial STDs, and will also include a summary of Manitoba - specific information.

Special populations are disproportionately affected by STD morbidity. In Western society and under developed countries, sex workers, homeless persons, adolescents and adults in detention, and migrant workers are among special populations with high STD morbidity.

High rates of STDs are seen in homeless populations. A study in Chicago revealed that 26% of homeless females had trichomoniasis, 6% were infected with gonorrhea and 5% had developed PID (Johnstone *et al.*, 1993). In Baltimore, 8% of homeless males and 11% of homeless females were positive for gonorrhea and syphilis while one third of these individuals reported having a prior STD (Breakey *et al.*, 1989).

Adolescents and adults in detention facilities have high levels of morbidity. In 1994, rates of infection for gonorrhea in male adolescents in detention were 152 times greater than those among males in the general population. Females in detention have reported rates for gonorrhea that are 42 times higher than females in the general population (CDC, 1996).

Incarcerated individuals frequently experience high STD incidence and morbidity rates. A jail facility for men in Los Angeles County reported syphilis rates of 507 cases per 100,000 persons. This number is 11 times the rate observed in the general population

(Cohen *et al.*, 1992). In 1996, Cook County Jail screening of female arrestees identified 22% of the 803 early syphilis cases identified in females in the county (CDC, 1998). In 1993 and 1994, after routine STD testing, United States correctional facilities reported 17% of inmates were infected with syphilis, 32.5% were positive for gonorrhea and 4.4% were positive for chlamydia (Bickell *et al.*, 1991).

Worldwide, STDs are a major health problem among migrant workers (CDC 1992, Jones *et al.*, 1991). This may possibly be due to limited access to health care systems, language and cultural barriers, and limited economic resources, combined with sexual behaviors that expose migrant workers to high-risk partners which tend to perpetuate high levels of STD morbidity among this group (Bechtel *et al.*, 1995; Brewer *et al.*, 1998). Epidemics of STDs are frequently seen in men who temporarily have to stay away from their families due to heterogeneity in economic growth rates in a country, leading to high levels of geographic mobility (Aral *et al.*, 1991; Hunt 1989; Carswell *et al.*, 1989). Long familial separations have resulted in a disintegration of long established marital and sexual patterns. The unequal sex ratio in towns resulting from the in-migration of male workers encourages prostitution. The combination of migrant labor and prostitution has led to epidemics of STDs among these groups all over the world (Aral *et al.*, 1991; Hunt 1989; Carswell *et al.*, 1989). Prostitutes also travel long distances to offer their sexual services elsewhere. This international migration of prostitutes has now become an important part of the sex industry (Koenig 1989).

services elsewhere. This international migration of prostitutes has now become an important part of the sex industry (Koenig 1989).

Monitoring the incidence or prevalence of STDs among female sex workers (FSW) and the proportion of men who report having had sex with FSW, provides information on the contribution of sex work to the spread of STDs. Men involved in this practice are typically characterized as being foreign, military servicemen, long distance truck drivers and/or migrant workers. In less developed countries, men purchase sex from female prostitutes more frequently. Both males and females involved in several STD outbreaks among migrant workers and travelers, have involved the exchange of money for sex and vice versa. In 1986 to 1987 in Seattle, an outbreak of a unique strain of penicillin producing *Neisseria gonorrhoea* was seen within a few months to have spread throughout the community (Handsfield *et al.*, 1989). Approximately 80% of those infected were individuals who had sex with FSWs or were themselves FSWs, or who had used illicit drugs around the time infection occurred. Globally, CSWs are most commonly found in settings characterized by poverty and social disintegration (Day 1988; Padian 1988). Contact with CSWs and FSWs is clearly a major factor in the epidemiology of STDs in both developed and developing countries (D'Costa *et al.*, 1985; Schachter *et al.*, 1983; Ryan *et al.*, 1998).

Adolescent and young females are greatly affected by chlamydial infections. Rates in the United States demonstrate that 1 in 10 adolescent girls tested for chlamydia are infected (Judson, 1985). Teenage girls have the highest rates of infection where girls between the

ages of 15-19 represent 46% of infections and 20-24 year old women make up another 33%. Infections are also widespread geographically with economically disadvantaged females between 16 to 24 affected greatly.

In Canada, a similar scenario is seen where the majority of infected females are between the ages of 15-19 (Orr *et al.*, 1994). In 1995, rates in this group were 1,109.1 per 100,000. This is almost nine times the national rate for males and females combined and six times the national rate for females (all ages). Females 20-24 had the second highest rate with 1,041.7 cases per 100,000, over eight times greater than the national rate. Rates for males significantly differed from females in both age groups and prevalence levels. Those males aged 20-24 had the highest rates of 335.6 cases per 100,000 while males between the ages of 15-19 had 169.6 cases per 100,000. These rates were 2.6 and 1.3 times greater than the national rates. The gender differentials may be an artifact of screening and may not reflect true differences between males and females. Higher rates in females likely reflect the fact that screening programs frequently focus on the testing of females. Although contact tracing should bring males in contact with the health care system, the reluctance to provide a urethral swab sample limits the number of diagnosed male cases. The introduction of nucleic acid amplified testing of urine specimens should provide opportunities for increased testing of male contacts and for male-oriented screening programs. In addition, gender differentials may be a result of females having a biological disposition for infection (Cohen *et al.*, 1985; Holmes, 1998).

Studies from both Canada and the U.S. report that North American Indians, Inuit, African Americans and Hispanics suffer disproportionately from chlamydia infections, compared to individuals who are of other ethnic backgrounds (Jolly *et al.*, 1995; Orr *et al.*, 1994; Beller *et al.*, 1992, Cullen *et al.*, 1990). There is no known biological reason for the different STD rates in different race and ethnic groups, therefore, race and ethnicity in the United States are presumably markers that correlate with fundamental determinants of health status such as poverty, access to health care, health care seeking behavior, illicit drug use and living in communities with high prevalence of STDs (CDC, 1996).

The growth of STD incidence in the general population has been attributed largely to adolescent or young adult groups (Holmes 1994). As a result of WWII, Canada underwent a "baby boom", where birth rates increased over those normally seen. In the late 1960's, these babies began to reach late adolescence, ushering in an increase in STD incidence. STD rates declined in teens and young adults around 1980 and has continued through to the 1990's. However, the children of the "baby-boom" generation are now reaching adolescence. The size of the teenage population is expected to increase by 20% in the decade from 1996 to 2005, with greatest increases in the African American and Hispanic populations creating a new wave of demographic pressure for STD transmission.

Of significant importance to STD research are behavioral risk factors that include: early age of intercourse, having multiple sex partners, non-use of condoms, men having sex with men, sharing injecting drug equipment, injecting in shooting galleries and the use of cocaine (Health Canada 1994).

Trends toward earlier and more liberal sexual behavior may have recently come to a halt (U.S. National Center for Health Statistics 1997). Reports show that 50% of U.S. teenagers (15 to 19 years of age) in 1995 had ever had sexual intercourse compared with 53% in 1988 and 47% in 1982. Age at first intercourse has two epidemiological functions (Sanchez *et al.*, 1996; Hofferth *et al.*, 1987). First, it is a true risk factor causally related to disease outcome. Etiologically, age at first sexual intercourse has been independently associated with an increased risk of the development of cervical cancer. Second, age at first intercourse serves as an indicator of other aspects of sexual activity. For example, it is correlated with sociodemographic factors such as race and socioeconomic status, sexual behavior variables, the number of sex partners and specific STDs.

Some general conclusions of studies of condom use demonstrate that condom use is highest in men. Men report having higher number of partners and highest rates of partner change in young, teenage groups (Tanfer *et al.*, 1993). Females have similar rates of condom use compared to males, however, condom use by females becomes less frequent with an increase in partner number. Steiner (1999) suggested that the main problem with condoms is the high level of non-use by people who are at the highest risk of HIV/STD. Maticka-Tyndale (1991) note that individuals between the ages of 15-24 have a pattern for sexual relationships of serial monogamy. This behavior leads to impediment of condom use as it provides a means to justify their behaviors based on their perception that their partners are well known and trusted. The CDC reported that less than 25% of young aged people, from 15-24 used a condom in the last year.



Increased use of crack cocaine spread throughout the United States and the Caribbean during the early and mid-1980's. Estimated numbers of current cocaine users has dropped from 5.8 million in 1985 to 1.3 million in 1982. However, the number of weekly users has not dropped since 1985 (Golub and Johnson, 1997). It is crack use by adolescents and young adults that has been of greatest interest to those concerned about the impact of cocaine on the spread of STDs. The perspective on the epidemic of cocaine use takes on a similar paradigm as that used in STD epidemics. It has been noted that a sub-group of hard core drug users are the first users of cocaine, followed by a quick expansion of use to close friends and associates (Golub and Johnson, 1997). If the drug becomes popular, teenagers coming into the circle will adopt the use of cocaine resulting in a rapid increase in cocaine use. Teenage social behavior, such as crack cocaine use, clearly promotes the spread of STDs since crack cocaine use leads to several STD risk behaviors, including exchanging sex for drugs or money (Caldwell and Caldwell, 1983). Gonorrhea has now been linked to membership of street gangs, which is another factor involved in STD transmission and drug network dynamics (Golub and Johnson, 1997). In addition, crack use is directly associated with syphilis, gonorrhea and HIV infection rates (CDC, 1993).

Aral *et al* (1990) noted 3 important behavioral factors involved in risk: 1) the number and selection of partners, 2) frequency of intercourse and 3) type of intercourse. The number of sexual partners and the selection of partners by an individual reflects the level of risk to which those individuals are exposed. Persons with a high number of partners, who in turn also have a high number of partners, are at very high risk of infection

compared to individuals who do not have high number of partners. Type of intercourse influences STD risk as receptive anal intercourse is associated with high risk, while oral sex is relatively low risk. Vaginal intercourse is almost universally practiced within partnerships (Laumann *et al.*, 1994). However, vaginal intercourse with an infected partner puts a woman at greater risk for infection with HIV and other STDs than a man. Frequency of intercourse affects risk, as infected individuals with a high frequency of intercourse are more likely to transmit and acquire infections.

Risk of exposure to an STD is directly related not only to the number of infected sexual partners, but also to the prevalence of STDs within one's pool of potential choice of sex partners. The number of partners within a specific time period, often one to three months, has been shown to be a risk factor for acquiring gonorrhea, chlamydia, genital herpes and HPV infections (D'Costa *et al.*, 1985; Handsfield *et al.*, 1986; Schachter *et al.*, 1983., Syrjanen *et al.*, 1984). Both in the general population and clinic attendees, marked gender differences exist in the recruitment of sex partners. Females meet their partners through less casual association, and know them better and for longer periods of time, compared to males (Sanchez *et al.*, 1996). A nondiscriminatory approach to sex partner recruitment increases the probability of sexual contact with members of high risk core groups and thus, exposure to STDs. Better understanding of patterns of partner recruitment in population subgroups would help to explain some of the variability in STD risk.

In Manitoba, aboriginal individuals suffer disproportionately from infections with chlamydia in comparison to other ethnic groups. Geographically, many aboriginals in

Manitoba inhabit rural areas of the province accessible only by aircraft or boat. Elliott *et al* (1999) found that overall trends in Manitoba indicate a concentration of high STD risk behaviors and STD infections among high risk sub-populations including aboriginals in northern remote communities. This may be due, in part, to the limited contact with the health care system by these communities, which can have a profound effect on the prevalence of STD (Fairely, 1997).

Epidemiology due to *C. trachomatis* infections in Manitoba was evaluated in 1994 where, after the implementation of a screening program, the annual incidence of *C. trachomatis* was highest in females aged 15-24 years. Recurrent infections were most common in females in this age group, registered North American Indians, and individuals with concurrent infections with gonorrhea (Orr, 1994)

In summary, based on these findings, persons infected with *C. trachomatis* are typically adolescent and belong to ethnic minorities or lower socioeconomic classes (depending on population). Other populations infected include homeless, adults in detention, migrant workers, commercial sex workers and substance abusers. Risk markers associated with risk of acquiring an infection include large numbers of sexual partners, coinfection, as well as concurrent partnerships. The importance of partner number on disease dynamics is discussed in the next section.

### **1.3 Beyond the individual: Core group and the reproductive number**

Previous approaches to delineate the epidemiology of STDs were more frequently based on the individual as the unit of analysis. As illustrated in the previous section on *C. trachomatis* epidemiology, these approaches have provided much information on the risk factors and risk markers associated with STDs. However, although providing valuable data, they fall short of fully capturing the risk an individual faces with respect to acquiring an STD.

The main drawback of individual based approaches to STD epidemiology is that information is not sought on the behaviors of an individual's sexual contacts. Regardless of an individual's behavior, that individual is not at risk of infection, unless they come in contact with an infected person. Therefore both individual behavior and the behaviors of their sexual partners, and their partner's partners, must be considered to fully understand STD epidemiology and transmission dynamics at an individual and population level.

Several authors have reviewed the importance of population versus individual based approaches in epidemiological studies of STDs (Shiboski and Sussman, 1996; Koopman and Lynch, 1999). An individual based approach does not fully capture the social forces involved in disease transmission. It only focuses on individual characteristics of infected persons, and not their surrounding social environment. Population based approaches, on the other hand, see the bigger picture. They incorporate information on both cases and their contacts to provide a depiction of the social relationships occurring within populations. The organizational structure of a population and the systematic interactions among its members help to determine the speed with which an STD epidemic spreads

through a population, the magnitude it reaches, and the paths through which preventive interventions diffuse.

The basic reproductive number,  $R_0$ , of an infection encapsulates the parameters and processes influencing the transmission dynamics within defined populations (Brunham, 1997; May, 1981; Anderson and May, 1992). The reproductive number is defined as:

$$R_0 = \beta c D$$

where  $\beta$  represents the probability of transmission of an infectious agent between infected and susceptible individuals;  $c$  represents the rate of contact between infective and susceptible individuals; and  $D$  represents the duration of infectivity.  $R_0$  represents the average number of secondary cases generated by one primary case in a population of sexually active individuals. If  $R_0$  is greater than one, then the infection will spread in a population. When  $R_0$  is less than one, the infection cannot sustain itself and disappears out of the population. Infection is maintained at an endemic level when the value of the reproductive number remains at one.

Therefore, the reproductive number is defined by three cooperating variables that incorporate biological ( $\beta$  and  $D$ ), and behavioral ( $c$ ) concepts.  $\beta$  can be affected by such variables as the number of sex acts per unit time, the type of sex act and the genotypic/phenotypic characteristics of the infecting agent and the potential host. Given their potential heterogeneity, estimates of  $\beta$  must be taken with caution, however it is generally accepted that  $\beta$  for chlamydia is near a value of 0.2, suggesting a relatively low probability of transmission in comparison to other bacterial STD's (Anderson, 1999).

Similarly, estimates of the duration of infection,  $D$ , are also subject to variability. Duration of infection for untreated cases of chlamydia are relatively long (estimated at approximately one year) and would act to counter balance the relatively low transmission efficiency (McCormack, 1979). This parameter has many external variables that can disrupt interpretation of results because many infected are asymptomatic. Furthermore, it is difficult to determine true incident infections apart from re-infection, recurrent infection and prolonged infection where studies show that women, without treatment, are still infected over a year after initial culture (McCormack, 1979).

The behavioral determinants, or the rate of change of sex contact, is measured based on peoples' behaviors. These measurements are more variable than the biological determinants which would therefore explain the variation in STD rates that can occur in a given population. It is also a key variable, as the sex partner change rates are clearly central in determining the potential number of secondary infections a given case can generate.

The measurement of contact between infectious individuals and susceptible individuals is denoted by  $c$ . Nagelkerke (1994) states that for a random mixing population,

$$C = m + \frac{\sigma^2}{m}$$

where  $m$  is the mean rate of partner change and  $\sigma^2$  is the variance. Therefore, for chlamydia this translates as the average number of partners exposed per unit of time, weighted by the extent of heterogeneity in numbers of partners. The only limitation of this equation is based on the partner selection process where people in higher sexual

activity classes offer more sexual liaisons than do those in lower classes. Hence, they have a higher likelihood of being selected (Anderson, 1991). Several complex formulae have been devised which describe the effects of different criteria in choosing sexual partners. There are a great number of people who will not have adequate numbers of partners to sustain a disease in a population. There are, however, a few who act to sustain a disease based on their having sufficient number of partners.

Rearrangement of the formula for  $R_0$  by Brunham (1997) reflects the importance of sexual behavior on STD epidemiology :

$$C = \frac{1}{\beta D}$$

From this new equation,  $c$  is the critical threshold of sex partner change rates that would keep the infection in an epidemic state (or  $R = 1$ ). This simple derivation highlights the significance of the rates of sex partner change as a determinant of the rate and degree of spread of infection.  $C$  for chlamydia is four, demonstrating how sexually active individuals play a disproportionate role in STD transmission. (Discussed under core groups). Individuals with extremely high numbers of partners have a higher likelihood of either being exposed to, acquiring, or transmitting pathogens.

### *Core groups*

Traditionally, a common assumption when assessing an individual's risk of acquiring an STD was that an individual has a random and uniform risk of infection based on promiscuity. Hethcote and Yorke (1984) moved beyond the random nature of STD infection and developed a hypothesis that any given population can be divided into two

responsible for maintaining the endemic state of infection (Brunham, 1997). Theoretically, these populations are defined as containing individuals who have high number of partners or a high partner change rate and are responsible for maintaining  $R_0$  above one. As such, it is thought that these individuals provide multiple transmission opportunities for movement of pathogens throughout the population. In order for the infection to be present in the population as a whole, connections must exist between the core group and the rest of the population. The degree of mixing between the core and non-core will have a profound impact on the pattern and the spread of infection.

Core group members who mix with the portion of the population with low sexual partner number (i.e. one) would result in frequent transmission termination events and potentially a low  $R_0$  for the population as a whole. Core group members who mix with individuals who have moderate numbers of sexual partners, would elevate  $R_0$ . The scenario where core and non-core group members mix is an example of dissortative mixing. A third scenario where core members mix exclusively amongst themselves can create a high  $R_0$  within the core but relatively little disease outside that group. This situation, where individuals in a certain group mix only with people in that same group, is referred to as assortative mixing. The extent and rate of spread of the infection through the population is a consequence of the mixing behaviors exhibited within and between the groups (Anderson, 1992; Gupta *et al.*, 1989).

Anderson and Garnett (2000) used mathematical models to assess the significance of sexual mixing between core and non-core groups. They used previously published data



Anderson and Garnett (2000) used mathematical models to assess the significance of sexual mixing between core and non-core groups. They used previously published data from contact tracing studies in which core groups were defined by area of residence (i.e. high levels of gonorrhea infection) to derive estimates of the degree of mixing within and between groups. Their results revealed high within-group mixing in the core, which suggests that within group transmission ensures the persistence of gonococcal infection in the community. Therefore, by identifying behavioral patterns of core group members, the degree of infections found within a population can be explained.

Bridge individuals may be critical for mixing between groups. Morris (1996) demonstrated this concept in Thailand, where young men and women and truck drivers, were responsible for linking high risk core groups with non-core groups. This can lead to a rapid increase in infections, followed by spread of the disease throughout the population. Morris (1996) and other authors (Kretzchmer *et al*, 1996) suggest that bridge individuals and core group members should be considered essential target groups for prevention strategies.

Although it is clear that core groups appear to have a disproportionate influence on disease maintenance and transmission, questions still remain. What portions of the population will be part of a core group and how can these individuals be identified for STD prevention efforts? It is evident that only a small segment of the population contains critical threshold rates to sustain infections in a population. This segment of the

STDs in a community and are the sites from which infections in the community originate. Such networks are termed core groups and by theory only within the core group are productive chains of infection maintained. Infections can enter and exit the core into adjacent populations, however, these infections ultimately terminate within the adjacent populations. Therefore, it is the existence of core groups within communities which are responsible for endemic persistence of an STD.

### ***1.3.1 Core groups in Manitoba***

As noted by Brunham when explaining the core group theory, a high rate of sex partner change is a key behavioral factor in maintaining high STD rates (Brunham, 1997). In Manitoba there appears to be geographic segmentation of core groups. High STD rates tend to cluster in two main geographic locales: the northern rural section of the province and the downtown core area of the largest city in the province, Winnipeg. Both of these areas are characterized by high unemployment rates, a disproportionate number of young people, first nations people, and single parent families, racial, ethnic and cultural diversity, language barriers and high amounts of migration (Blanchard, 1998).

Blanchard *et al* (1998) described and compared the transmission dynamics of chlamydia and gonorrhea in Winnipeg, and assessed these implications for control programs. In response to a control program implemented in 1970's and mid 1980's for gonorrhea and chlamydia, respectively, incidence levels and transmission dynamics changed over time. Both organisms were present within core areas of the province, defined geographically. Chlamydia became more core dependent with higher proportions of individuals reporting

chlamydia, respectively, incidence levels and transmission dynamics changed over time. Both organisms were present within core areas of the province, defined geographically. Chlamydia became more core dependent with higher proportions of individuals reporting contact with at least one core member. For gonorrhea, infections declined evenly among the core and the non-core segment of the population with a decline in the proportions of individuals reporting contact with a core member. Therefore, these authors enhance the notion that control efforts should respond to changes in the epidemiology of STDs caused by the effects of a control program. Core groups are an essential point of intervention as chlamydial infections should be targeted through these core groups. Gonorrhea prevention efforts should also focus on core groups; however, if incidence levels continue to decrease and geographic segmentation continues, it will be difficult to target gonorrhea infections through the core groups. Intervention can then be assessed to interrupt the spread of infection within specific sexual networks rather than specific geographic areas.

#### **1.4 Social networks and STD epidemiology**

Several recent papers have suggested a paradigm shift is occurring in STD epidemiology (Aral, 1999; Wasserheit and Aral, 1996). As noted earlier, many investigations into STD epidemiology are moving beyond the individual to encompass population – level approaches. This shift to a more direct focus on the social context of individuals is encapsulated by the seminal paper by Wasserheit and Aral (1996). These authors have outlined STD transmission within the context of social networks.

cultural groups. Measures can also be made of the extent of concurrency and the average level of risk engaged in by members of a network. This perspective has the potential to further our understanding of STD epidemics and contribute to the development of more effective targeted control strategies. This section will provide background on social network concepts.

The new paradigm calls for consideration of an individual's risk environment, as opposed to a focus on an individual's risk behaviors. The concept is based on the idea that the significance of an individual's risk behaviors are increased or decreased depending on the infection status and behaviors of the people with whom they interact (i.e. a risk environment). This set of interactions, in conjunction with other types of personal interactions, defines a person's social and sexual network.

A network can be defined as a compilation of nodes that are all connected together, directly or indirectly, by a common connection. A component is referred to as a distinct subset of a larger network. The nodes may represent people, groups, events or other units connected by a relationship (Klovdahl, 1985). Links between the nodes are referred to as ties, arcs or edges and define the relationship between the nodes (Scott, 1991). Links can be such things as roles (student, co-worker, executive); cognitive relationships (love, hate, friendship); active relationships (sexual, monetary); or geographic relationships. Social networks can represent an array of relationships and therefore can be applied to any form of social analysis. To generate a sexual network, the nodes represent people,

and the links represent sexual contact. Figure 1.2 represents a visual depiction of a network of six people.

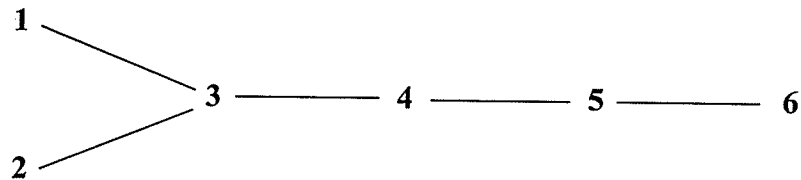


Figure 1.2: Network consisting of six members derived from the adjacency matrix depicted below.

The following illustrates this network in the form of an adjacency matrix.

	1	2	3	4	5	6
1	0	0	1	0	0	0
2	0	0	1	0	0	0
3	1	1	0	1	0	0
4	0	0	1	0	1	0
5	0	0	0	1	0	1
6	0	0	0	0	1	0

There are six people that make up the network and the connections linking them together are embedded within the matrix. A 1 indicates a direct connection between those persons in the network while a 0 indicates that two nodes are not directly connected.

Two types of networks can be analyzed. The first is termed an egocentric, or personal network (Figure 1.3). An egocentric network consists of a single person, those persons who are in direct contact with that person, and the social relationship that links these people together.

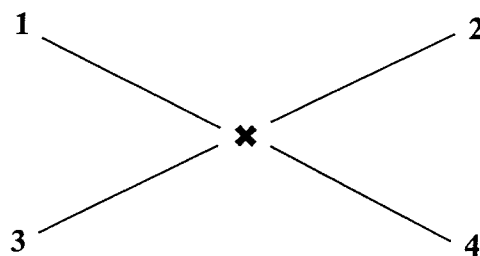


Figure 1.3: Egocentric or personal network where information is generated for person x only. Here, person x has four direct connections

The second type of network is termed a sociometric network and refers to a larger set of individuals within a particular population and their connecting links. Figure 1.4 places the egocentric network of person X within the context of a larger sociometric network. In this example, a consideration of the sociometric network serves to highlight the central role person X plays in connecting two separate groups of people.

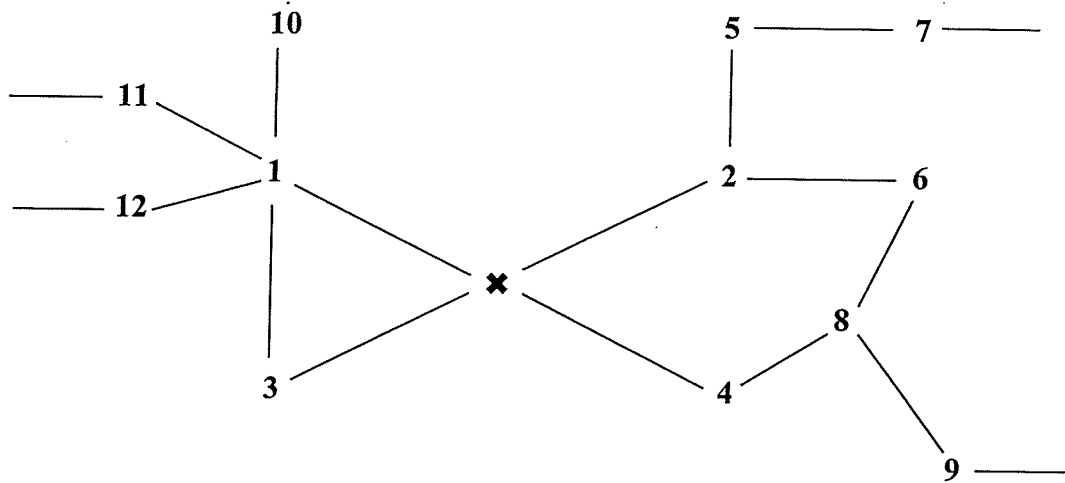


Figure 1.4: Comparison of the social network for person x and their personal network shown in figure 1.3. A social network encompasses more contextual information than a personal network.

The framework for social network analysis was built in the 1930's by a sociologist who was interested in measuring interpersonal relationships within groups of individuals (Berkowitz, 1982). In the 1950's, the branch of mathematics known as graph theory provided powerful tools for analyzing the properties of networks (Harary *et al.*, 1953; Berkowitz, 1982).

In the early 1980's, Auerbach *et al* (1984) published results of an intensive investigation of individuals who were infected with HIV even before researchers knew they were dealing with a virus, and illustrated the potential application of network analysis for the investigation of infectious disease transmission. These results helped to establish the infectious nature of HIV, and demonstrated that the infected individuals were not just connected by chance. Using the same data, Klovdahl *et al.* (1985) took a few steps

forward to demonstrate the usefulness for epidemiology of explicitly conceptualizing human populations in network terms. Klovdahl showed how the inconsistent temporal on-set of disease between partnerships within networks lead to assumptions of more than one point of entry for infection and that possibly more than one strain of the infectious agent was involved. The results from this study was the template for more intense network analysis in the study of STD epidemiology.

#### ***1.4.1 Social network concepts***

Using the network from figure 1.4 several network measures currently used in describing social networks can be defined. If two people are directly joined together, they are said to be adjacent to one another. The number of persons directly connected to a given individual is called the degree for that individual. For example, person 1 has a degree of 5 while person 2 has a degree of 3. A path refers to the lines that connect an individual to any other one while the length of a path is measured by the number of lines connecting those individuals. A component (frequently called a "network") refers to a group of people who are all directly or indirectly linked to one another by a known relationship. The distance between two people is the length of the shortest path which connects them. For example, person 5 and person 6, are only connected indirectly via person 2, therefore, person 5 has a distance of 2 from person 6.

The density of a network refers to the actual number of connections between the points, relative to the maximum number of connections that could exist within that network. Density values can range in value from 0 to 1 (The density for the network in figure 1.4 is



0.17). Centrality measures reflect the potential importance of an individual within a network. A person is locally central if they have a large number of direct connections to others in the network. A person can be globally central when they have a position of strategic significance in the overall structure of the network. A globally central person is positioned at a short distance from many other points in the network and depending on infectious status or closeness to an infection, that person can be influential in causing rapid transmission. A locally central individual is one that has a high number of partners, however, only a few have more than one partner that branches out into a larger network. Locally, that individual is critical in disease transmission to their direct contacts. Looking at the network as a whole, however, that individual does not pose a high risk for disease transmission. Betweenness measures the extent to which a particular point lies 'between' the various points in the graph. People who are centrally located, or are between a large number of other people can greatly influence the dissemination of disease.

Overall measures of size, shape and structure of networks are very useful in network analysis. Networks can be large or small, linear, branched or radial, and consist of dense or sparse connections (Figure 1.5). The degree of these measures effects transmission dynamics throughout a population. A person who connects individuals who are part of different subgroups within a network, are referred to as bridges. For example, bridge individuals may have one contact who lives in a high risk area, and another contact who lives in a low risk area, or they may link individuals who are within different age groups. (Everette, 1973). In figure 1.6, person 5 links two densely connected subgroups together.

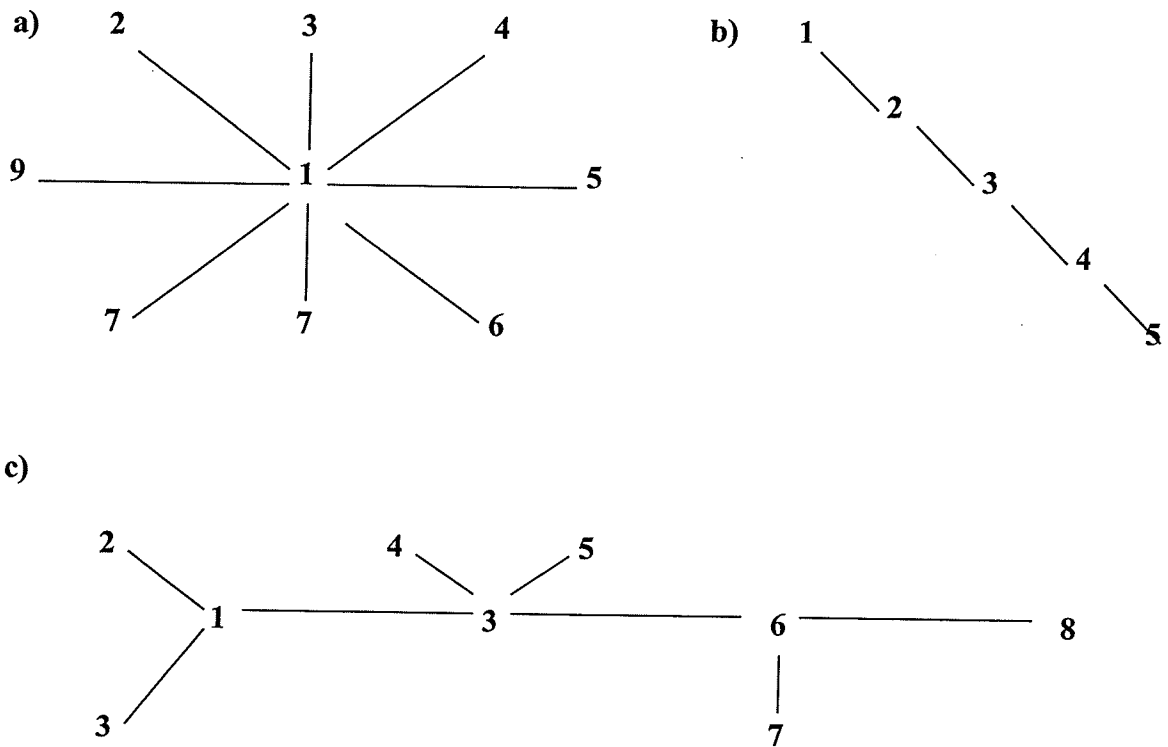


Figure 1.5: Structural features of networks. A) radial network, B) linear network, C) branched network. Each structure has their own relevance to network analysis.

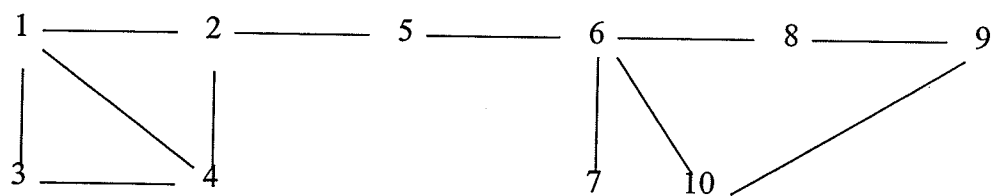


Figure 1.6: Example of a bridged network. Person 5 links individuals 1,2,3, and 4 to individuals 6,7,8,9 and 10. Removing person 5 will create two separate components.

Within a single network, nested subgroups of closely tied individuals can be present (Figure 1.7). A  $k$ -core is the number of people with  $k$  contacts who are connected to each other. For example, a 3k core is identified as a group of connected individuals with no less than a degree of three. A clique is a subset of people in which every possible pair of people are directly connected. A 3-plex is identified as three people connected in a row.

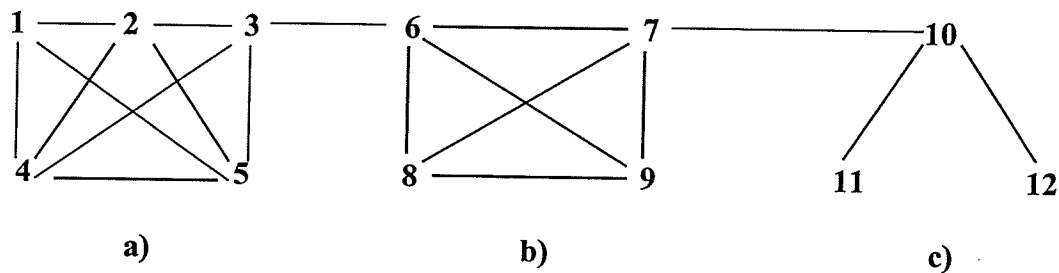


Figure 1.7: Social network measures used to study patterns of network structure and disease transmission. This figure shows a network with three distinct types of subgroups. For the subgroups above, each respective letter represents a 3k-core (a), clique of size 4 (b), and 3-plex (c).

Sociometric data permits calculation of network measures for purposes of identifying persons with important roles in a network. Identification and characterization of core groups has been a critical turning point in STD epidemiology. They have served as the foundation for the way networks function in the spread of disease throughout a population (Brunham, 1997; Rothenberg *et al.*, 1996). There has been increasing attention devoted to the role of uninfected members within networks, and how core members mix with non-core members (Aral *et al.*, 1999; Sattenspiel, 1987). Defining

core group structure has led to an explanation regarding population network structure and interconnectedness of individuals in a given population (Rosenberg *et al.*, 1999; Thomas and Tucker, 1996). Densely connected segments of a network suggest a higher probability that disease will spread quickly through these segments. Network measures can provide insight into persons heavily involved in maintenance and spread of infection, based on network position. Therefore, network analysis is used to analyze the parameters that influence pathogen transmission dynamics, as well as a tool for providing essential information for the development of prevention and control programs. The next three sections will review the influence of network structure, temporal changes within networks, and mixing patterns between individuals on STD transmission dynamics.

#### ***1.4.2 Structure of networks***

Several studies have highlighted the importance of network structure in relation to transmission dynamics for STDs. A landmark study conducted in Colorado Springs was one of the first and largest studies to focus on the transmission of HIV with respect to network analysis. Study participants recruited included CSWs, injecting drug users (IDU) and their associates. More than 600 individuals were found to be connected to one another, with each individual in the network being only three steps from an HIV infected individual (Klovdahl *et al.*, 1994; Rothenberg *et al.*, 1995). These authors were the first to demonstrate the ability to construct and use network analysis in studying STD transmission dynamics in a population (Rothenberg, 1995).

Descriptive analysis of networks also has direct implications for disease control. Sexual networks are not isolated networks that exist separately as a sub-component of the general population. A key issue in STD spread is the occurrence of bridging between networks such that low risk networks (on average having low partner numbers and low risk behaviors) are connected to high risk networks causing movement of infections beyond the two types of networks. A critical point for intervention is the small set of individuals who bridge the different networks.

Morris *et al* (1996) provided evidence on the importance of bridge populations in Thailand. Sexual network data was used to identify men who had sex with both CSW and non-CSW. Twenty-five percent of these men were found to be truck drivers. They were more likely to be HIV positive, have had at least one prior STD in the past year, be young in age and exposed to more female contacts than wives. The CSW reported less condom use than non-CSW. They concluded that bridge populations may be as important to identify as core groups due to their connecting different sub-populations.

Similarly, Calzavara *et al.* (1999) identified bridge individuals amongst aboriginal populations in Ontario. These authors found a high degree of geographic bridging amongst some aboriginal individuals. Approximately half of the study participants reported having partners only within the same community, while 21% reported having contacts both within and outside of their community. These latter individuals had more sexual partners and serve as sexual bridges between communities, possibly providing pathways for the transmission of HIV through this population. The pattern of sexual

partnering reported in this study, suggests that if the pool of infection among the partners from an outside community increases, HIV will spread rapidly within and across communities.

Several studies have focused on the risk of infection based on the position of individuals within a network and partner choice within and between networks. Friedman *et al* (1997) identified groups of tightly linked individuals at the center of drug sharing networks, who were at a higher risk of becoming infected with HIV, and who engaged in a wide range of high risk behaviors compared to other network members. The seronegative individuals in this group were at a higher risk of serconversion than individuals in peripheral sections of the network, based on their highly connected position within the network. These authors note that if HIV infection were to enter this tightly linked subgroup, their position within the network would provide an opportunity for rapid transmission of HIV outwards toward many peripheral network members.

In Colorado Springs, Woodhouse *et al* (1994) studied prostitutes and IDU's, and their risk for HIV based on the interconnectedness of these individuals within networks. One large component was identified containing 3658 connected individuals. There were 19 reported HIV infections in the population of which 11 were identified within this large component. The HIV positive cases in the component were confined to a less interconnected group practicing low risk sexual behavior. This aspect of the network suggests that the isolated positions of HIV positive individuals may serve as a barrier to HIV transmission in this region. Therefore, as opposed to finding a situation of epidemic

spread of infections, network analysis can demonstrate the failure of an infectious agent to propagate in certain populations with a specific social structure.

Latkin (1993) found that increased network size was associated with multiple partners for males and exchanging money and drugs for sex, while network density was inversely associated with exchanging money and drugs for sex. These findings are concordant with the social opportunity hypothesis and the social integration hypothesis, respectively. The former theory states that larger personal networks may allow for the opportunity to meet more potential partners, while the latter theory states that persons who are less integrated in the network may be less intimate with other network members, and hence more easily purchase sex in place of a supportive network.

Rosenberg (1999) applied the core group concept to identify syphilis core group members in Louisiana. Forty two percent of contacts (and with low numbers of partners) were connected directly or indirectly to a core group member. An individual was identified as a core group member if the individual was 1) a female who a) received money or drugs for sex or b) had more than or equal to five sexual partners and 2) a male who had more than or equal to five sex partners. The network that maintained syphilis infections in this community was comprised of many persons with relatively small numbers of partners, but linked closely to core transmitters. These contacts also exhibited moderate drug use risk behaviors. These characteristics put the whole network at high risk for syphilis with the potential to elevate network members risk of infection.

In summary, several network characteristics have been shown to affect the transmission dynamics of STDs throughout a population. Individuals that bridge populations, such as migrant workers or CSWs, and between different ethnic groups, result in increased spread within these groups, as well as within the groups which they bridge. Increased network density as found in core groups and tightly knit networks generates a structure with higher potential of disease spread, in addition, larger networks increase risk of disease transmission, where there is an increased potential to meet new sexual partners. By identifying and targeting these characteristics, prevention programs have the potential to effectively decrease the rate of transmission among infected populations.

#### *1.4.3 Temporal changes in networks*

Various epidemic patterns can be explained by network effects over time. Epidemics with multiple peaks can result as an infectious agent jumps from one distinct network to another. Each jump, followed by rapid spread in that network, coincides with a respective peak in case numbers. The extent to which an infection spreads throughout a network depends on sexual behaviors of the individuals within that network. Over time, these sexual behaviors can change such that more sexual links exist to bridge populations together, resulting in epidemic peaks in the bridged populations. Temporal aspects of partnerships within a network can also affect incidence such that concurrent partnerships within networks result in faster spread of infections. This next section will detail studies that have revealed important issues in network change over time, and issues relevant to the development of prevention and control strategies.



Monitoring network stability can help to explain the social structure of networks. Hoffman (1997) discovered in some networks of IDU's, significant movement of individuals in and out of networks over time. Some networks were consistently the same size but with new members. This was felt to be due to the need to maintain a constant network number so enough money was available to buy drugs. This, presumably, is most common amongst poorer IDUs who have little income. The end result is that risk in this group would be high because of the constant, potential for needle sharing with new members.

Potterat *et al.* (1999) used concurrency, a network measure of overlapping partnerships, as a predictor of disease transmission for chlamydia in Colorado Springs. This study revealed how infections can spread through segments of the population with high rates of concurrent partnerships because they lead to larger groups of people connected at any one time through chains of sexual partnerships. Concurrency was also shown to be a better predictor of infection than some other measures such as partner number and mixing between ethnic groups. These findings have important public health implications for STD control.

Microstructures was the main network characteristic implicated in an increased syphilis transmission in Atlanta (Rothenberg, 1998). A network infected with syphilis was monitored for over a year and revealed a dramatic change in network structure. Microstructures such as n- cliques, k-plexes and k-cores increased in number over time and indicated an increased potential for transmission. The changing configurations

identified in this study and in the literature discussed above, are likely to assist disease transmission and also provide a basis for tying program intervention activity to declining transmission. Routine assessment of networks can be used as a surveillance tool for providing warning of epidemiologic changes that may increase incidence.

#### ***1.4.4 Mixing patterns in networks***

Social network analysis has been used to study several different concepts in STD research. Based on sections 1.3 and 1.4, social network analysis focuses on the connections between people within the population and the influence their connections have on the spread of STD's. Various authors have noted that network structure is maintained by three variables: 1) The mean partner number per person per unit of time, 2) the duration of the partnerships and 3) the distribution of partnerships over the population. The latter point relates to mixing patterns occurring between and among subgroups within the population. Several publications suggest that mixing across subpopulations contributes to the spread of STDs in the population (Gorbach *et al.*, 2000; Morris *et al.*, 1996; Havanon *et al.*, 1993).

STDs follow routes of sexual contact, and the incidence and pattern of their transmission depends on who has sex with whom (Anderson *et al.*, 1990; Anderson *et al.*, 1991; Ellen *et al.*, 1997). Anderson and May (1992) state that "... most sexually transmitted diseases cannot be understood without acknowledging the marked heterogeneity in the degrees of sexual activity within the overall population". Within the concept of mixing, infections may spread from high incidence to low incidence, through links between individual

members or indirectly through bridge populations. Mixing patterns can also help to define the risks faced by specific individuals.

Aral *et al.* (1999) computed mixing patterns to determine the risk of infection based on characteristics of the study participants. These authors measured mixing patterns between partnerships in terms of race/ethnicity, age, education and numbers of partners. Within these categories, dissortative mixing predominated and was associated with significant risk of gonorrhea and chlamydia infections spreading throughout different segments of the population. Mixing patterns based on geographic location revealed that individuals residing in areas with low levels of infection showed assortative mixing. These individuals were mainly infected with chlamydia. Individuals residing in areas of low levels of infection with dissortative mixing to high prevalence populations were mainly infected with gonorrhea. This study demonstrates how mixing patterns within defined populations can influence the risk of infection. Dissortative mixing across age, sexual activity class, and geographic locations allows the potential for infections to spread from one group to another, placing individuals seen as low risk at an actually high risk.

Other authors have analyzed mixing patterns of spatial sexual mixing, or mixing patterns between individuals in different geographic locations (Ellen, 1997; Rothenberg, 1983; Potterat, 1985). In Upstate New York, Rothenberg evaluated STD patterns based on incidence levels in definable geographic areas. Gonorrhea incidence was neither uniformly nor randomly distributed among this population, but rather exhibited geographic clustering. The majority of infections occurred in the central segment of the

region, or the core area. This core area was surrounded by geographic areas with smaller incidence levels, or adjacent areas. Encompassing these locations, were areas with the least amount of infections, or peripheral areas. An interesting finding was that these core areas were characterized as inner city, high- poverty areas, with high population density, as compared to the other areas. In addition, the core areas also had a 20% prevalence rate, sufficient to sustain infections in the geographic core segment, as well as the rest of the population. In conclusion, analyzing infection rates geographically served to be highly relevant to identify populations with geographic clustering and to focus on individuals important in maintaining high STD rates. These observations provide an opportunity to narrow the focusing of epidemiological resources, as a major disease control strategy.

Service and Blower (1995) analyzed sexual mixing patterns with relation to risk of HIV infection. These authors found that a critical variable for assessing infection risk was based on the age group with whom an individual mixed. Seropositive and seronegative individuals had different age-stratified sexual mixing patterns. Seronegative persons were more likely to have had sex with men under the age of 30, while seropositive individuals were more likely to have had sex with men over the age of 30. This association was independent of the number of partners each case had. Young gay men who mixed with individuals in their own age group, had a low risk of seroconversion, while gay men who chose older partners (>30) were at high risk of serconversion. In this population, older aged males were more likely to be infected with HIV. These results suggest that STD control efforts should target age groups with high STD rates, as well as the individuals they are most likely to choose as sexual partners.

Haraldsdottir *et al* (1992) generated mixing matrices derived from sexual network data. Within a homosexual population in Iceland, these authors attempted to demonstrate the influence of sexual partner mixing with respect to sexual activity class on the spread of HIV. Data revealed an overall pattern of dissortative mixing, where persons with high numbers of partners typically mix with individuals who have low numbers of partners. Individuals with high numbers of partners are at an increased risk of acquiring HIV. As mixing occurs between these two groups, infections can readily spread to groups of individuals with low numbers of partners, contributing to the spread of HIV. Compared to mathematical models for assortative and proportionate mixing, results from this study suggested the epidemic in this population would reach a greater magnitude, but with a slow initial start.

Ethnicity mixing may be important to STD transmission. Ethnic factors may act as a barrier to STDs in that different ethnic groups may be separated by geography, language, social class, social customs and prejudice. Once ethnic barriers are lessened, transmission of STDs across ethnic groups could increase (Catania, 1996). In the United States, African Americans, and to a lesser extent Hispanics, experience higher rates of STDs than other ethnic groups. In Manitoba, analogous situations occur within the aboriginal population (Orr *et al*, 1994). There is no known biological reason for these findings, suggesting that ethnicity must be a marker for other determinants of health status such as poverty, access to health care, or drug use (Lauman, 1999). Using network analysis, Lauman (1999) has provided an explanation for this phenomenon. They found that dissortative mixing occurred between low risk and high risk African Americans. Low risk

(based on number of partners) African Americans were five times more likely to choose high risk African American partners than Caucasians or Hispanics. This factor elevated the overall incidence rates experienced in this population compared to other ethnic groups whose mixing trends tend to be assortative (low risk with low risk, high risk with high risk). In addition, the limited sexual mixing between African American and other ethnic groups contributes to the isolation of STDs within this population.

### **1.5 Molecular Epidemiology**

The use of molecular typing in conjunction with network analysis has been suggested as a novel approach to further increase our understanding of STD epidemiology (Day 1998, Hesse 1995). Combining these two approaches may demonstrate whether a proposed transmission route between a pair of infected partners based on epidemiologic data, accurately reflects the underlying transmission pattern of an STD through a population. Additionally, molecular data may reveal unknown or undisclosed sexual partners (Day 1994). This discussion will focus only on bacterial STDs, as they are most relevant to the research described in this thesis. Due to the lengthy incubation period of some viral STDs, such as HIV, the link between molecular epidemiology and sexual network data is much more complex. Also viral mutation rates are frequently much higher than bacterial mutation rates and therefore molecular analysis of viral STDs in a network context is, again, more complex.

Several studies have focussed on the molecular epidemiology of *N. gonorrhoea* within social networks. Older phenotypic techniques for gonococcal typing (auxotyping and serotyping) have not proved sufficiently discriminatory, however (Catlin 1973, Knapp 1984) *opa*-typing has achieved high levels of discrimination and data to date suggests that individual genotypes are unique, unless part of a transmission chain (Ison 1998, O'Rourke M 1995).

Ward *et al* (1996) obtained interview and *opa*-typing data on 215 cases of gonorrhoea from two UK clinics in London and Sheffield. In London, *opa*-typing results revealed a small number of epidemiologically linked individuals and *opa* clusters, suggesting a highly diverse, transient, and generally unlinked network. A different result was observed in Sheffield. *Opa*-typing data for this region was highly correlated to epidemiologic data and did not exhibit high diversity of *opa*-types suggesting that this region was more densely connected and less transient compared to sexual networks in London. In both cases, there was a strong correlation between *opa*-typing and epidemiologic data suggesting that combining methods is advantageous in identifying real sexual links.

Another study in Sheffield for gonorrhoea-infected individuals also incorporated social and genetic techniques (Day, 1998). Social or genetic data alone did not yield as detailed a network as that produced using both techniques together. Microbiological results were able to confirm linkages, but, for discordant linkages, suggested where more detailed social analysis was necessary. Incorporating social data, such as places where infected persons would normally socialize and meet new people, was helpful in extending

possible sexual connections. This information revealed that many of the infected individuals had met at a common bar in their community. Their gonorrhea genotypic profiles were highly diverse suggesting many unidentified links. Targeting areas of socialization may prove valuable for interrupting chains of transmission in this community.

Molecular epidemiology using *C. trachomatis* strains is achieved by genotyping the predominant antigen that defines individual serovars. This antigen, called the major outer membrane protein (MOMP), consists of five domains alternating with four variable domains (Stephens 1987, Yuan 1989). These four variable domains are the basis of serovar assignment, traditionally conducted through serological methods. Newer molecular approaches (nucleotide sequencing analysis) have provided even greater discrimination by identifying serovar subtypes (see chapter 3 for further discussion of these techniques).

To date, molecular epidemiology studies for genital infections by *C. trachomatis* within networks have been limited to pairs of sexual partners (Quinn, 1996). Fifteen couples were enrolled in this study. Serovar typing of epidemiologically-linked infected people revealed identical types within a partnership, verifying the use of these techniques for chlamydia. In our study, we will expand this approach by conducting molecular epidemiology studies of *C. trachomatis* within large networks.



## 1.6 Research Goals

The goals of this research are to demonstrate the ability to generate sexual networks from contact tracing data, to describe the molecular epidemiology of *C. trachomatis* within the context of sexual networks and to describe the demographic, geographic, and behavioral patterns found within sexual networks in Manitoba.

This study is intended to act as a starting point for the application of sexual network analysis to STDs in Manitoba. The findings of this study will identify future research priorities and will potentially provide data necessary to develop new targeted STD control programs in Manitoba.

## CHAPTER 2

### METHODS AND MATERIALS

#### 2.1 STD control in Manitoba and data source for sexual network generation

In 1987, a selective screening program for *C. trachomatis* was initiated in Manitoba. This program recommended the screening of women who had multiple sex partners or recent history of, or exposure to, an STD, and for women undergoing therapeutic abortion or intrauterine device insertion. In Manitoba, STDs are diagnosed and treated by primary care providers, including physicians and nurses practicing in medical clinics and hospitals, and nurses employed by Health Canada, Medical services Branch, in Health centres on First Nation reserves. All chlamydia and gonorrhoea cases identified in the province by both primary care givers and laboratories are required to be reported to Manitoba Communicable Disease Control (MCDC) Unit. Once these cases have been identified and reported, the primary care givers may also interview clients for information on sex partners and complete the management of these sex partners. In addition, case interview and contact notification could be accomplished by public health nurses employed by either the City of Winnipeg, the province of Manitoba, or Medical Services Branch, Health Canada depending on the client's jurisdiction. Information on cases and their contacts are entered into the STD surveillance database of the MCDC unit. Demographic information is included in this database for all cases and includes age, gender, geographic residence, aboriginal status, known alias, dates of diagnoses, and pathogen identified. A second contact database is maintained which contains the names

of the sexual partners of infected cases, nominal information on the case who named them, contact birth date, gender, address and whether the contact was located, tested and treated, the dates during which exposure took place, and the disease to which the contact was exposed. This database provides a provincial overview of chlamydia and gonorrhea infection and transmission in Manitoba and allowed for the generation of the networks used for the present study.

## **2.2 Data Management**

For the purpose of this study, a 6 month block of data was used from the case and contact STD registry from November, 1997 to May 1998. In addition, data for three prior months were obtained from August, 1997 to October, 1997 in order to include contacts entered into the database after November but named by cases prior to that time. Within this database, there would be individuals who were named multiple times, but who could have provided slightly different nominal information each time they were contacted. For this reason an algorithm was devised in order to resolve this issue. Records were considered to belong to the same individual if the first and last names (or alias) matched exactly and a) two of the three components of the birth date (day, month, year) were identical or b) month and day were identical but reversed, or c) address including house or apartment number were identical. On occasion, if an individual's first name was shortened or last name was spelled differently, they were also considered to be the same person if one of the above criteria (a,b,c) held.

### **2.3 Manitoba Sexual Network construction**

Once the data management was completed and multiple records had been verified as belonging to the same individual, unique numerical identifiers of 5 or 6 digits were assigned to contacts and cases, respectively. A new spreadsheet file was created which included these unique numbers and was formatted as a line-listing of all pairs of connected cases and contacts. This was then entered into PAJEK (Batagelj, 1998), a social network analysis program capable of identifying direct and indirect connections between all of the individuals in the dataset. From the original STD database from the MCDC unit, the demographic data retained included an individual's gender, date of birth, infection status, and geographic residence. Geographic residence for individuals outside Winnipeg was identified as a city, town or First Nation Reserve. For addresses that were within the city limits of Winnipeg, postal codes were used to provide a finer breakdown of network connections between different parts of the city.

### **2.4 Specimen Collection and DNA extraction**

During the time of this study, Cadham Provincial Lab was conducting approximately 60,000 diagnostic tests per year for chlamydia and gonorrhea. This accounts for approximately 94% of the annual chlamydia and gonorrhea testing performed in the province. Diagnostic tests utilized at this time were Chlamydiazyme (Abbott Laboratories) and Gonorrhea PACE 2 (GenProbe). Corresponding to the six month time frame of information collected from the MCDC database, all of the positive

Chlamydiazyme specimens identified at CPL were retained and frozen at -20 C. Specimens were removed from storage as required for DNA extraction and nucleotide sequence analysis.

DNA extraction was done using QIAamp DNA minikits (Qiagen, Inc., Mississauga, ON) according to the manufacturer's instructions. For each extraction, 50 µl of specimen was used. In some cases, urine specimens were available for extraction rather than a Chlamydiazyme swab specimen. In these instances, 50 µl of urine was used and treated in an identical manner as the Chlamydiazyme swab specimens. The extracted DNA was resuspended in 50 µl of sterile distilled H<sub>2</sub>O and stored at -20°C until DNA was required for PCR.

## **2.5 PCR amplification**

PCR was conducted to isolate the *omp1* gene of *C. trachomatis* which encodes the major outer membrane protein. PCR and sequencing primers have been previously described (Yuan *et al.*, 1989) and are listed in table 2.1. Primers were obtained from the Regional DNA synthesis Laboratory, University of Calgary, Alberta. For the initial PCR reaction, primers labeled CT1 and CT2 were used to amplify a 1142 bp fragment containing all four variable domains of *omp1*. Nested PCR using internal primers CT3 and CT4 amplified a 879 bp fragment.

Table 2.1: Summary of primers used for PCR amplification and DNA sequencing of the *C. trachomatis omp1* gene. All sequences are 5' to 3'.

Primer	Primer sequence	Strand	Nucleotide position
CT1	GCCGCTTTGAGTTCTGCTTCCTC	Sense	34-56
CT2	ATTTACGTGAGCAGCTCTCTCAT	Antisense	1145-1167
CT3	TGACTTTGTTTTTCGACCGTGTTTT	Sense	199-216
CT4	CCGCAAGATTTTCTAGATTTTCATCTTGT	Antisense	1048-1076
CT6	CTT(GT)A(TC)TTTAGGTTTAGATTGAGC	Antisense	649-673
CT7	CCTTACATTGGAGTTAATGGC	Sense	859-881

For the PCR reaction, 100  $\mu$ l of the reaction mixture contained 30 pmol of a given primer, 0.2 mM dNTP's, 10  $\mu$ l of DyNaZyme<sup>TM</sup> EXT optimized buffer (MJ research, Waltham,NC), 1.5 mM MgCl<sub>2</sub>, and 2.5 units of DyNAzyme EXT DNA polymerase. The initial PCR reaction contained 5  $\mu$ l of extracted Chlamydiazyme (or urine) DNA. Nested PCR was accomplished by using 0.5-2  $\mu$ l of the initial PCR reaction mix. For the initial PCR, the first round of PCR consisted of heating the mixture at 95°C for 7 minutes followed by 9 cycles of denaturation (95°C, 1 min), annealing (60°C, 1 min) and extension (72°C, 1.5 min), and an additional 24 cycles with the annealing temperature reduced to 55°C. Conditions were the same for the nested PCR however the number of cycles for the latter part of the program was increased from 24 to 30. Final PCR products were stored at - 20°C until needed for precipitation.

PCR products were purified by ethanol precipitation. Fifty  $\mu\text{l}$  of PCR reaction was added to 125  $\mu\text{l}$  of 95% ethanol and one  $\mu\text{l}$  of 3M sodium acetate, pH 3.5. The mixture was left at room temperature for 15 minutes followed by centrifugation at 14,000 rpm for 20 minutes. Supernatants were removed and 500  $\mu\text{l}$  of 70% EtOH was added and centrifuged at 14,000 rpm for 15 minutes. Following the removal of the supernatants, the pellets were air dried and resuspended in 500  $\mu\text{l}$  of  $\text{dH}_2\text{O}$  and stored at  $-20^\circ\text{C}$  until needed for nucleotide sequencing reactions.

## **2.6 Nucleotide sequencing analysis**

Sequencing reactions were carried out with purified PCR products using BigDye<sup>TM</sup> Terminators cycle sequencing kits (Applied biosystems, Foster City, CA). The sequencing reaction contained 1-3 ng of template, 1.6 pmol of primer, 2  $\mu\text{l}$  of 5X dilution buffer, 2  $\mu\text{l}$  of Big Dye<sup>TM</sup> termination and water to a final volume of 10  $\mu\text{l}$ . DNA labeling was conducted using a Perkin Elmer GeneAmp 9600 PCR system (Perkin Elmer, BranchBurg, NJ). Thermocycling conditions were  $96^\circ\text{C}$  for 10 seconds,  $50^\circ\text{C}$  for 5 seconds, and  $60^\circ\text{C}$  for 4 mins, for a total of 25 cycles. Labeled products underwent ethanol precipitation as outlined above. The final purified DNA was resuspended in 15  $\mu\text{l}$  of Template Suppressor Reagent (TSR) and denatured at  $95^\circ\text{C}$  for 5 mins.

Prepared specimens were loaded on an ABI genetic analyzer 310 (Applied Biosystems, Foster City, CA). Each sequencing reaction was conducted under the Big Dye<sup>TM</sup> terminator, rapid sequencing chemistry using Performance Optimized Polymer (POP-6),

1X sequencing buffer, and 47cm X 50 µm capillary. Primers CT3 and CT7 were used for forward sequencing reactions, while CT4 and CT6 were used for sequencing in the reverse direction. Comparison of *omp1* sequences was conducted with ALIGN PLUS (Scientific and Educational Software, Durham, NC). Reverse sequencing was only carried out if potential mutations were identified in the forward direction.

Traditional serotyping identifies distinct serovars of *C. trachomatis* (designated A-L) based on amino acid variability in the externally exposed, variable domains (VD) of the major outer membrane protein (MOMP). Sequence analysis of the *omp1* gene, encoding the MOMP protein, from each distinct serovar has identified the corresponding differences in the nucleotide sequence of the VDs of the gene (Yuan *et al.* 1989). Chlamydia genotypes based on sequence analysis are assigned using the same letter designation as that used for serovar status. We used the *omp1* VD DNA sequences reported by Yuan *et al.* (1989) as prototype strains. If DNA sequence variants were identified they were assigned sub-genotype status by identifying the prototype sequence with the highest degree of homology and designating the isolate with a numerical subscript (e.g. D<sub>1</sub>, D<sub>2</sub>, D<sub>3</sub>).

## 2.7 Statistical Analysis

Univariate analyses was conducted for characterization of the study population based on age, gender, partner number, risk area and transmission success. In addition, closeness



and betweenness values, both network measures derived from UCINET, were calculated for cases within networks. Univariate analyses were also carried out for comparisons between chlamydia genotypes (common vs. rare) and network concordancy (with respect to chlamydia genotypes present) with demographic and network characteristics. Descriptive analyses (chi-square and Kruskal-Wallis) were used to identify associations between the variables described above (Epi Info v. 6.01, 1996 © Centers for disease control, Atlanta Ga).

Mixing matrices were computed by cross classifying each partnership by a characteristic of the study participant and the corresponding characteristic of the partner as reported by the study participant. The counts were then converted to proportions by dividing by the total row. To evaluate the degree of assortative (like with like) vs. disassortative (like with unlike) mixing in each matrix, we computed the Q statistic. The formula for Q is  $(\sum_i P_{ij} - 1) / (n-1)$  where  $P_{ij}$  is the diagonal elements of the mixing matrix. If study participants choose their partners at random, then Q would be zero. Negative values of Q indicate disassortative mixing while positive values indicate assortative mixing.

Linear regressions were computed for comparison of component size vs. 3-plex number against positivity to determine if one model is a better predictor than the other. For this analysis, components of size two were not entered as they do not contain any 3-plex structures. For each component, data was entered for size, number of 3-plexes and the number of positive cases identified. JMP IN (v. 3.0, © SAS Institute inc.) statistical package was used to carry out the linear regressions.

## CHAPTER 3

### CONCORDANCY BETWEEN MOLECULAR AND SOCIAL DATA

(This chapter is an excerpt of a paper written for submission to a scientific journal)

#### 3.1 Introduction

In this section, we present an analysis linking *C. trachomatis* molecular genotyping data with sexual network data. The main objectives of this analysis were to examine the genotypic diversity of chlamydia in Manitoba, assess the agreement between sexual network data and molecular data, and analyze the factors associated with discrepancies between the two types of data. As contact tracing data is currently the most readily available source of STD epidemiologic data available to us for this and future studies, and, as Ghani *et al.* (1998) note is routinely available in other areas, we chose this data type to construct the sexual network for comparison with our molecular data. Given the known shortcomings in contact tracing as a data source (*e.g.* potentially incomplete/inaccurate sexual histories), agreement between the contact tracing and molecular data would increase our confidence in secondarily constructing sexual networks from routinely collected data of this type.

## RESULTS

### 3.2 Overview of the Manitoba sexual network

Manitoba is a Canadian province with a population of 1.14 million. The majority of the population (app. 660,000) lives in Winnipeg, the capital and only large urban center in the province. Most of the population outside Winnipeg lives in the southern portion of the province. At the time of this study, 61 aboriginal reserves were present in the province. Eighteen of these reserves in the eastern and northern parts of the province are accessible by air only. Provincially, approximately 10% of the Manitoba population are registered treaty members. This under-represents the number of aboriginal individuals, as not all eligible aboriginals register. Within Winnipeg, approximately 3% of the population are aboriginal.

The complete data set contained 4544 STD cases and contacts with demographic data available for 3565 individuals. The gender breakdown of the study population, for cases and contacts inclusive, was 56% male and 44% female. STD cases were 22% male and 78% female. Reflecting the concentration of the Manitoba population in Winnipeg, 53.6% of individuals in the dataset provided Winnipeg addresses. The sexual network of the 4544 individuals contained 1503 distinct components varying in size from 2 to 82 people (Table 3.1). Two hundred and two cases had not provided any information on named sexual partners, and therefore were not identified as belonging to any components of 2 or more people. The majority of the network consisted of dyads and triads,

accounting for 911 and 366 components, respectively. A detailed analysis of the 23 largest components, ranging in size from 10 to 82 individuals, has been published (Wylie and Jolly, 2001).

### 3.3 Genotypic diversity of chlamydia in Manitoba

In total, we genotyped 359 chlamydia specimens. Of this total, 35 were part of a random sample, 297 were entered in an analysis of concordancy between epidemiologic and molecular data, and 50 were part of a geographic analysis (individual analyses detailed below). The three subsets do not total 359 as some chlamydia specimens were entered in more than one analysis.

The total number of each chlamydia genotype is outlined in table 3.2. The nucleotide sequences of the variant genotypes identified are shown in table 3.3. Prototype genotypes identified were D, E, F, G, and J. Variants of 3 of these prototypes were identified and designated D<sub>1</sub>, D<sub>2</sub>, D<sub>3</sub>, D<sub>4</sub>, G<sub>1</sub>, and J/C. Additional variant strains were identified for which the respective prototype sequences were not identified in our study population. These variants were designated Ba<sub>1</sub>, Ba<sub>2</sub>, I<sub>1</sub>/H, and K<sub>1</sub>. I<sub>1</sub>/H and J/C indicate *omp1* sequences that are potentially the result of recombination events between *omp1* sequences from two different chlamydia genotypes. VD1 and VD2 of I<sub>1</sub>/H most closely match the prototype I sequence, however, point mutations were present in both VDs (Table 3.3). VD4 of this strain matched prototype H sequences. For genotype J/C, VD1

Table 3.1: Summary of the number of components of size  $n$  identified within the Manitoba sexual network. The number of cases and contacts within the database was 4544. A total of 1503 components with 2 or more people were identified.

Number of individuals	Number of components	Number of individuals	Number of components
2	911	12	1
3	366	13	2
4	116	14	6
5	39	16	1
6	20	17	1
7	15	18	1
8	8	19	1
9	5	39	1
10	2	41	1
11	5	82	1

Table 3.2: Summary of the number of *C. trachomatis* genotypes identified in Manitoba and their abundance. In total, 359 specimens were analyzed for this study (column A). Thirty-five of these specimens were initially selected as part of a random sample (B), 297 were included in the analysis of concordancy between molecular and epidemiologic data (C), and 50 were included as part of the geographic analysis (D).

Genotype	A	B	C	D
Ba <sub>1</sub>	2	1	0	1
Ba <sub>2</sub>	9	0	9	0
D	73	9	54	17
D	37	2	37	0
D	2	1	2	0
D	1	0	1	0
D	1	0	1	0
E	93	10	73	15
F	41	3	36	4
G	4	1	4	0
G	10	0	10	0
I <sub>1</sub> /H	14	2	10	4
J	44	4	37	5
J/	1	0	1	0
K	27	2	22	4
	<b>359</b>	<b>35</b>	<b>297</b>	<b>50</b>

**Table 3.3.** Summary of the nucleotide sequences of the variant *C. trachomatis* genotypes identified in Manitoba. Shown are portions of the variable domains of the *omp1* gene where point mutations occurred within a given genotype. Prototype sequences are shown on the upper line, while the point mutations found in the variant strains are shown beneath the prototype sequence. Numbers indicate the nucleotide sequence position of the respective prototype *omp1* sequence.

Variant genotype designation	VD 1	VD 2	VD4
D <sub>1</sub>	244 256 TTTCAGATGGGTG C		
D <sub>2</sub>		482 491 ATAATGAAAA	995 1005 CTGGCGCAGAG A
D <sub>3</sub>	244 256 TTTCAGATGGGTG C	482 491 ATAATGAAAA	995 1005 CTGGCGCAGAG A
D <sub>4</sub>	244 256 TTTCAGATGGGTG C		973 984 GCTGGAGCTGGC A
G <sub>1</sub>		481 491 GATGGTGAAAA A	
Ba <sub>1</sub>		507 527 AAATAGTACGTTTGTACCAAA G G	
Ba <sub>2</sub>		507 527 AAATAGTACGTTTGTACCAAA G	
I <sub>1</sub> /H	267 275 AAGGATGTA A	514 522 AAGCTTGGTT A	
K <sub>1</sub>			966 975 GACCATCAC G

and VD4 appeared to match prototype J sequences, while VD2 matched the prototype C sequence. The nucleotide sequence of this strain is not listed in table 3.3 as only one specimen was identified with this sequence and the minimal volume of this specimen prevented us from fully verifying the nucleotide sequence.

### **3.4 Concordancy between epidemiologic and molecular data**

We assessed the extent to which a given component, constructed from routinely collected contact tracing data, was concordant (containing only one chlamydia genotype) or discordant (containing two or more chlamydia genotypes). Concordancy between the sexual links established by contact tracing and the molecular data would be consistent with a component tracing the transmission route of a single type of chlamydia.

To establish a baseline for genotype diversity in the province, we had first randomly selected 35 of our stored specimens for genotype analysis. Table 3.2 shows the diversity of chlamydia genotypes found within this sample. In this first sample, the most common genotype, E, accounted for 28.6% of the sample, while the rarest genotypes, Ba<sub>1</sub>, D<sub>2</sub>, and G, each totaled 2.9% of the random sample. This initial data verified that genotypic diversity did exist within the chlamydia circulating in the province and concordancy between epidemiologic and molecular data should not simply occur due to a limited number of circulating genotypes.



We examined the entire sexual network to identify all components which had at least two specimens available for genotyping. In total, 124 components were identified, containing 297 specimens for genotyping. The total number of each genotype within this sample is shown in table 3.2 and largely reflects the diversity seen in our random sample.

Figure 3.1 illustrates the extent to which components were concordant with respect to chlamydia genotypes circulating within them. Overall 90 of 124 (72.6%) components contained only one chlamydia genotype (3 components containing both prototype D and variant D genotypes were classed as concordant - discussed below). Concordancy dropped as a function of component size, ranging from 92.6% for components of size 2, to 24% for the largest components containing 8 or more people.

For the two largest components with a relatively large number of specimens available for analysis, two types of discordant results were evident. Fig. 3.2 shows component 23 with 82 people and 11 available specimens, and component 30 with 18 people and 7 available specimens. Although designated discordant, component 23 does appear to have largely captured the circulation pattern of two chlamydia genotypes, E and J. The former type appears to have been circulating largely in the Northern Manitoba portion of this component and may have crossed a geographical bridge into (or out of) Winnipeg by way of one central individual, while the J genotype was circulating in a subset of the Winnipeg individuals.

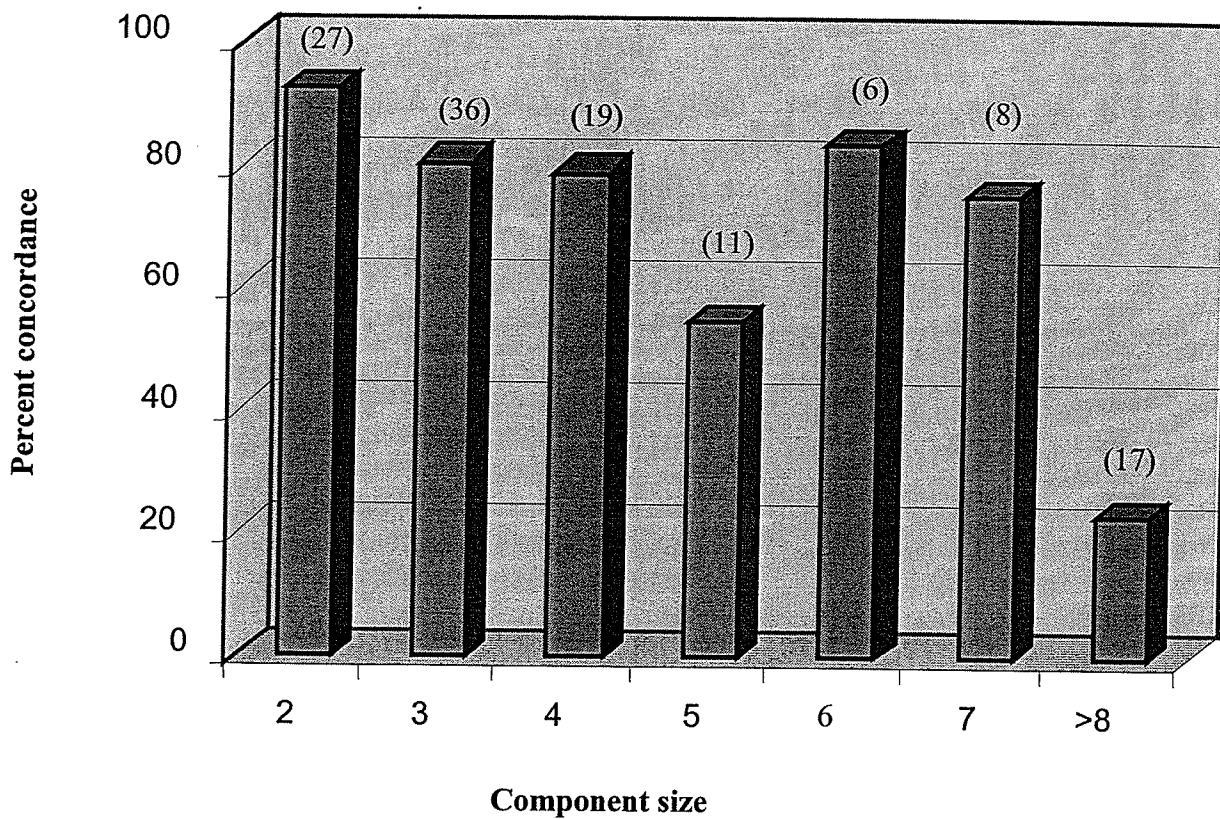


Figure 3.1: Graph illustrating the extent of concordancy for different sizes of components. Concordant components were those containing only one chlamydia genotype. Discordant components contained two or more chlamydia genotypes. The number of components in each category is shown above the respective bar. The average number of available specimens for components of size >8 was 3.9; size 7, 2.8; size 6, 2.3; size 5, 2.3; size 4, 2.1; size 3, 2.1. For components of size two only those components with two available specimens were analyzed.

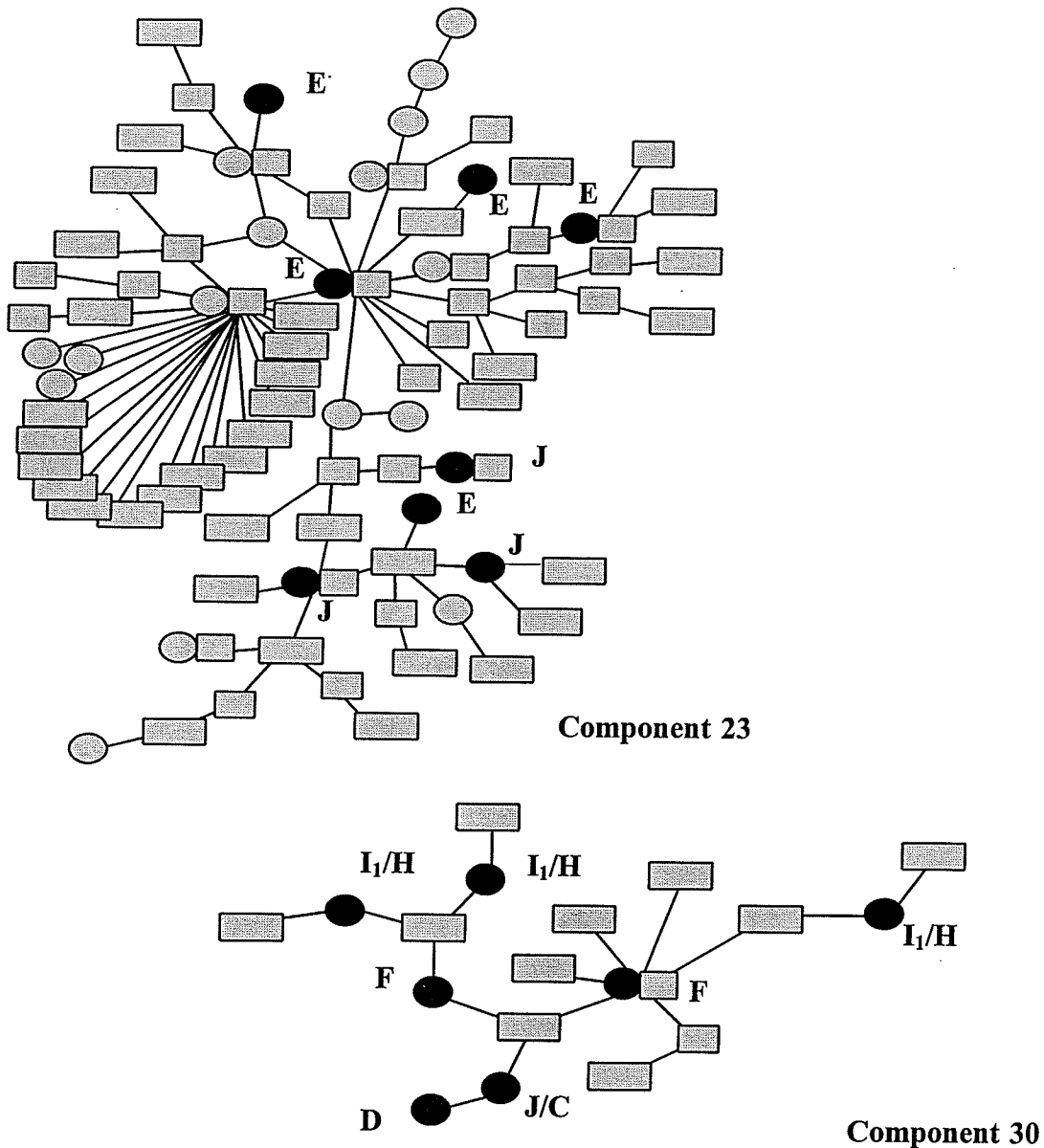


Fig. 3.2. Illustration of the distribution of chlamydia genotypes within two of the largest components. Component 23 contains 82 people. Component 30 contains 18 people. For its size, component 23 had a limited number of specimens available, as many of the cases were diagnosed in the 3 month period preceding storage of chlamydia-positive Chlamydiazyme specimens. Symbols: chlamydia-infected cases from which a chlamydia specimen was available for genotyping, ●; chlamydia-infected cases from which a specimen was not available, ○; gonorrhea-infected cases, □; co-infected cases, ●□ or ○□; individuals with negative or unknown test results, □. Letters represent the chlamydia genotypes isolated from a given case.

In contrast, component 30, with 18 people, contains 4 different chlamydia genotypes with no clear pattern to their distribution within the component. In this case, the contact tracing data upon which this component was built appears to fall short of fully capturing the transmission routes of the chlamydia infecting these individuals.

We conducted a univariate analysis to determine if there were any significant differences between concordant and discordant components (Table 3.4). Five variables were analyzed, with a detailed description of each variable provided in table 3.4. Briefly variables were based on component size and on the co-infection status, partner number, ethnicity, and infection dates of the individuals within a given component. Discordant components were larger, and had a higher likelihood of containing an individual that was aboriginal, co-infected with gonorrhea and chlamydia, or who had named 3 or more sex partners.

### **3.5 Geographic patterns**

Geographic approaches have been presented as a possible means of targeting core groups for STD control programs (Zenilman *et al.*, 1999; Becker *et al.*, 1998; Han *et al.*, 1999). Our previous analysis of large components in Manitoba suggested that geographical bridging via sexual contact may frequently connect STD core groups in different areas (Wylie and Jolly, 2001). These latter results suggest that a coordinated approach spanning different geographic areas may be necessary when developing a targeted STD control program.

Table 3.4. Univariate analysis of the variables associated with discordant genotyping results within components. A component was classified as “co-infected” if at least one person in the component was co-infected with chlamydia and gonorrhea. A component was classified as “aboriginal” if at least one person was of aboriginal descent (ethnicity information in our dataset permitted only this limited type of categorization). Components were classified as one partner, two partners, or 3 or more partners, based on the largest number of sexual partners named by any case within the component. Time interval is the average length of time between the earliest and latest diagnosis dates within components. “Component size” is the average number of individuals within components.

Variable	Number of Discordant components (percent)	Number of concordant components (percent)	p value
Co-infected chlamydia infected	12 (35) 22 (64)	5 (6) 85 (94)	0.00002*
Aboriginal Non-aboriginal	26 (76) 8 (23)	33 (37) 57 (63)	0.00008*
1 partner 2 partners ≥3 partners	2 (6) 12 (33) 20 (61)	25 (28) 38 (42) 27 (30)	0.004*
Time interval	53.6 days	34.5 days	0.188
Component size	10.8	3.9	0.00001*

We conducted an analysis to determine the extent to which sexual contact between individuals living in different geographic areas may contribute to successful bridge transmission of chlamydia. Of the components spanning more than one geographic area, we determined which contained only one chlamydia genotype. In total, 55 components spanned two or more geographic areas. For 34 of the 55 components specimens were available from the different communities represented within a given component. Of these 34 components, 20 (59%) were concordant across communities with respect to genotype. This data suggests that many of the geographic linkages identified by network analysis do potentially reflect successful transmission.

We conducted a preliminary geographic analysis of genotype data to assess the genetic diversity of chlamydia present in small population centres. We analyzed all of the chlamydia specimens available in our stored collection from 7 towns and aboriginal reserves in Manitoba (50 specimens; Table 3.2), ranging in population from 1,212 to 15,145 people. The number of chlamydia genotypes varied from 4 in communities B and E (populations 8,950 and 1,212, respectively) to 5 in communities D and C (population 2,126 and 8,039, respectively). This data suggests that chlamydia diversity in small communities is relatively high, and that more than one chlamydia genotype simultaneously circulates in these small towns or reserves.

### 3.6 Discussion

The sexual network we analyzed was based on routinely collected data for 4544 STD cases and contacts. In total, the network contained 1503 components with the largest containing 82 people. Some initial analysis of this data has been published (Wylie and Jolly, 2001; Jolly and Wylie, 2001). Here, we describe a linkage of chlamydia genotype data with sexual network data. To our knowledge, this is the first study to examine the molecular epidemiology of *C. trachomatis* within the context of large sexual networks.

The main objective of this analysis was to observe the extent to which molecular typing data for *C. trachomatis* would agree with the proposed transmission routes within the sexual networks constructed from contact tracing data. The majority of the components containing individuals linked either directly or indirectly by sexual contact were concordant with respect to chlamydia type, indicating that network analysis based on routinely collected data frequently does reflect the transmission routes of individual chlamydia types. Additionally, of the components which contained individuals from different geographic areas and for which specimens were available, 59% were concordant with respect to chlamydia type. This data indicates that despite the physical distances involved, the geographic connections identified by network analysis do frequently reflect transmission events. Network analysis could therefore be used to identify the different areas in a region where it would be useful to coordinate STD control efforts.

Several factors were associated with the presence of two or more genotypes within a component. These components were more likely to be larger in size, or to contain at least one individual who was aboriginal, co-infected with chlamydia and gonorrhoea, or who named 3 or more sex partners. Size, co-infection status, and partner number would be expected to be surrogate measures of core group status. In network terms, STD core groups have been defined as groups of individuals with high rates of sex partner change and a high degree of interconnectivity between those individuals (Brunham, 1997). A network diagram of a core group would therefore tend to be large, reflecting the higher number of sexual connections between individuals with multiple sex partners. Similarly, Jolly (1998) identified co-infected individuals as central to core groups, therefore components containing both chlamydia and gonorrhoea would also be more likely to potentially represent all or part of a core group. Within the core, the high number of sexual pathways would create multiple transmission routes for different genotypes of chlamydia (or other STDs), hence components that include core group members would be more likely to contain criss-crossing transmission pathways of different chlamydia types. This scenario was evident in fig. 3.2, showing the chlamydia types within our largest identified component of 82 people. Although containing two types of chlamydia, E and J, the location of these types within this component suggests that we captured two distinct transmission paths within one component.

The association between multiple chlamydia types within a component and aboriginal status may reflect the difficulties associated with STD control in small communities. Some aboriginal communities in Manitoba are small with populations of less than 1,000



people. This characteristic is true of component 30 in fig. 3.2, which is centered on a small, largely aboriginal, community in Manitoba. The random dispersal of chlamydia types in this component suggests that either additional sexual connections have been missed or delays in locating individuals have put the molecular data out of synch with the social data. Anonymity could be a concern in small close-knit communities and individuals may, therefore, be reluctant to provide a complete list of their sex partners. Thus, components based on contact tracing from these communities would fail to fully capture the actual transmission pathways. Anonymity issues may also contribute to individuals not attending their local medical clinic, but instead waiting until they have an opportunity to seek medical care in a larger urban center. Depending on the rate of sex partner change, these delays may ultimately result in spurious connections being identified with respect to the chlamydia genotypes actually present at the different time points. When a combination of molecular and social data reveals patterns of this type, it may indicate areas or circumstances in which STD control may not be meeting its goals.

Although geographic approaches to STD control have proven effective in lowering incidence rates (Han *et al.*, 1999), this approach does not address the effects of connections between communities. A strict geographic approach focussed on one area could result in an equilibrium being reached between local control efforts and re-seeding of the targeted area by chlamydia from geographically remote, but socially continuous core groups in other areas. Our analysis of chlamydia types in small geographic centres in Manitoba demonstrated that even relatively small population centers do not act as a focus of infection for a single chlamydia type. In combination with the extensive

geographic connections revealed by network analysis (see map in Wylie and Jolly, 2001), the picture that emerges of STD transmission is one of individual communities acting as staging points for the maintenance and transmission of multiple chlamydia types within the context of the larger sexual network in the area.

In conducting this analysis, we generated a large collection of genotype data for chlamydia, consisting of 359 typed specimens. In conjunction with earlier Manitoba data collected by Yang *et al.* (1993), a dataset of this size is useful for addressing the question of the rate at which *omp1* undergoes nucleotide sequence variation. Although immune selection is the likely underlying factor for amino acid and nucleotide sequence variation in *omp1*, the temporal dynamics of this variation has been debated. The isolation of apparently identical variant strains at distant geographic points and the long term persistence of these strains on a global scale implies a slow tempo for *omp1* variation. Conversely, the identification of numerous variable types in a STD core group in Africa (Brunham *et al.*, 1994) and the rapid temporal shifts in the genotypes present in a limited geographic area (Brunham *et al.*, 1996) suggests a more rapid flux in amino acid and nucleotide sequence variation in response to immune pressure.

Within our sample, 15 distinct chlamydia types were identified; 5 types contained *omp1* VD sequences identical to the prototype serovar strains analyzed by Yuan *et al.* (1989) and 10 were variant genotypes. A search in GenBank and the published literature indicated that none of the variants we observed are newly identified. Four of the variants were present in Manitoba 6 years prior to our study, when Yang *et al.* (1993) conducted

their molecular survey of chlamydia types in Manitoba (a fifth variant we identified, J/C, may match the C-like/J-like variant identified by Yang *et al.* [1993], however, we were unable to obtain sufficient nucleotide sequence data to confirm this possibility). An E and D variant type seen in 1992 was absent in our sample, while we identified 5 variants, D<sub>1</sub>, D<sub>2</sub>, D<sub>3</sub>, D<sub>4</sub>, and G<sub>1</sub>, not seen in 1992. Given the smaller sample size of Yang *et al.* (1993), it is uncertain whether these latter 5 types of chlamydia have emerged in Manitoba since 1992, or were already present in low numbers at that time. We identified only 4 components that contained both a prototypic genotype (D), and a variant of that genotype (D<sub>1</sub>, D<sub>4</sub>). This is the only combination of epidemiologic and molecular evidence we found to suggest the generation of *omp1* diversity within the time frame of our study. Overall, this data suggests an endemic presence in Manitoba of several chlamydia genotypes, with little in-migration of new strains and/or little occurrence of *omp1* nucleotide sequence variation in the province between the time points of the two studies. Genotype D chlamydia may be the exception in our study area, as this genotype appears to have undergone more variation than other chlamydia genotypes.

An extensive molecular survey of *C. trachomatis* in Kenya, using *omp1* genotyping (Brunham *et al.* 1994; 1996) obtained results that were somewhat different than in Manitoba. Although some patterns were similar in the two areas (the total number of genotypes was similar [Manitoba, 15; Kenya, 18] with 4 genotypes making up approximately 70% of the total of both samples [Manitoba, genotypes E, D, J, F; Kenya, genotypes E, L<sub>2</sub>, K, L<sub>1</sub>/L<sub>2</sub>] and some temporal changes in genotype prevalence was evident in both locations, with some genotypes either appearing or disappearing or

decreasing or increasing in prevalence over time) the Kenyan site differed from Manitoba in that 62% of the specimens were variant genotypes, while only 29% of the specimens in Manitoba were variants. Additionally, 10 chlamydia genotypes appeared in Kenya over a two year period, while only 5 appeared in Manitoba over a 6 year period. In Kenya, these temporal fluctuations involved 7 chlamydia prototypes and 3 variants of these types (A, D, D<sub>variant</sub>, G, H, I, L<sub>2</sub>, L<sub>2variant</sub>, L<sub>2</sub>/L<sub>1</sub>, M), while in Manitoba most changes were due only to genotype D (D<sub>1</sub>, D<sub>2</sub>, D<sub>3</sub>, D<sub>4</sub>, G<sub>1</sub>).

The lower proportion of variant strains in Manitoba and the lower number of chlamydia prototype strains undergoing variation in Manitoba may reflect the phase of the chlamydia epidemic in this part of North America. In Manitoba, the chlamydia epidemic is believed to be at phase III, "decline" (Blanchard *et al.* 1998), using the terminology of Wasserheit and Aral (1996). The larger number of variant strains and the more extensive temporal shifts in chlamydia genotypes observed in a Kenyan STD core group (Brunham *et al.*, 1994; 1996) may reflect the positioning of the chlamydia epidemic in that country at an earlier, "hyperendemic" phase where transmission events within the core may occur at a more rapid rate and individual chlamydia genotypes could more frequently encounter immunologically-experienced hosts. If so, molecular typing in combination with sexual network analysis may help to pinpoint the epidemic phase of chlamydia within a given setting. Large networks, a high proportion of chlamydia variants in the population, and more frequently observed genotype variation in epidemiologically-linked pairs of sexual partners or within networks, may represent a hyperendemic phase of the epidemic curve as opposed to the relative stability observed in Manitoba. Additional research conducted

in different geographic areas with similar observational designs would be necessary to obtain further evidence to test this hypothesis.

In summary, in this analysis we identified considerable genetic diversity in the number of chlamydia genotypes circulating within the province of Manitoba. Many of these chlamydia genotypes have persisted in the province over approximately a 6 year period. Only in a few instances did we observe what appeared to be the generation of point mutations within the time frame of our study. Genotypic data showed a high degree of concordancy with epidemiologic data, suggesting that routinely collected contact tracing data is a valid means of constructing sexual networks. In some instances a lack of concordancy may reflect the identification of a core group or, alternatively, identify areas where STD control programs may need to be changed or modified. The next chapter provides an overview of our characterization of the Manitoba sexual networks with respect to demographic, geographic and behavioral variables.

**CHAPTER 4**  
**MIXING PATTERNS AND CHARACTERISTICS OF CHLAMYDIA INFECTED**  
**CASES PRESENT IN MANITOBA**

**4.1 Introduction**

This chapter has been divided into two sections, and consists of an analysis of the behavioral and demographic characteristics of individuals within a network context. It serves to display the social environment of cases and their contacts in order for characterization of persons involved in disease maintenance and spread within Manitoba. From chapter three we concluded that the molecular and social data have a high degree of concordancy, thereby suggesting that the constructed networks generated from contact tracing data represents an accurate depiction of transmission events.

Characteristics of individuals were analyzed at an individual level as well as in a social network context. The first section of this chapter is related to partner selection, or who mixes with whom, for partnerships found throughout the province, based on mixing matrices computed for several different variables. A mixing matrix is a table that contains information on both a case and their contacts. This matrix allows us to monitor how individuals within a certain population choose their partners based on particular characteristics. For example, determining mixing patterns based on partner number requires the number of partners for a case and their respective contacts. Situations may arise where cases have many sexual contacts and their contacts, in turn, have many partners. This pattern of mixing represents situations in which rapid spread occurs as a

result of the high number of transmission opportunities. On the other hand, a mixing matrix can identify a population in which cases have many partners but these partners have few contacts. This scenario can result in disease spread from cases to their contacts, however, the spread of infections outside these partnerships is limited. Therefore, the elements of a mixing matrix represent the proportions of individuals with certain characteristics who form partnerships with individuals with the same or different characteristics. Mixing matrices are what we are interested in estimating to acquire knowledge about mixing patterns. Several documented studies have reported the importance of mixing patterns on disease spread and prevention (Laumann and Youm, 1999; Aral *et al.*, 1999; Rothenberg, 1983). These studies have shown how strong assortative mixing can result in an initial infection spreading quickly through a population, followed by the epidemic rapidly declining. A population that mixes dissortatively has a slow initial onset of an epidemic, however, that epidemic lasts for a longer period of time.

The second section of this chapter focuses on general population characteristics of the whole data set of infected cases to provide baseline data for the province. Here we attempt to find distinguishable characteristics for different segments of the infected population. For example, what age groups are mostly affected by STD infections, what segments of the population have more sexual partners resulting in high levels of STD incidence and what are the characteristics of infected persons within high and low risk areas. Results from this chapter will provide information on the behaviors of chlamydia infected individuals, their contacts, and information on network characteristics of the

population during the period that the data was collected for this study. This data is important in understanding the sexual network dynamics of the Manitoba population, which will aid in future epidemiologic comparisons and development of effective prevention and control programs.

## RESULTS

### 4.2 Mixing patterns among study participants and their sex partners

This section was conducted to further our understanding with respect to partner selection of STD-infected persons. For this analysis, mixing matrices were constructed. As many of the cases have more than one sexual partnership, each partnership was entered individually into the matrix. The Q statistic, a measure of within class mixing, is defined as  $(\sum p_i - 1)/(N - 1)$ , where  $p_i$  is the proportions from the diagonal elements of the matrix and N is the number of rows present in the matrix (Aral *et al.*, 1999). This was used to measure the degree of assortative and disassortative mixing within the population. Several studies have revealed how information on mixing patterns of partnerships is important for targeting groups that are most likely responsible for disease acquisition and spread (Wasserheit and Aral, 1996; Rothenberg, 1996).



#### **4.2.1 Stratification of the data**

Following construction of an overall matrix for a given variable (data available allowed age groups and partner number to be analyzed), the matrix was first constructed for the population as a whole, and then stratified by gender and risk area.

Manitoba has three areas in the rural part of the province as well as four areas in the city of Winnipeg that we classified as high risk areas. This classification was based on the chlamydia incidence levels for those areas (Figure 4.1; for a detailed description of the geographic breakdown in Winnipeg, see Blanchard *et al.*, 1998). Area boundaries in rural areas are based on regional health authority (RHA) boundaries. Areas within Winnipeg are based on forward sortation areas. The high risk rural areas include the RHAs of Burntwood, Norman and North Eastman. The forward sortation areas in Winnipeg are R2W, R3A, R3B and R3H (Blanchard *et al.*, 1998). The remainder of the geographic areas in the province were categorized as low risk areas. This analysis was conducted to monitor whether there are differences in sexual mixing patterns between these two areas. Each case for which place of residence data was available was included in the analysis.

#### **4.2.2 Mixing across age categories**

In total, information on 966 partnerships was available where the age of both the case and their respective partners were known. A mixing matrix was constructed based on three categories of age for both case and contact. The pattern of mixing between cases and their reported partners by age was assortative.

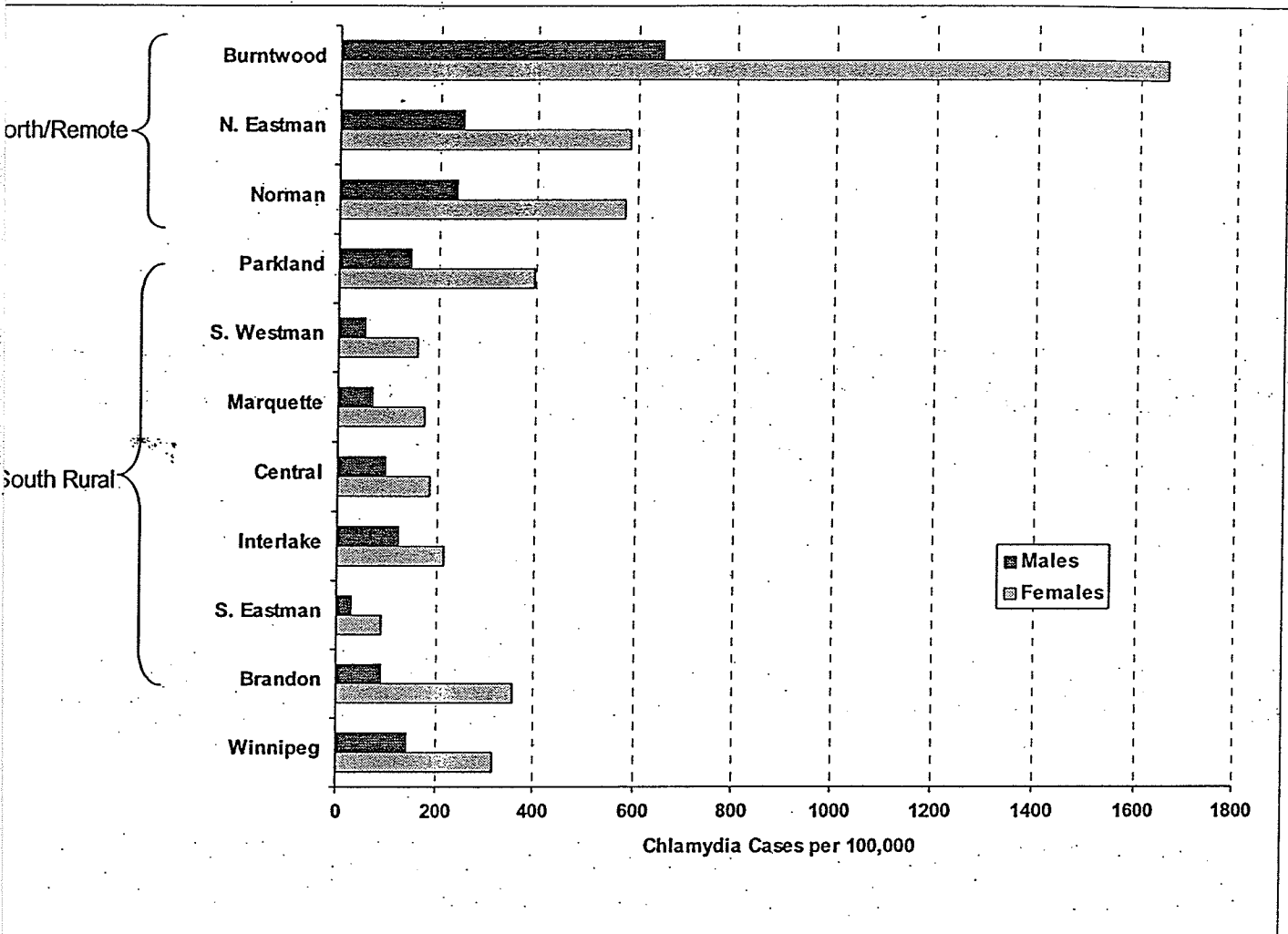


Figure 4: 1999 incidence levels of Chlamydia in Manitoba by RHA. Norman, Burntwood and North Eastman have the highest incidence levels and are categorized as high risk areas. (Data obtained from Debbie Nowicki, CDC Unit in Manitoba).

Cases <20 and between 20-29 years of age had a majority of their partnerships with people in their own age group while, individuals >29 years of age reported having similar numbers of partners in the same group (44%) and in a younger age group (49%)(Table 4.1A). Overall, the Q statistic for the matrix was  $Q = 0.34$  reflecting the assortative relationship outlined above.

Stratification by gender revealed several differences between age based partnerships for males and females (Table 4.1B). Both male and female cases tend toward assortative mixing, with a Q of 0.37 and 0.34, respectively. However, female cases tend to choose older partners compared to males, hence the lower Q value. For example, 44% of female cases <20 years of age, report having contacts who are between 20 and 29 years of age, while the same characteristic is reported by only 17% of males.

When further stratified by risk area some differences in mixing patterns appear to be present for individuals in the two areas (Table 4.1C). Cases were divided into two groups based on their geographic location. In total there were 891 cases with place of residence data and age data available. Of these, 375(42%) were in high risk areas and 516(58%) were in low risk areas. Males in high risk areas show stronger assortative mixing than males in low risk areas (0.41 vs. 0.32, respectively). Females, in contrast, show less assortative mixing than females in low risk areas (0.32 vs. 0.37, respectively).

Table 4.1 : Mixing matrices for cases and their contacts by age. Age was categorized into three groups: <20, 20-29 and >29 years of age. A) Overall mixing matrix for cases and their contacts by age. B) Age mixing matrices for age, stratified by gender. C) Age mixing matrices for cases in high risk areas stratified by gender. D) Age mixing matrices for cases in low risk areas stratified by gender. Q values are located directly below the respective matrix.

Contact age		Case characteristic					
		<b>Age</b>					
A)		<b>&lt;20</b>	<b>20-29</b>	<b>&gt;29</b>			
	<b>&lt;20</b>	151(56%)	140(24%)	7(7%)			
	<b>20-29</b>	106(40%)	393(67%)	55(52%)			
	<b>&gt;29</b>	11(4%)	56(10%)	43(41%)			
		<b>268</b>	<b>589</b>	<b>105</b>			
		Q= 0.32					
		<b>Gender</b>					
B)		<b>Males</b>			<b>Females</b>		
		<b>&lt;20</b>	<b>20-29</b>	<b>&gt;29</b>	<b>&lt;20</b>	<b>20-29</b>	<b>&gt;29</b>
	<b>&lt;20</b>	43(86%)	64(38%)	5(14%)	109(50%)	76(18%)	2(3%)
	<b>20-29</b>	7(14%)	97(58%)	21(58%)	99(45%)	296(70%)	33(48%)
	<b>&gt;29</b>	0	7(4%)	10(28%)	11(5%)	48(11%)	34(49%)
	<b>50</b>	<b>168</b>	<b>36</b>	<b>219</b>	<b>420</b>	<b>69</b>	
		Q= 0.36			Q=0.34		
		<b>High risk area</b>					
C)		<b>Males</b>			<b>Females</b>		
		<b>&lt;20</b>	<b>20-29</b>	<b>&gt;29</b>	<b>&lt;20</b>	<b>20-29</b>	<b>&gt;29</b>
	<b>&lt;20</b>	24(92%)	35(45%)	4(24%)	43(44%)	41(25%)	0
	<b>20-29</b>	2(8%)	37(47%)	8(47%)	49(51%)	113(66%)	20(54%)
	<b>&gt;29</b>	0	6(8%)	5(29%)	5(5%)	16(9%)	17(46%)
	<b>26</b>	<b>78</b>	<b>17</b>	<b>97</b>	<b>170</b>	<b>37</b>	
		Q= 0.59			Q=0.30		

Table 4.1, continued

Low Risk Area						
D)	Males			Females		
	<20	20-29	>29	<20	20-29	>29
<20	19(79%)	29(32%)	2(12%)	66(54%)	35(14%)	2(5%)
20-29	5(21%)	60(67%)	10(59%)	50(41%)	183(73%)	20(51%)
>29	0	1(1%)	5(29%)	6(5%)	32(13%)	17(44%)
	24	90	19	122	250	39
	Q=0.36			Q=0.36		

Overall, the female pattern of greater age mixing in high risk areas may serve to disperse the epidemic over a greater range of ages and hence, increase incidence.

#### 4.2.3 *Mixing across sexual activity class*

The potential for the rate of spread of an STD is highly embedded in characteristics such as number of sexual partners. The number of partners that a case has, and in turn, the number of partners that their contacts have, is an essential component in determining the rate of spread and the course of an epidemic.

In our study, contacts who had not become cases and had been named by only one case were positioned at the peripheral points within networks. These individuals either tested negative or were not located by a public health nurse. In either case, since they have not been interviewed by a public health nurse, we have no means by which we can assess how many sex partners they had, and therefore, no basis for placing them at a given

location in the matrix. This would potentially under represent the number of cases whose sex partner in turn had no other partners. However, as one of our main interests is to compare the Q statistic for partner number mixing between high and low risk areas (see below), we would at least carry the same limitation throughout the comparison.

Table 4.2A shows the mixing patterns between the different sexual activity classes for the entire dataset. Data was grouped into three sexual activity classes for cases or contacts with one, two or greater than two partners. In total, 825 partnerships were entered for this analysis. It appears that, contrary to the assortative mixing seen for the different age groups, overall, relationships based on partner number approaches proportionate mixing (i.e. partnerships are formed randomly;  $Q=0.025$ ). For the majority of the data, cases in all sexual activity classes were connected to partners who had two sexual partners, reflecting the limitation outlined above.

The same analysis of mixing by partner number was conducted following stratification by gender (Table 4.2B). There was 825 partnerships entered for this analysis with 516 (63%) females and 309(37%) males. Results were similar to the findings above. There appears to be a slight trend for male cases to exhibit more assortative mixing than female cases ( $Q=0.045$  vs  $Q= 0.005$ , respectively).

As above, the data was further stratified by risk area (Table 4.2C and 4.2D). In total, place of residence data was available for 568 cases with 284 (50%) representing high risk areas and 284 (50%) representing low risk areas. For both males and females, high risk area cases tend towards more assortative mixing than low risk area cases. Q values for

Figure 4.2: Mixing matrices for cases and their contacts based on partner number. Individuals were categorized as having one, two or greater than or equal to three sexual contacts. A) Overall partner number mixing matrix for the dataset. B) Partner number mixing matrices stratified by gender. C) Partner number mixing matrices for cases present in high risk areas stratified by gender. D) Partner number mixing matrices for cases in low risk areas stratified by gender.

<b>Contact partner number</b>		<b>Case characteristic</b>					
A)		<b>Partner number</b>					
		<b>1</b>	<b>2</b>	<b>≥3</b>			
<b>1</b>		58(19%)	63(23%)	45(18%)			
<b>2</b>		174(57%)	137(51%)	118(47%)			
<b>≥3</b>		71(23%)	71(26%)	88(35%)			
		<b>303</b>	<b>271</b>	<b>251</b>			
		<b>Q=0.025</b>					
		<b>Gender</b>					
		<b>Males</b>			<b>Females</b>		
		<b>1</b>	<b>2</b>	<b>≥3</b>	<b>1</b>	<b>2</b>	<b>≥3</b>
B)	<b>1</b>	29(24%)	29(29%)	21(24%)	29(16%)	34(20%)	24(15%)
	<b>2</b>	67(55%)	45(45%)	32(36%)	107(59%)	92(53%)	86(53%)
	<b>≥3</b>	25(21%)	25(25%)	36(40%)	46(25%)	46(27%)	52(32%)
		<b>121</b>	<b>99</b>	<b>89</b>	<b>182</b>	<b>172</b>	<b>162</b>
		<b>Q=0.045</b>			<b>Q=0.005</b>		
		<b>High risk area</b>					
		<b>Males</b>			<b>Females</b>		
		<b>1</b>	<b>2</b>	<b>≥3</b>	<b>1</b>	<b>2</b>	<b>≥3</b>
C)	<b>1</b>	9(26%)	11(30%)	7(16%)	11(24%)	16(24%)	12(21%)
	<b>2</b>	18(51%)	17(46%)	11(24%)	25(55%)	37(56%)	25(45%)
	<b>≥3</b>	8(23%)	9(24%)	27(60%)	9(20%)	13(20%)	19(34%)
		<b>35</b>	<b>37</b>	<b>45</b>	<b>45</b>	<b>66</b>	<b>56</b>
		<b>Q=0.16</b>			<b>Q=0.12</b>		

Table 4.2, continued

D)	Low Risk Area					
	Males			Females		
	1	2	≥3	1	2	≥3
1	10(22%)	11(28%)	11(33%)	28(40%)	22(45%)	11(19%)
2	24(53%)	24(60%)	16(48%)	28(40%)	18(37%)	23(40%)
≥3	11(25%)	5(12%)	6(18%)	14(20%)	9(18%)	23(40%)
	45 Q=0.0	40	23	70 Q=0.09	49	57

high risk males and low risk males are 0.16 and 0.0, respectively, while high risk females and low risk females are 0.12 and 0.09, respectively.

#### 4.3 Characterization of the chlamydia infected population within Manitoba

Traditional epidemiologic studies in Manitoba and elsewhere, have focussed on the individual as the unit of analysis. In contrast to these studies, our data is network-based. In this section, we present data intended to provide baseline measures for individual and network variables for chlamydia infected cases in Manitoba. This will help to provide baseline preliminary measures of our local population and will serve as a basis for comparison for both future studies and for comparing against future STD incidence trends. Demographic data available for incorporation into this analysis included three variables: age, gender and geographic location.



#### *4.3.1 Case characteristics based on partner number*

It has been argued that the course of an epidemic is partially a result of the number of partners cases have within a population. Several studies have demonstrated differences within and between sub-populations and how high numbers of partners contained within some groups are responsible for disease maintenance and spread (Potterat *et al.*, 1999; Brunham, 1997; Anderson, 1992). In this section, we provide data on our infected population in Manitoba grouped by age, gender and risk area for the number of sex partners reported by each group.

Table 4.3A shows the number of reported partners corresponding to the respective age group of chlamydia infected cases. For this analysis, and subsequent analyses, three categories of age groups, <20, 20 to 29 and >29 years of age were analyzed. In total, 1004 cases had age data available. A  $\chi^2$  statistical analysis revealed that there is a significant difference between age and the number of partners that a case reports ( $P = 0.029$ ). Eighteen percent of cases <20 years of age and 14% of cases between 20 to 29 years of age reported greater than two partners as compared to only 7% of cases >29 years of age. The majority of cases, irrespective of age, reported one partner ranging from 53% in the <20 group to 70% in the >29 group.

Table 4.3: Characterization of chlamydia infected cases based on partner number. Cases were grouped as having one, two or greater than or equal to three sex partners. Data was analyzed for A) age groups, B) gender and C) risk area categories.

Case partner number					P value
		<b>Age</b>			
		<b>&lt;20</b>	<b>20-29</b>	<b>&gt;29</b>	
A)	<b>1</b>	138(53%)	331(58%)	100(70%)	0.029*
	<b>2</b>	74(29%)	158(28%)	33(23%)	
	<b>≥ 3</b>	46(18%)	78(14%)	10(7%)	
		<b>258</b>	<b>567</b>	<b>143</b>	
		<b>Gender</b>			
		<b>Males</b>	<b>Females</b>		
B)	<b>1</b>	307(56%)	875(63%)		0.029*
	<b>2</b>	167(30%)	355(26%)		
	<b>≥ 3</b>	74(13%)	153(11%)		
		<b>548</b>	<b>1383</b>		
		<b>Risk area</b>			
		<b>High</b>	<b>Low</b>		
C)	<b>1</b>	242(56%)	321(60%)		0.265
	<b>2</b>	117(27%)	149(28%)		
	<b>≥ 3</b>	71(17%)	68(13%)		
		<b>430</b>	<b>538</b>		

Partner numbers for cases were also assessed by gender (Table 4.3B). Gender data was available for the majority of our dataset (1141 of 2120 cases). For unknown genders, in this analysis and subsequent analyses, assumptions were made if the genders of adjoining cases and contacts were known. Our results illustrate a significant dissimilarity in that female cases tend to report fewer partners, with 63% reporting only one partner compared to 56% reported by males ( $\text{Chi}^2$ ,  $P= 0.029$ ).

The number of partners for cases present within high or low risk areas was also examined. In total, there was 968 cases with recorded place of residence, 44% of these were present in high risk areas and 56% present in low risk areas. The categories for case partner number were kept identical to those above, such that a case can have one, two or greater than or equal to three contacts. This data is shown in table 4.3C, where there was no significant difference for high and low risk areas with respect to the mean partner number ( $P=0.265$ ). The majority of cases in either area reported one partner, while 17% in high risk areas and 13% in low risk areas had greater than 2 partners.

Overall, cases in our population with the most number of reported sexual partners tended to be younger in age and/or male. These groups may serve as an important foci of disease transmission throughout the province. Analysis of risk area did not reveal any differences such that cases residing in high and low risk areas reported having similar numbers of partners. In the previous section, the analysis based on partner number mixing within the two risk areas revealed a different pattern of partner selection, such that individuals in

high risk areas are more likely to select partners with numerous contacts. It seems that the difference between the two areas does not lie in the average number of partners any given case has, however it is the pattern of partner selection that differs and contributes to the spread of STDs throughout the population.

#### ***4.3.2 Case characteristics based on partner location***

Many studies have demonstrated the differences that exist in terms of the geographic location of sex partners (Rothenberg, 1983; Aral *et al.*, 1999). Sex partners can potentially be chosen from within the same community as that in which an individual lives, from a different community, or both. This concept ties in the idea of mobility within and between communities, which may affect the probability of transmission from one location to another. This also relates to bridging between communities and the linking together of networks. For our next analysis, we evaluated case partner location stratified by gender, age group and risk area. Choosing partners within the same community would help to maintain infections within that community. In contrast, choosing partners in a different community would help to seed STDs into new communities and potentially new networks.

Table 4.4A shows the results of this analysis for cases within different age groups. If multiple partners were named by a case, all contacts with geographic data available were included. Three categories were used in which a case can have 1) all contacts within the

Table 4.4: Partner selection of chlamydia infected cases with contacts either within the same community, outside the same community, or contacts both within and outside their community. Case partner location was analyzed for A) age groups, B) gender and c) risk area categories.

Case partner location			P value		
<b>Age</b>					
A)		<b>&lt;20</b>	<b>20-29</b>	<b>&gt;29</b>	0.184
	<b>Within</b>	111(47%)	273(52%)	74(60%)	
	<b>Outside</b>	98(41%)	192(36%)	36(29%)	
	<b>Within and Outside</b>	28(12%)	63(12%)	13(11%)	
		<b>237</b>	<b>528</b>	<b>117</b>	
<b>Gender</b>					
B)		<b>Males</b>	<b>Females</b>		0.124
	<b>Within</b>	101(50%)	358(51%)		
	<b>Outside</b>	67(33%)	261(37%)		
	<b>Within and Outside</b>	34(17%)	81(12%)		
		<b>202</b>	<b>700</b>		
<b>Risk area</b>					
C)		<b>High</b>	<b>Low</b>		0.0016*
	<b>Within</b>	231(57%)	223(46%)		
	<b>Outside</b>	123(30%)	201(42%)		
	<b>Within and Outside</b>	52(13%)	57(12%)		
		<b>406</b>	<b>481</b>		

partner location was also used for subsequent analyses. Data for age and geographic location was available for 882 cases. No significant association was found between age and partner location ( $\text{Chi}^2$ ,  $P = 0.184$ ). The majority of cases within any of the age groups were most likely to have partners that were inside the same community. Infected persons >29 years of age had the highest number of reported contacts within the same community (60%), followed by individuals in the 20-29 age group with 52% of contacts in the same community, and lastly, 47% of cases <20 years of age reported contacts in the same community.

A large group of cases also reported contacts outside of their communities ranging from 29% in the >29 group to 41% in the <20 group. The <20 age group appear most likely to choose partners outside their community (41%). All age groups may be important bridges between communities since they have similar reports of contacts who are either outside their community or both within and outside.

Partner location based on gender was also analyzed as above. Table 4.4B shows no significant difference with respect to males or females having partners in a particular community ( $P = 0.124$ ). There was a trend for females to report partners outside the community (37%), in comparison to males (33%) while the percentage of males choosing

partners both within their own community and in a different community was higher than females (17% vs. 12%).

Table 4.4C demonstrates the partner location of high and low risk area cases. There is a significant difference identified in that cases in high risk areas have the majority of their sex partners within the same community compared to low risk area (57% vs. 46%). Fewer high risk area cases (30%) reported having partners outside their community, in comparison to low risk area cases (42%).

Overall, the only statistically significant difference seen with respect to partner location was seen in high risk area cases, who report a majority of their contacts within the same community. This may be one of the factors responsible for maintaining infections in these high incidence areas, as endemic infections through intra-community mixing may be enhanced.

#### ***4.3.3 Case characteristics based on transmission success***

We wished to determine whether long chains of chlamydia transmission (defined as the number of STD-infected cases directly connected together within a network) showed any association with specific age groups, gender or geographic area. For example, are more males than females typically present in chains of 3 or more people, or are long transmission chains more frequent in high vs. low risk areas? The length of these chains we took as a measure of “transmission success”. This measure is similar to  $R_0$ , in that it helps to define the number of secondarily infected persons present within a network.

Overall, in our dataset, we identified 1,490 instances of one infected person, 127 chains of 2 infected individuals, 24 chains of 3 people, 7 chains of 4 people, 3 chains of 5 people, and 1 chain each consisting of 6,7, and 9 people.

Results for the number of individuals of a certain age within a transmission chain is presented in table 4.5A/B. The majority of cases, regardless of their age, were more likely to be involved in a transmission chain of one. In other words, they were connected to a non-infected person who at the time tested negative for chlamydia, or could not be identified or located. Differences were observed in cases <20 and 20-29 years of age compared to cases >29 years of age, as no case >29 years of age were identified within transmission chains greater than three people. Cases <20 tended to be present in longer transmission chains compared to older cases. Cases >29 years of age did not appear in transmission chains greater than 3 persons long. ( $\text{Chi}^2$ ,  $P = 0.00042$ ).

Next, we measured the transmission success of chlamydia based on gender. Our analysis revealed a very strong correlation between gender and transmission chain length (Table 4.5C/D,  $\text{Chi}^2$ ,  $P = <0.000001$ ). The majority of both males (66%) and females (85%) are directly connected to a non-infected or non-located/tested person. However, the marked difference between the two is seen within longer transmission chains where 34% of males, compared to 15% of females represent transmission chains of 2 or more connected cases.



Table 4.5: Transmission success measured by the length of chains consisting of directly connected cases characterized by A/B) age groups, C/D) gender and E/F) risk area categories. Tables B,D and F are grouped data from tables A,C and E, respectively, to demonstrate differences within transmission chains greater than two people.

**Transmission Success**

<b>A)</b>		<b>Case age</b>			<b>B)</b>		<b>Case age</b>		
<b>Chain length</b>	<b>&lt;20</b>	<b>20-29</b>	<b>&gt;29</b>	<b>&lt;20</b>	<b>20-29</b>	<b>&gt;29</b>	<b>&lt;20</b>	<b>20-29</b>	<b>&gt;29</b>
<b>1</b>	177(68%)	363(66%)	110(79%)	<b>1</b>	177(68%)	363(66%)	110(79%)		
<b>2</b>	46(18%)	112(21%)	28(20%)	<b>2</b>	46(18%)	112(21%)	28(20%)		
<b>3</b>	17(6.6%)	43(8%)	1(0.7%)	<b>&gt;2</b>	34(13 %)	81(13%)	1(1%)		
<b>4</b>	4(1.5%)	18(3.3%)	0						
<b>5</b>	3(1.2%)	6(1.0%)	0						
<b>6</b>	4(1.5%)	1(0.2%)	0						
<b>7</b>	4(1.5%)	2(0.4%)	0						
<b>9</b>	4(1.5%)	1(0.2%)	0						

<b>C)</b>		<b>Males</b>	<b>Females</b>	<b>D)</b>		<b>Males</b>	<b>Females</b>
<b>1</b>	358(66%)	1132(85%)		<b>1</b>	358(66%)	1132(85%)	
<b>2</b>	127(23%)	126(9%)		<b>2</b>	127(23%)	126(9%)	
<b>3</b>	29(5.3%)	43(3.2%)		<b>&gt;2</b>	61(11%)	77(6%)	
<b>4</b>	13(2.4%)	15(1.1%)					
<b>5</b>	6(1.1%)	9(0.7%)					
<b>6</b>	3(0.6%)	3(0.2%)					
<b>7</b>	4(0.7%)	3(0.2%)					
<b>9</b>	6(0.6%)	3(0.2%)					

<b>E)</b>		<b>High risk</b>	<b>Low risk</b>	<b>F)</b>		<b>High risk</b>	<b>Low risk</b>
<b>1</b>	258(65%)	398(71%)		<b>1</b>	258(65%)	398(71%)	
<b>2</b>	77(19%)	109(19%)		<b>2</b>	77(19%)	109(19%)	
<b>3</b>	32(8%)	31(5.6%)		<b>&gt;2</b>	60(16%)	52(10%)	
<b>4</b>	8(2.0%)	14(2.5%)					
<b>5</b>	5(1.3%)	4(0.7%)					
<b>6</b>	4(1.0%)	2(0.4%)					
<b>7</b>	7(1.8%)	0					
<b>9</b>	4(1.0%)	1(0.2%)					

There was a total of 621 transmission chains identified with geographic locations known for each individual in a specific transmission chain. A significant association was identified between risk area and chain length (Table 4.5E/D,  $\text{Chi}^2$ ,  $P = 0.008$ ). The highest number of cases in both risk areas are in transmission chains of one or two persons. However, a higher proportion of individuals in high risk areas (16%) were involved in longer transmission chains ( $>3$ ) compared to low risk area cases (10%).

#### ***4.3.4 Characteristics of cases present in certain network sizes***

Large networks contain more potential transmission routes and may be critical in maintaining infections. Network size was measured for each of the cases in the data set for the three variables of interest (Table 4.6). The first variable consisted of age. There was a significant difference in network size for different age classes ( $\text{Chi}^2$ ,  $P = 0.000002$ ). Younger cases were more likely to be part of a larger network where the average network size was 7.3. Cases  $>29$  years of age were found at the other end of the spectrum where the average network size for this group was only 2.8. The remainder of cases 20-29 years of age, were found in the middle of these two extremes with an average network size of 5.5.

Statistical analysis identified a significant difference between males and females with respect to network size (Table 4.6B, KW,  $P=0.00028$ ). Females were more likely to be in smaller networks (mean network size of 5.1) compared to males (mean size of 6.0). These results coincide with earlier findings that males on average, tend to have more partners than female cases.

Statistical analysis revealed a significant difference between high and low risk areas and network size (Table 4.6C, KW, P = 0.010). High risk area cases are found in somewhat larger networks than cases in low risk areas (network size of 7.0 vs 4.6, respectively).

	Case average network size			P value
a) Age	<b>&lt;20</b>	<b>20-29</b>	<b>&gt;29</b>	
	7.3	5.5	2.8	0.000002*
b) Gender	<b>Males</b>	<b>Females</b>		
	6.0	5.1		0.00028*
c) Risk area	<b>High</b>	<b>Low</b>		
	7.0	4.6		0.010*

Table 4.6: The average network size for cases within our dataset based on A) age groups, B) gender and C) risk area categories.

#### ***4.3.5 Characteristics of cases based on network measures of closeness and betweenness***

The next two measures for our population, were based on network measures of closeness and betweenness (refer to section 1.4.1 for an overview). “Closeness” measures the centrality of persons within a network by measuring how close each person is to every

other person in that same network. "Betweenness" measures how many times a person lies in a path between two other people. These measures are important in that they provide some information on network structure. Using UCINET (Freeman *et al.*, 1999 © Analytic technologies inc.) closeness measures for each person in a network were calculated and stratified by age, gender and geographic location.

Analysis based on age indicated that older age groups were more likely to be more central as they had the highest average closeness value of 81.2 (Table 4.7A, KW,  $P=0.000123$ ). No significant difference for betweenness and age was identified (Table 4.8A, KW,  $P = 0.059$ ). However, as for closeness, the trend was for older age groups to have higher betweenness values.

Average closeness values were calculated based on gender. There were no significant differences between males and females (Table 4.7B, KW,  $P = 0.687$ ). Betweenness values for males and females also did not differ (Table 4.8B,  $P = 0.737$ ).

We calculated the average levels of betweenness and closeness for cases within high and low risk areas (Table 4.7C and 4.8C). Significant differences were observed for closeness ( $KW = 0.000045$ ). For cases in these respective areas, the average closeness was 66.4 and 75.8. Cases in low risk areas are more closely connected to everyone else in a network while the reverse is true for high risk areas. Betweenness was not significantly different between the two risk areas (KW,  $P=0.053$ ).

Table 4.7 : Average closeness for cases present in our dataset measured by A) age groups, B) gender and C) risk area categories.

	Case average closeness			P value
a) Age	<b>&lt;20</b>	<b>20-29</b>	<b>&gt;29</b>	
	67.8	71.3	81.2	0.000123*
b) Gender	<b>Males</b>	<b>Females</b>		
	73.7	73.1		0.687
c) Risk area	<b>High</b>	<b>Low</b>		
	66.4	75.8		0.000045*

Table 4.8: Average values of “betweenness” for cases analyzed by A) age groups, B) gender and C) risk area categories.

	Case average betweenness			P value
a) Age	<b>&lt;20</b>	<b>20-29</b>	<b>&gt;29</b>	
	52.5	53.0	65.8	0.059
b) Gender	<b>Males</b>	<b>Females</b>		
	57.6	58.3		0.737
c) Risk area	<b>High</b>	<b>Low</b>		
	51.1	57.4		0.053

#### 4.4 Discussion

##### *Mixing patterns between cases and their contacts within the Manitoba sexual network of chlamydia infected cases*

The degree of mixing based on several different variables has been essential in explaining STD maintenance and spread (Ellen *et al.*, 1997; Rothenberg, 1983; Garnett *et al.*, 1996). Several studies have revealed that the spread and maintenance of infections is largely influenced by the mixing patterns between cases and their contacts. This concept emerged as a result of a changing paradigm in STD research that integrates information on both cases and their contacts. Several authors have drawn attention to the importance of mixing between core and non-core groups (Zenilman *et al.*, 1999; Lauman and Youm, 1999; Rosenberg *et al.*, 1999). In principle, selection of sexual partners can vary between two extremes: assortative, or like with like, and dissortative. Each pattern has an effect on the course of an epidemic with assortative mixing resulting in rapid early spread of disease with rapid decline, and dissortative mixing resulting in slower onset of disease with a longer lasting epidemic (Anderson, 1992). Sociologists report that in general, human sexual behavior is assortatively selective. Garnett and Anderson (1993) reviewed epidemiological studies reporting STD transmission among STD cases and their contacts. These authors observed that in a heterosexual population, sexual networks transmitting gonorrhea and chlamydia infections exhibited assortative selectivity in partner choice. However, a homosexual network in Iceland in which HIV transmission occurred was dissortative (Haraldsdottir, 1992). Thus, data suggests that this characteristic varies among different sexual networks, and, in turn, affects the course of a given epidemic.

In this chapter, we evaluated the degree of mixing of STD cases and contacts in Manitoba. The data available to us to assess mixing was age, partner number and geographic location for cases and contacts. Mixing was based on the Q statistic, which measures the degree of assortative and disassortative mixing.

Anderson (1992) provides a model that highlights the need for quantitative data on age-dependency in rates of sexual partner change. For example, in their model, a scenario with age mixing between older males with younger females resulted in the highest impact on the spread of HIV. In a population with men having sex with men, Service and Blower attributed the rapid spread of HIV to the formation of partnerships between older males with younger males (Service and Blower, 1996). As a result of the increasing evidence that age mixing patterns provide important information on STD transmission dynamics, it has been noted that age mixing can now be added to other correlates of disease (Catania *et al.*, 1996).

Results show that overall, our study population tended to choose partners assortatively based on age, such that STD cases are mainly within the same age category as their partners (Table 4.1). Stratification by gender showed that both male and female cases display assortative mixing patterns; however, several gender-based differences were noted. Almost all of the male cases <20 years of age reported having sexual contacts who are within the same age category while older males tend to choose younger female partners. Female cases tend to choose partners who are in the same age group. Extensive contact between cases and contacts within the same age category serves to keep

infections within that group. However, there were partnerships between individuals within and between age groups which can account for movement of infections from one age group to another.

Analysis based on geographic segmentation of populations infected with STDs have proved beneficial for targeting portions of the population responsible for disease persistence and spread. In Manitoba, high risk geographic locations have been identified primarily in Northern regions of the province and central areas in Winnipeg (Blanchard *et al.*, 1998). Blanchard *et al.* (1998) reported that these regions have the highest chlamydia incidence rates and exhibit high risk behaviors and risk markers for disease status. As such, high risk areas are associated with core group membership, high numbers of sexual partners, low socioeconomic status, and have poor access to the health care system (Shahmanesh *et al.*, 2000; Potterat *et al.*, 1985). Therefore, the degree of mixing between our high risk and low risk areas is important in the magnitude of disease spread throughout the province. For our geographic analysis, we incorporated the Winnipeg geographic segmentation of high and low risk areas based on chlamydia incidence levels developed by Blanchard *et al.* (1998). For rural Manitoba, high and low risk areas were based on chlamydia incidence levels reported by the CDC (Manitoba Health, CDC Unit; Table 4.1). Three remote northern regions, or RHA's, as well as four areas within Winnipeg served as our high risk area population.

There is a slight difference for females in high and low risk areas. Females in high risk areas tend to choose older partners if they are <20 years of age. Females 20-29 years of



age in high risk areas have a higher percentage of <20 year old partners in comparison to low risk females. Males from high risk areas are more assortative than males from low risk areas. Females in high risk areas, have less assortative mixing than females in low risk areas. How this pattern balances out in terms of transmission is not clear. Although both males and females in low risk areas show similar Q values, males and females in high risk areas are disparate (Q of 0.57 for males and 0.30 for females). Although the male pattern would help to keep infections isolated in specific age groups, the female pattern would tend to disperse it throughout age groups and may contribute to a higher overall incidence in the population. This pattern may be significant since it may help STDs to move into a wider range of age groups and increase STD rates overall in high risk areas. The extent to which the discrepancy between male and female Q values is related to less than full disclosure of partner information by either one or both sexes is unknown at this time.

Passage of an STD through any population is largely based on the number of sexual partners contained in that population. Sexual behaviors between partnerships vary within and between populations and the measurement for the degree to which this occurs is captured in the equations for the basic reproductive number  $R_0$  and sex partner change rates proposed by Brunham (1997). Of equal importance are the contacts infected cases choose as their sexual partners. If an infection enters a population with the majority of individuals having no more than 1 sex partner, that population is at limited risk of the infection spreading, due to limited sexual transmission opportunities. Similarly, a population with the majority of individuals having high numbers of partners would result

in the infection spreading rapidly. Mathematical and empirical models have demonstrated that a population suffering from an epidemic has core groups, or segments of the population with high numbers of sexual partners, that serve to keep infections within that population (Gupta *et al.*, 1989; Anderson *et al.*, 1990; Aral, 1999). They have also demonstrated how mixing of these groups with the rest of the population would result in spread from the core to the non-core segment. Enough partnerships outside the core group would result in further maintenance and spread within the entire population. We attempted to capture these features in our Manitoba infected population by assessing mixing based on case and contact partner number.

Our results show that characteristics of partnerships are relatively proportionate with a  $Q$  value close to zero (Table 4.2). Cases, regardless of the number of partners they have, are most likely to be connected to individuals with two partners. In our dataset, this result largely reflects the removal of most single connections, due to a lack of information on those partners. Interestingly, on average about one quarter of the respondents (24%, 22% and 35% for cases with 1, 2 and 3 contacts, respectively) had sexual relations with individuals who in turn had at least three contacts. This may potentiate maintenance of disease in the population since in many cases promiscuous individuals are connected to equally promiscuous sex partners. Frequency of sexual contact and information on sexual behaviors would help to further place this data within a social network context.

Differences in gender for partner number revealed proportionate mixing for both males and females, with males tending towards a stronger assortative mixing pattern (Table

4.2). Female cases, regardless of partner number, were more likely than males to choose partners who had two sexual contacts. In comparison to females, male cases with one and two partners had an increased tendency to have sexual contact with individuals who have only one sex partner. However, compared with females, males with greater than or equal to three sex partner tended to choose partners who also had greater than or equal to three partners. Our data indicates that females, generally, choose individuals with two sex partners. This characteristic enables infections to be maintained in a population, since sexual transmission opportunities exist. Similarly, males with greater than or equal to three sex partners are of concern as our data demonstrates they are more likely to choose partners who also have many sex partners. This allows for rapid movement and spread of infections. Because substantial sexual mixing occurs between individuals with high numbers of partners, transmission of chlamydia within this population can occur.

Further stratification by risk area revealed that both males and females in low risk areas exhibited proportionate mixing (low risk males and females Q values, 0.02 and 0.05 respectively vs. high risk males and females Q values, 0.18 and -0.02, respectively). This characteristic may also be related to higher disease incidence in these areas as it indicates that promiscuous individuals in high risk areas tend to partner with other promiscuous individuals. This characteristic would ensure numerous transmission opportunities for STDs to maintain a high  $R_0$  value within this population.

Based on our results on mixing patterns of cases within the Manitoba sexual network, male and female cases present in high risk areas exhibit the most variation in partner

selection. Particularly, females in high risk areas have a tendency to choose partners who are within different age groups compared to low risk area females. Similarly, males in high risk areas also show this same characteristic, however, they also choose partners who are younger in age and partners with higher numbers of sexual contacts, in comparison to low risk area males. This highlights the extent to which control programs need to focus efforts specifically towards males, and cases within high risk areas.

*Demographic and network characterization of chlamydia infected cases in Manitoba*

In this section, we attempted to characterize chlamydia infected cases present in Manitoba. Epidemiological research relies on identification of certain segments of the population most burdened with infections in order for the development of accurate and effective STD prevention programs. Numerous studies have focussed on groups of individuals with high numbers of partners as an important determinant of STD risk. For example, CSW's, FSW's and persons within core groups (Brunham, 1997; Brunham and Plummer, 1990; Baseman *et al.*, 1999; Rothenberg *et al.*, 1998). Studies have also focussed on groups of infected persons based on age and gender where males and individuals young in age are more likely to be infected (Herold and Mewhinney, 1993; Orr *et al.*, 1994). We were able to characterize various network characteristics for the infected population in Manitoba stratified by age, gender and geographic location. Univariate analyses were used to determine which network characteristics showed significant differences between the various strata.

Wasserheit and Aral (1996) suggested that the epidemiology of STDs would change in a predictable pattern in response to disease control activities. Specifically, Brunham and Plummer (1990) postulated that control efforts serve to concentrate STDs in more sexually active populations that are less accessible to public health control programs. Blanchard *et al.* (1998) demonstrated that, as a result of STD control strategies in Manitoba, chlamydia infections had become more dependent on STD core groups over time. We characterized our population based on the number of sex partners reported by cases. Of interest are individuals with high numbers of partners who contribute disproportionately to chlamydia transmission. For the number of partners reported by cases, significant associations with age and gender were identified. For age, younger individuals reported more sex partners than other individuals. For gender, male cases reported more partners than female cases. Age characteristics are in keeping with the observation that STDs are primarily a disease of youth, however, the gender differences highlight the need to develop more male focussed STD control strategies.

An interesting phenomenon with respect to STD epidemiology is the pattern of recruitment of sex partners at the community level. Calzavara *et al.* (1999) provided results from an aboriginal community in Ontario. They determined whether cases chose partners from within their own community, outside their community or from both within and outside their community. In their study, data on sexual partnering indicate that many individuals chose partners from both within and outside their community. These individuals serve as sexual bridges between communities. As incidence of infection rises in one community, disease would spread rapidly to other communities. Based on the

findings of Calzavara *et al.* (1999), we wanted to assess sexual partnering of our population at the community level based on age, gender and risk area. There was no association between partner location and either age or gender (Table 4.4). A significant association was seen, however, between partner location and risk area.

Analysis of cases based on risk area revealed that 57% of high risk area cases chose partners within their own community. This scenario can result in intra-community maintenance of infections contributing endemic persistence of STDs. In contrast, cases in low risk areas tend to choose partners outside their community. The consequent lack of internal networks within communities in low risk areas may account for lower incidence levels in these areas. Therefore, although the potential for bridging exists in low risk areas in Manitoba, there may be fewer opportunities for maintenance of infection. Further classification of these bridge individuals may serve to identify subgroups of individuals for targeting in prevention programs.

Successful transmission of STD pathogens is affected by many factors such as condom use, frequency of intercourse and type of intercourse. We had no data available on sexual behaviors within a partnership, however, we were able to assess transmission success by identifying chains of connected cases. In our data set, we identified 1654 transmission chains ranging in size from one to nine and analyzed these chains with respect to age, partner number and risk area (Table 4.5). From the three variables analyzed, significant associations were seen with respect to length of transmission chain. Larger chains were

more likely to contain males, individuals <20 years of age, and to occur in high risk areas.

For gender, only 66% of males compared to 85% of females were in transmission chains of one. This observation is related to the fact that in a female-based screening program, male cases would typically only be entered in the provincial database after a female case has been diagnosed (hence by default, males should be skewed towards transmission chains of two). Whether transmission is more frequent for male to female or female to male is unknown for our population. Notably, however, males are more common in transmission chains of three or more, than females. Again, in a female based screening program, for a transmission chain of three, an initial female case should lead to a male, and then a subsequent female case. Ratios of females to males in transmission chains of three or more might be expected to be close to 2:1. In our dataset, an approximately equal number of males and females are seen in chains of three or more. This observation suggests that an infected male may be more likely to spread infections than an infected female, again highlighting the need for male-based control strategies.

Our results also indicate that cases young in age and present within high risk areas are responsible for efficient transmission to sexual partners as measured by the length of transmission chain. Successful transmission based on age, again, highlights the higher incidence within these age groups. STD control strategies may need to be developed which further targets high risk youth. The higher transmission success of chlamydia in high risk areas, also, reflects the higher incidence in these areas, and the larger number of

The remaining three variables analyzed by age, gender and risk area were network measures such as network size, closeness and betweenness. Network size and structure have important roles in disease dynamics. Larger network size reflects increased opportunities for disease transmission and spread. Evaluating network structure, such as closeness and betweenness, can help to relate incidence in a specific group to network structure. Network structure can play a role in the spread of infections within networks (Rothenberg *et al.*, 1998; Klodahl, 1994).

In the Manitoba network, larger networks were more likely to be associated with individuals young in age (<20 years of age), male (average network size of six compared with five for females) and to occur in high risk areas (average network size of 7 compared to 4 in low risk areas). High values of closeness were associated with increasing case age and were higher in low risk areas. Overall, closeness measures are higher for a radial vs. a linear network structure. Wylie and Jolly (2001) previously demonstrated that, at least for large networks, radial network structure is associated with a lower positivity rate. The above results also suggest that the differences in closeness values may reflect a greater tendency for linear networks to form in young age groups and in high risk areas. In turn, this network structure may contribute to higher STD incidence in younger age groups and in high risk areas.

In summary, several significant associations were found which contribute to high STD incidence rates. Cases young in age, male, and located in high risk areas, are on average, present within large sexual networks. In addition, high risk area cases have more partners



located within the same community, maintaining infections in the population. STD control efforts should focus on these groups, which in turn would lead to a decrease in the number of partners a given individual may have, interruption of chlamydia transmission and a decrease in network sizes present in this population. Our results also highlight the need for further studies on the role of mixing patterns and network size and structure on STD incidence patterns.

## CHAPTER 5

# DEMOGRAPHIC ANALYSIS OF MOLECULAR GENOTYPES AND GEOGRAPHIC DISTRIBUTION OF SEXUAL NETWORK CONNECTIONS IN MANITOBA

### 5.1 Introduction

Many studies have highlighted the correlation between geography and STD rates (Zimmerman *et al.*, 1990; Rothenberg, 1983; Potterat, 1992; Hamers *et al.*, 1995). Our social network analysis data provides the opportunity to represent the geographic connections that result from sex partner formation. In this chapter, we provide a general overview of all reported connections between communities to monitor where and how partnerships occur within Manitoba. Of interest are the patterns between different parts of the province. This analysis will help to identify which sections of the province are connected by sexual bridging, and in which parts of the province coordinated approaches to STD control may be advantageous. We also conducted a geographic analysis of the distribution of chlamydia genotypes. Identifying genotype distribution patterns within certain areas may allow identification of larger networks that would not have been identified through network analysis alone.

As part of our study, we identified several chlamydia genotype variants. These were graphed separately from prototypic genotypes. Past evidence suggests that chlamydia variants arise predominantly among core group individuals due to immune pressure on the infecting strain (Brunham and Plummer, 1990). Currently, it is believed that the

ecological success of these organisms in repeatedly exposed core group individuals, depends on sequence variation. For our geographic analysis of variant genotypes, patterns within high and low risk areas were of interest to determine whether chlamydia variants are more common in high or low risk areas. It was also of interest to determine if the presence of certain chlamydia genotypes correlated with specific demographic or network characteristics.

The final section of this chapter focuses on modeling the relationship between positivity rates and network structure. Network structure plays a role in the rate and spread of infections. For example, Rothenberg *et al.* (1998) demonstrated the effect of network microstructure on syphilis transmission. We assessed the relationship between chlamydia positivity, network size and the number of 3-plexes in a network.

## **RESULTS**

### **5.2 Overall geographic distribution of partnerships**

The connections that make up the various network components in this study were mapped to generate figure 5.1. This map shows all of the provincial connections between communities that were identified within our data set. For this figure, for visual clarity, connections were graphed only if at least two sexual partnerships occurred between individuals in different communities.

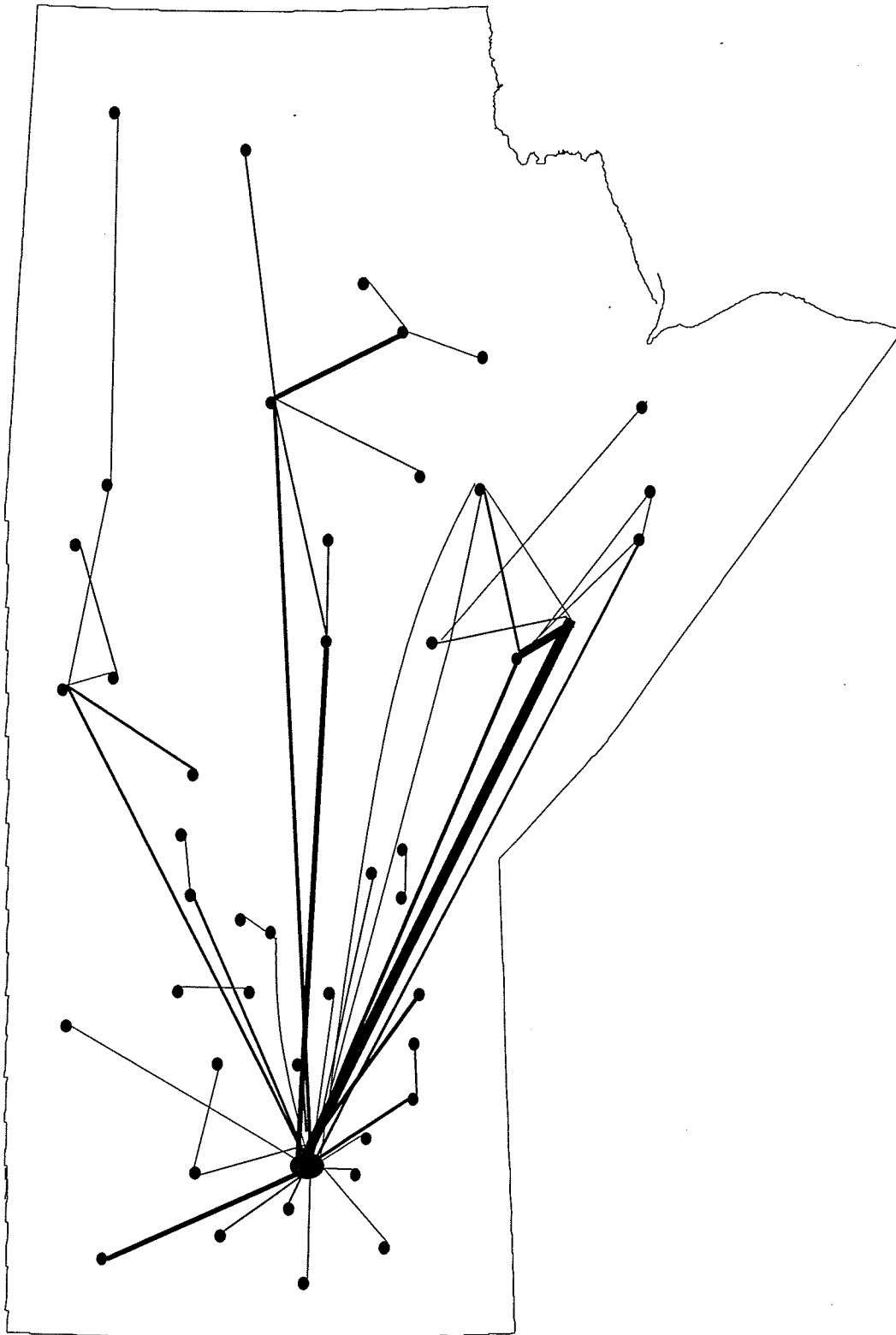


Figure 5.1: Geographic connections between communities connected by at least two partnerships. Connections represent all links identified from networks of size 2 to 82. The large black circle represents Winnipeg. Small circles represent communities within the province. Lines connecting communities represent sexual partnerships. Thin lines are few connections and thick lines are many connections. E.g.— represents 2 links, ——— represents 3 links, ————— represents 4 links, and ————— represents 12 links. This applies to figures 5.3 to 5.11.

As expected, based on its size and the fact that it is the largest urban center in the province, Winnipeg acts as a focal point for chlamydia transmission throughout the province. Communities present in low risk areas appear to be connected directly to Winnipeg with few connections to rural high risk areas or to other communities in low risk areas. Winnipeg is also directly connected to communities within high risk areas; however, in comparison to low risk areas, individuals in high risk areas appear to have sexual partnerships within and between the other communities in rural high risk areas.

Of interest are the connections that occur between Winnipeg and high risk areas such as Burntwood, Norman and North Eastman. Burntwood exhibits highest levels of chlamydia incidence in the province. Bridging within this area, as well as to other high risk areas, suggests a high probability of transmission between the connected communities. Overall, the majority of all connections identified occurred between Winnipeg and Burntwood. This suggests that infections have a high probability of moving between the two areas, which can lead to transmission across the province, as Winnipeg has extensive bridging to most other parts of Manitoba.

The above analysis focussed on geographic patterns between communities in Manitoba. However, it did not incorporate data on the extent to which sexual contact occurred between individuals within certain areas of Winnipeg, which was the focus of our next analysis. Additionally, connections to rural areas of the province from particular locations in Winnipeg are also incorporated to more fully represent the data. The geographic data was stratified based on high and low risk areas for both rural areas and within Winnipeg.

Core areas within Winnipeg were classified as high risk while adjacent and peripheral areas were classified as low risk areas as defined by Blanchard *et al.* (1998). Overall, the majority of individuals present in our dataset were located within high risk rural areas (35%) and low risk urban areas (40%) (Figure 5.2). In addition, the majority of intra-area sexual contact occurred within these two areas, such that 30% of all sexual contact occurred between individuals only within high risk rural areas and 31% between only low risk urban area individuals. Twenty four percent of all connections occurred between the four areas, with 11% of these between high risk areas in Winnipeg and low risk areas in Winnipeg. Therefore, core individuals of Winnipeg are responsible for the majority of connections to non-core areas of Winnipeg. This high degree of inter-area mixing between all four risk areas highlights the potential for geographic discontinuity.

### **5.3 *Geographic connections within co-infected networks***

Contained within our database was information on cases who were co-infected with both chlamydia and gonorrhea. Present evidence suggests that persons who are co-infected are more likely to be part of a core group (Jolly, 1998) and are therefore of particular interest. We mapped all network connections for networks that contained at least one co-infected individual. An overall distribution is shown in figure 5.3. Most of the connections in the low risk areas are lost since co-infected individuals are found within networks which are mainly located in high incidence areas, particularly the eastern part of the province. By targeting high risk areas and co-infected persons, core groups may be minimized.

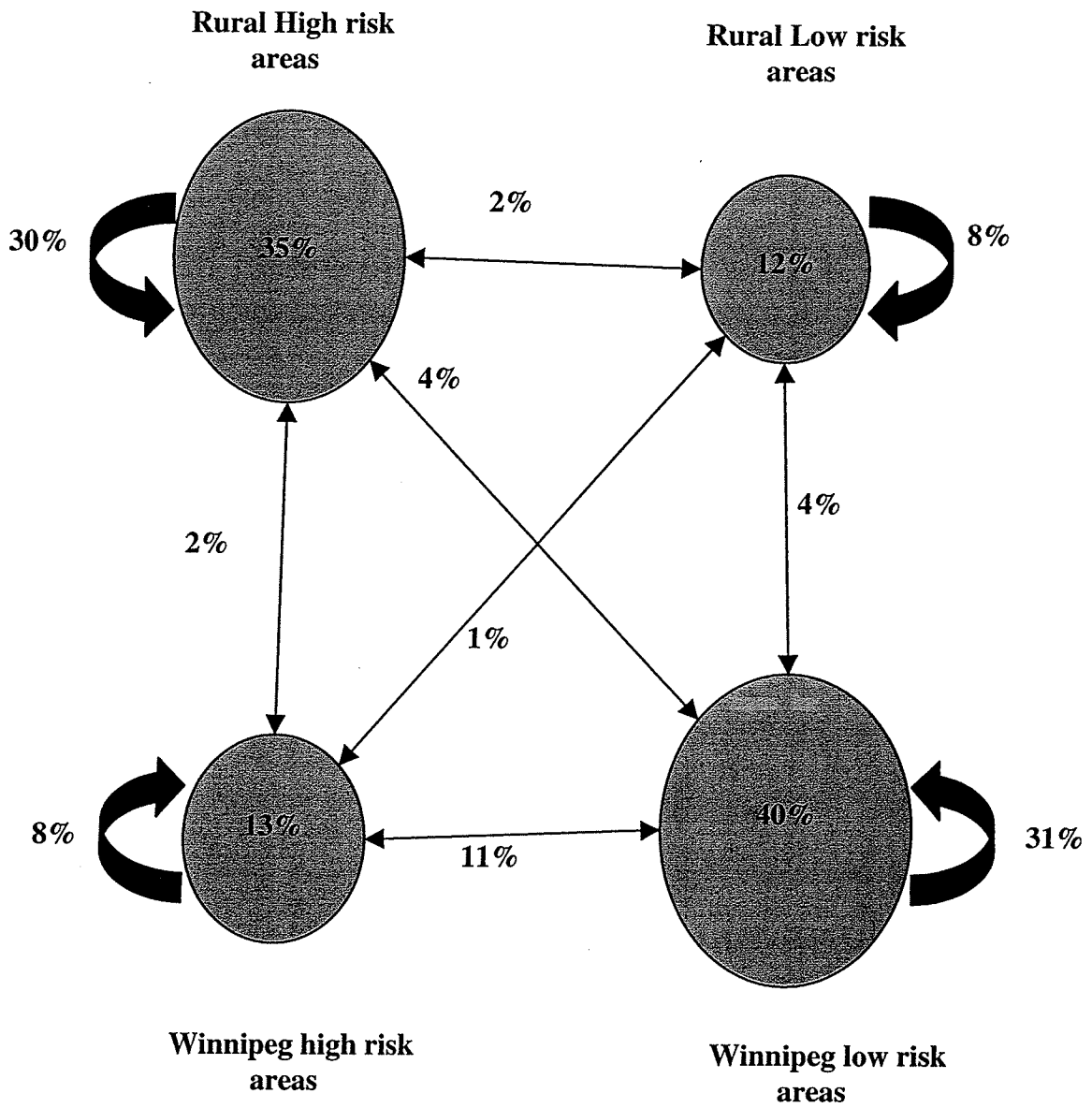


Figure 5.2: Sexual partner mixing between individuals located in high risk areas within and between rural Manitoba and Winnipeg. Numbers within circles represent the percent of individuals present within those areas. Numbers by thick arrows represent the percent of intra-community partnerships. Numbers by thin arrows represent the percent of partnerships occurring between those risk areas.

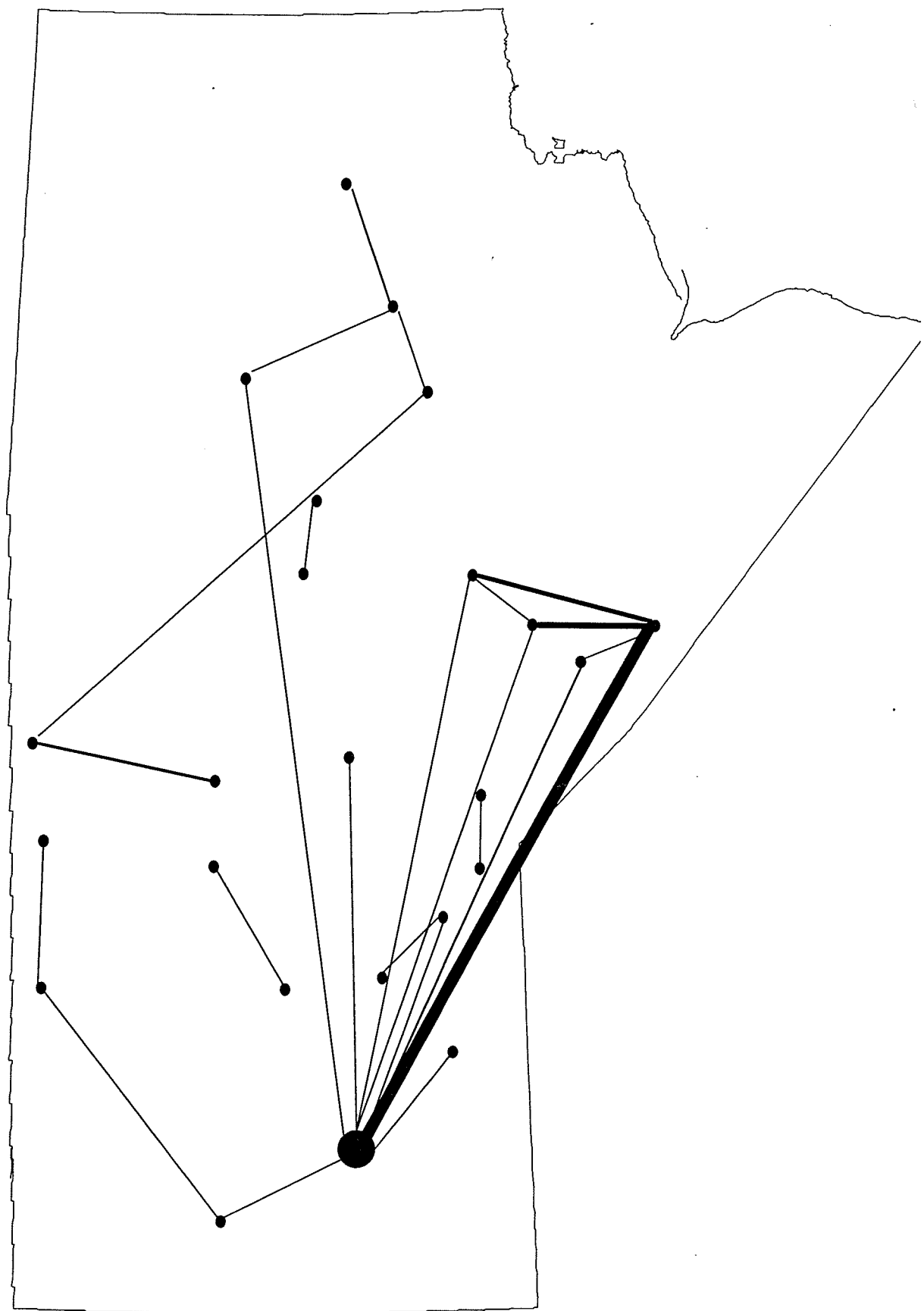


Figure 5.3: Geographic connections of networks containing at least one co-infected individual. See Figure 5.1 for symbol description.



#### **5.4 Geographic distribution of partnerships within different network sizes**

Breaking down connections based on network size provided a more detailed analysis of provincial connections. The geographic distributions can be seen from figures 5.4 to 5.11. Every possible network of a specific size was mapped ranging from two to 82. Small networks had many different community connections and were therefore graphed on separate maps. Larger networks were grouped together as they had fewer connections between communities. A comparison of the data suggests that as network size increases, connections between communities in the west part of the province decrease. That is, most of the connections within the low risk rural areas are lost. As shown in chapter 4, this is consistent with the larger average size of networks in high risk areas.

#### **5.5 Analysis of molecular genotypes identified in the province**

Molecular epidemiology studies of chlamydia in different parts of the world have revealed that types genotypes D, E and F, are more frequently isolated than other types of chlamydia (Ramirez *et al.*, 2000; Lan *et al.*, 1995). This consistent pattern suggests that there may be biological differences between chlamydia types. Although specific differences have not been identified, they could be related to factors such as differences in the quantity of bacterial cells shed, or differences in the infectious dose. In turn, these differences could manifest themselves as differences in transmission potential and hence, abundance in a population.

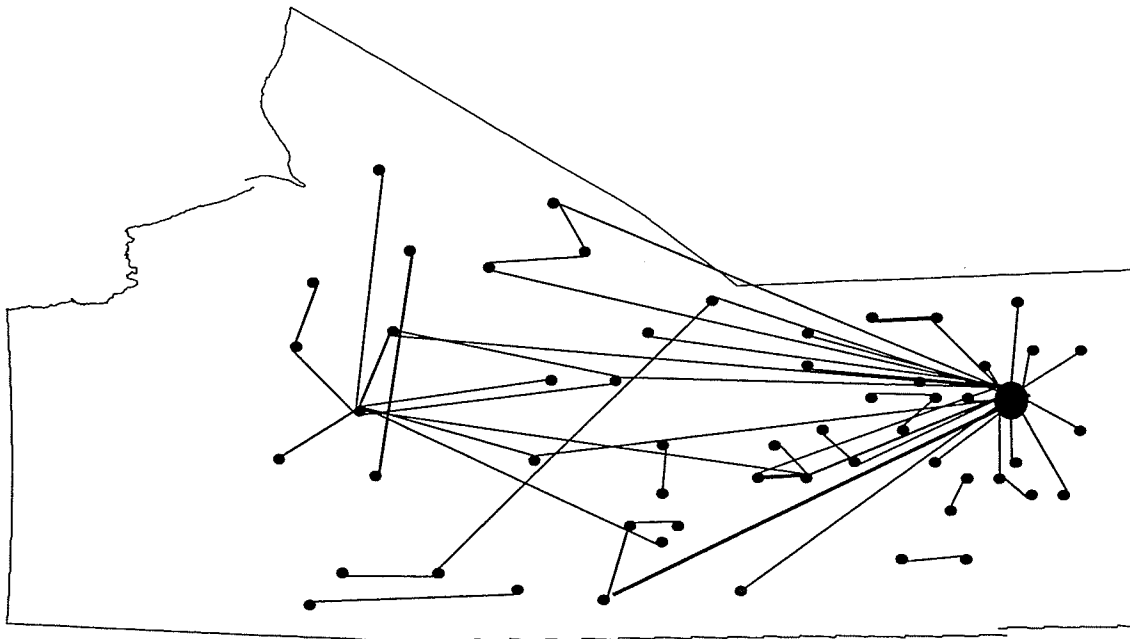


Figure 5.4: Geographic connections for networks of size 2.

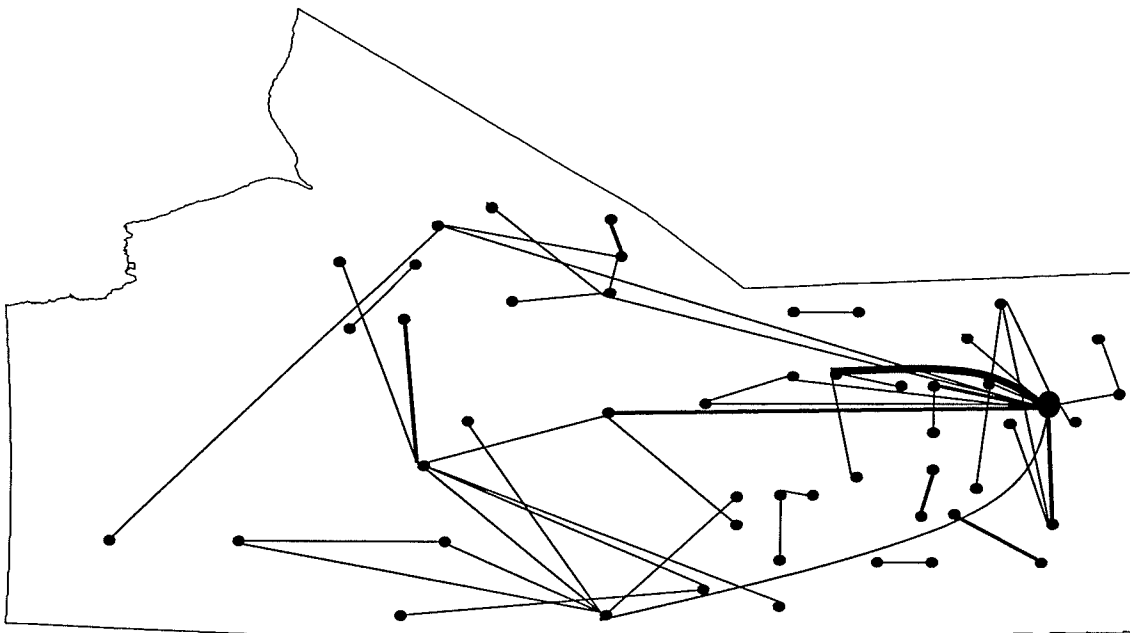


Figure 5.5: Geographic connections for networks of size 3.

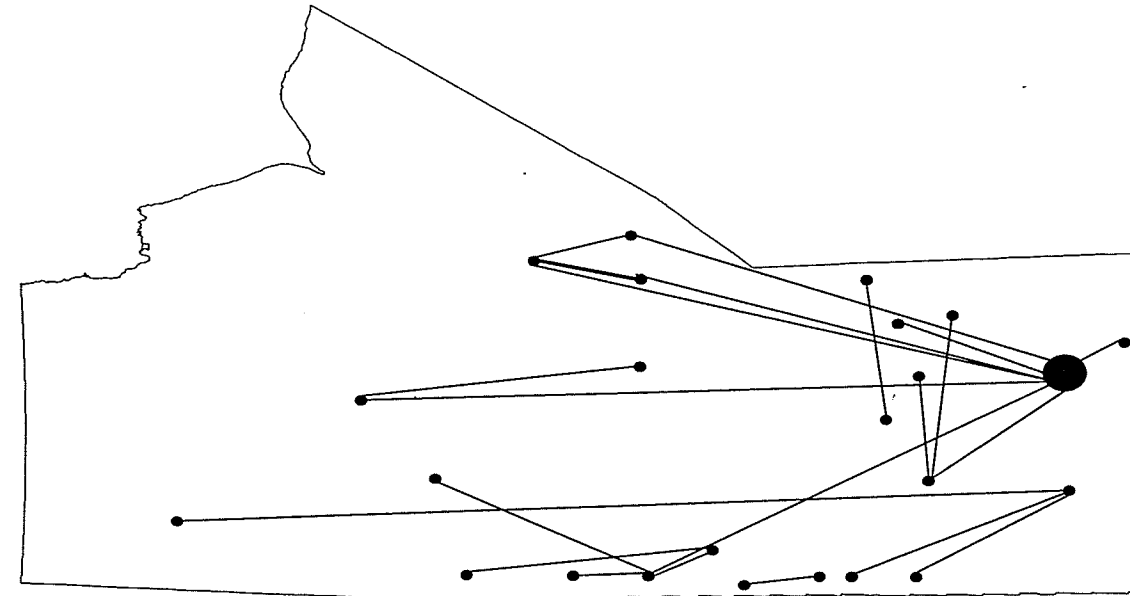


Figure 5.6: Geographic connections for networks of size 4.

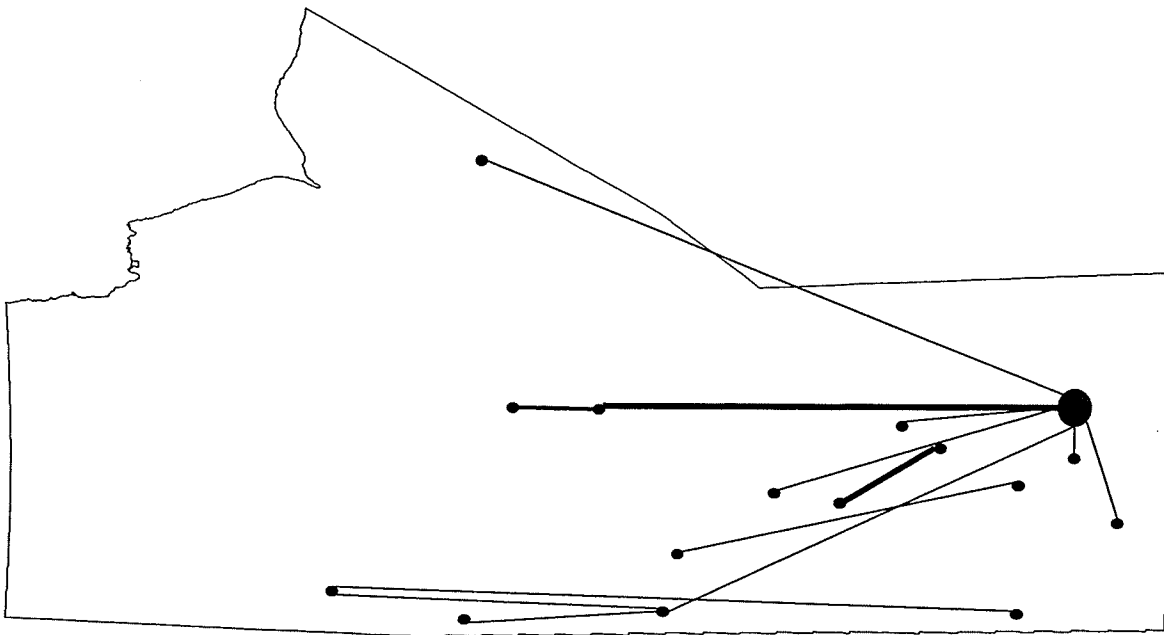


Figure 5.7: Geographic connections for networks of size 5.

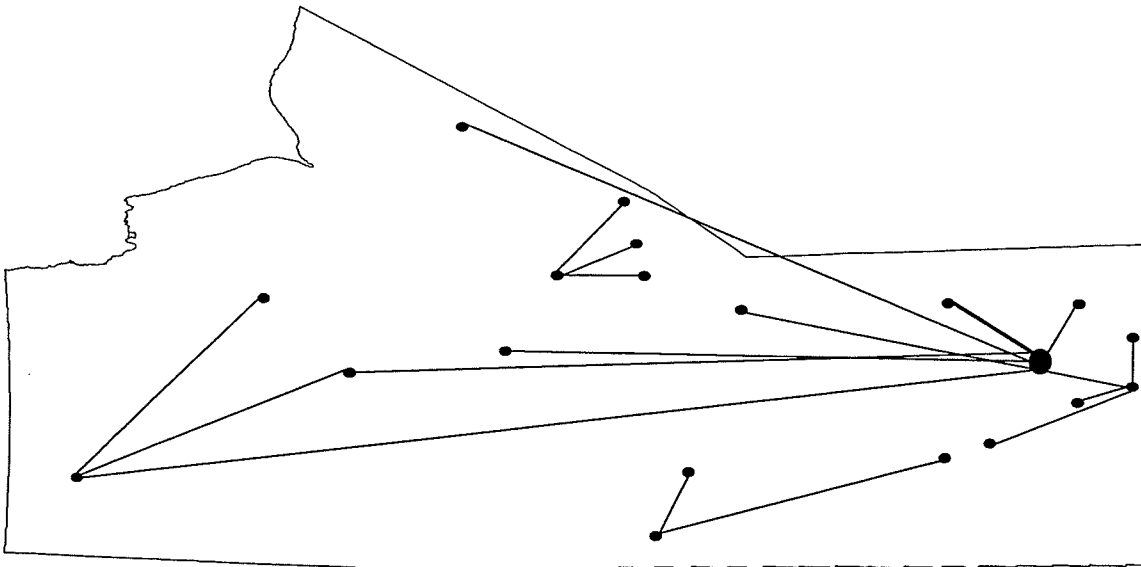


Figure 5.8: Geographic connections for networks of size 6.

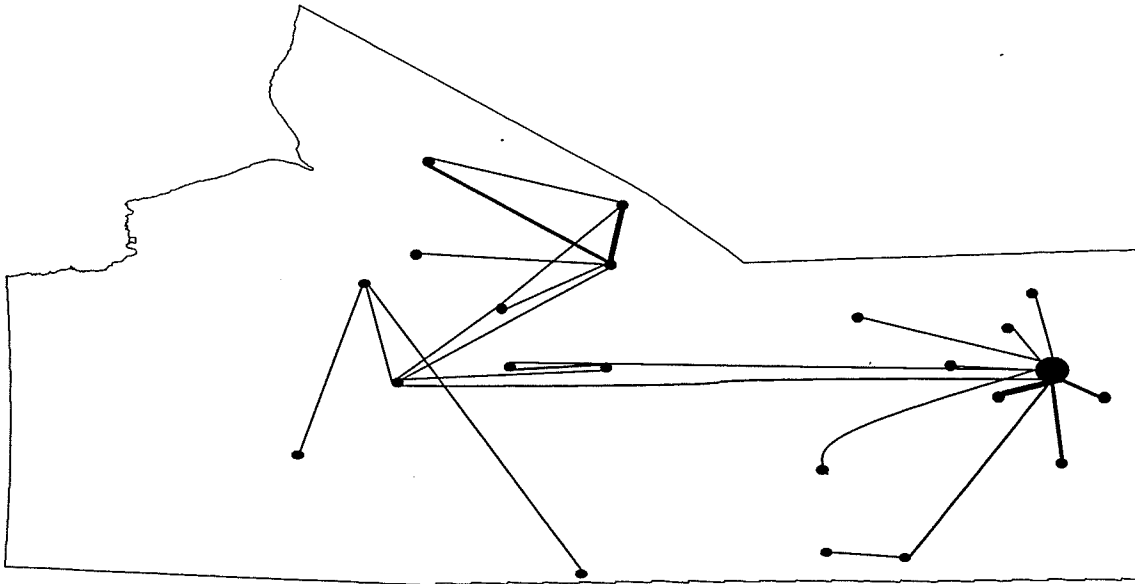


Figure 5.9: Geographic connections for networks of size 7, 8, 9 and 10.

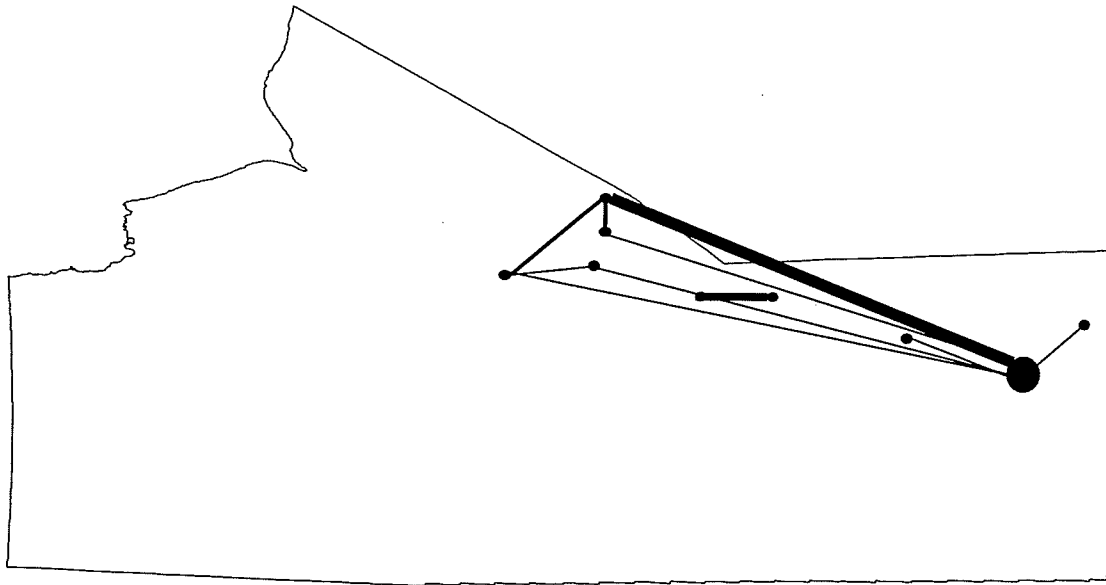


Figure 5.11: Geographic connections for networks of size 39, 41 and 82.

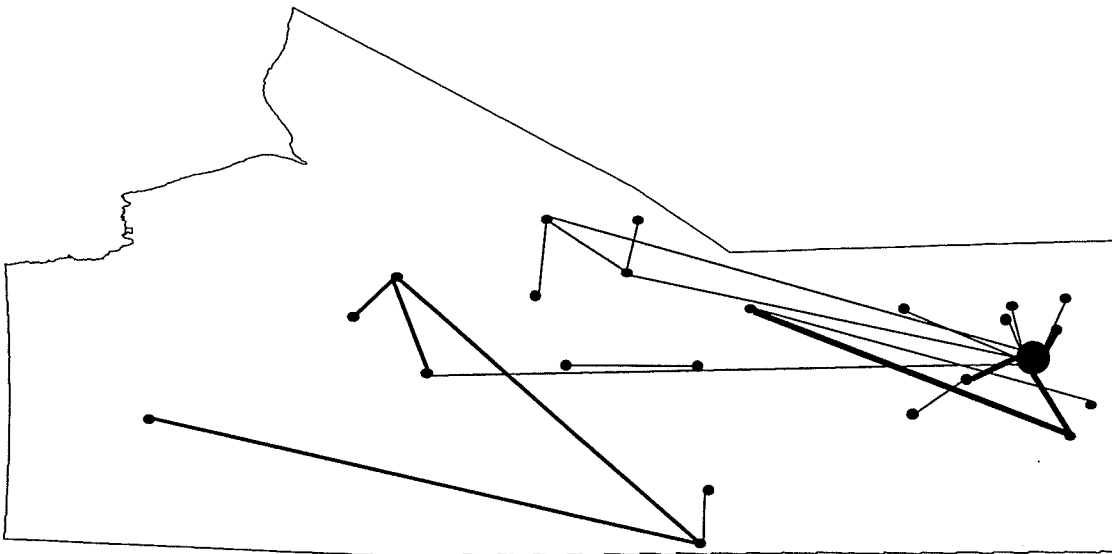


Figure 5.10: Geographic connections for networks of size 11, 12, 13, 14, 18 and 19.

Given these apparent biological differences, we reasoned that they may translate into ecological differences as well. We therefore examined our molecular data to determine if the different chlamydia types were distributed differently within networks. We were particularly interested in determining if the more common chlamydia genotypes might, for instance, be able to survive more readily in small networks, given their potentially higher transmissibility. In contrast, rarer chlamydia genotypes may need the frequent transmission opportunities provided by large networks to be able to persist in a population. Genotypes D and E were the most common types identified in our dataset. Given their lower frequency, genotypes F, J, K, G, I/H and Ba were initially classified as rare (see chapter 3).

We compared the presence of rare and common genotypes with several network and demographic variables. In total, we analyzed six different variables as shown in table 5.1. A detailed explanation of each variable is provided in the table caption. There was no association between common and rare chlamydia genotypes with age, network size, partner number, or risk area. There was however, a significant difference in the distribution of rare and common genotypes with respect to the partner location of an infected person. Thirty-nine percent of the common genotypes identified were isolated from cases whose partners were all located within the same community, while rare genotypes were more likely to be isolated from cases whose partners were located either outside their community or who had chosen partners both within and outside of their community ( $\text{Chi}^2$ ,  $P=0.037$ ).

Table 5.1: Demographic analysis of rare and common genotypes present within our dataset. "Average age" was classified as the average age of individuals infected with a rare or a common genotype. "Average network size" is classified as the average size of networks with either a rare or a common genotype present. "Partner location" was classified as the location of the sexual contacts for individuals infected with either rare or common genotype. "Partner number" was classified as the number of partners that an individual named. "Risk area" was classified as the geographic location of individuals infected with a rare or common genotype.

Variable	Genotype Class		P value
	Rare	Common	
Average age	22 years	22 years	0.355
Average network size	8	10	
Gender : males	42(35%)	68(40%)	0.30
Females	79(65%)	99(60%)	
Comm. Mix: within	47 (39%)	85 (51%)	0.037*
outside/both	74 (61%)	81 (49%)	
Partner number : $\geq 3$	26 (21%)	34 (21%)	0.98
2	30 (25%)	42 (25%)	
1	65 (54%)	88 (54%)	
Risk area* : High risk	53 (44%)	81 (49%)	0.43
Low risk	68 (56%)	86 (51%)	

## 5.6 Geographic distribution of chlamydia molecular genotype

The inclusion of chlamydia molecular genotypes in this study was a unique element in confirming sexual relationships identified by contact tracing. However, genotype information also aids in analyzing large scale patterns of sexual networks, based on the provincial distribution of genotypes. Specific genotypes clustered over several communities may reveal connections that link these communities together, which may not have been initially evident based only on contact tracing data.

Geographic mapping was conducted on the isolates we genotyped in our dataset . Overall each genotype was widely dispersed throughout the province. In order to clarify the data, three maps were produced for 1) the most common types, D and E, 2) less frequently observed types F, G, J and 3) the variant genotypes Ba, I/H, K<sub>1</sub>, G<sub>1</sub>, D<sub>1</sub>, D<sub>2</sub>, D<sub>3</sub> and D<sub>4</sub> (Figures 5.12 to 5.14).

Common genotypes, D and E, were widely distributed throughout the province, across both high and low incidence areas, however they appear to be more prevalent in high risk areas. Type D was largely found in the eastern and northern parts of the province while type E was mainly found in northern and western areas. In general, these patterns reflect the known network connections. Type J partially overlaps genotype E in being present in the southwestern part of the province, however, at the time of this study, they did appear mutually exclusive as type J is present in communities slightly to the southeast of the communities containing E. Type F largely overlaps type D, being present in the northern and eastern parts of the province.

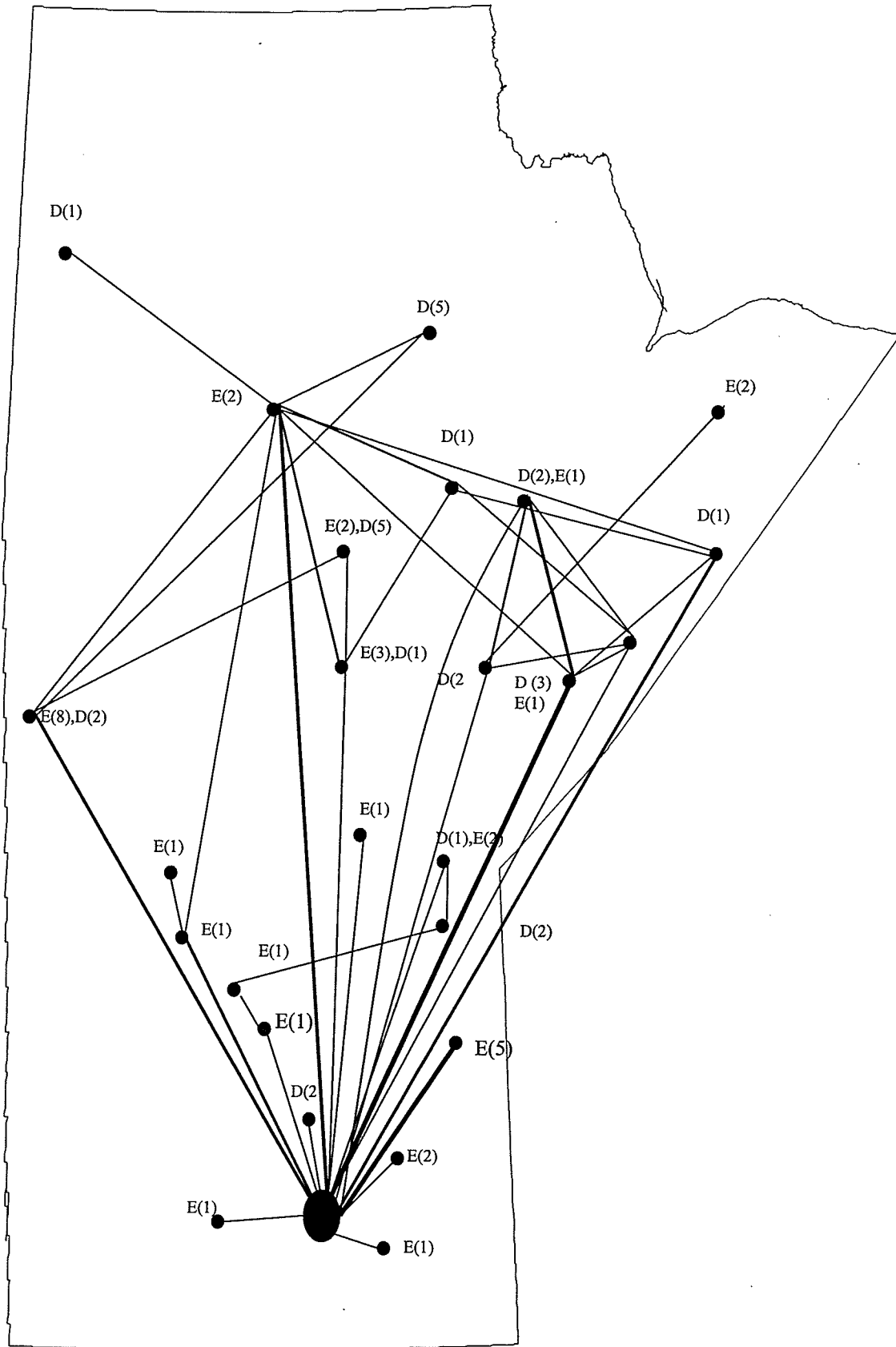


Figure 5.12: Geographic representation of common genotypes present in the province with connections between communities. Numbers in parentheses represent the number of genotypes identified at a given location.





Figure 5.13: Geographic representation of genotypes J, F and G, the least identified genotypes present in the province. Numbers in parentheses represent the number of genotypes identified at a given location.

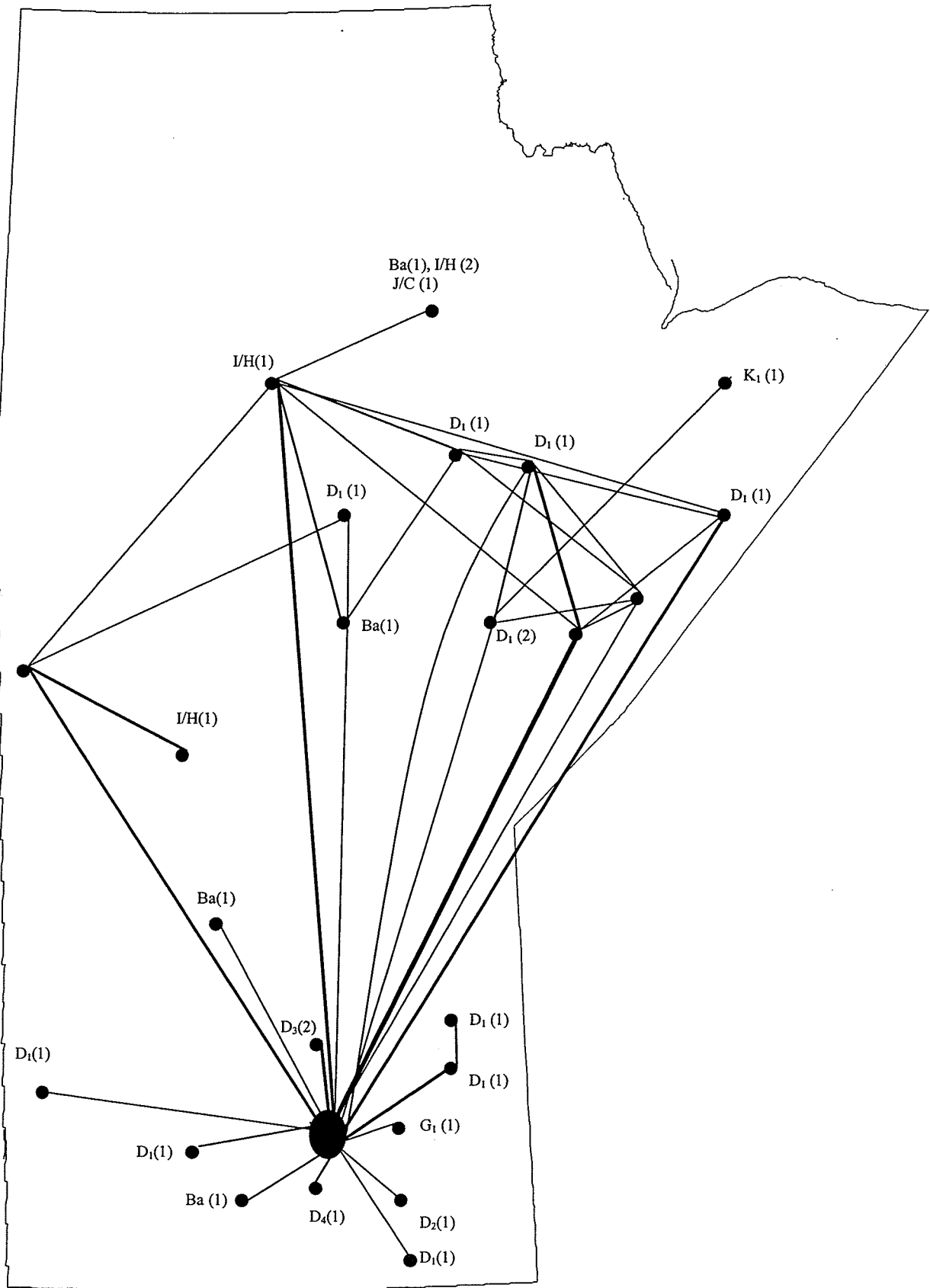


Figure 5.14: Geographic representation of chlamydia variant genotypes present in our dataset. Numbers in parentheses represent the number of genotypes identified at a given location.

Type G is restricted to the rural parts of the province to a small number of communities in central Manitoba. Variant D<sub>1</sub>, the most common D variant isolated, was largely confined to rural high risk areas, but was also relatively common in rural low risk areas. The variant genotypes isolated less frequently (D<sub>3</sub>, D<sub>4</sub> and G<sub>1</sub>) were identified within rural low risk areas.

Genotypes isolated from within Winnipeg were also graphed and are represented in figure 5.15 and 5.16. Types D, E, F and J were generally isolated from central and northwestern areas of the city. These latter two genotypes are relatively rare in rural Manitoba, suggesting that they would have been circulating for some time in Winnipeg prior to their relatively recent transmission to rural Manitoba. Given the minor presence of either genotype isolated from rural areas, network connections between the two areas allowed for the transmission of both types across the province.

Within Winnipeg, variant genotypes were dispersed much more than the common genotypes (Figure 5.16). D<sub>1</sub> has been transmitted throughout the majority of the city, while genotypes I/H and Ba are more confined to central and northern sections of the city. K<sub>1</sub> was frequently isolated from different sections of the city as well, in particular central and southwestern portions.

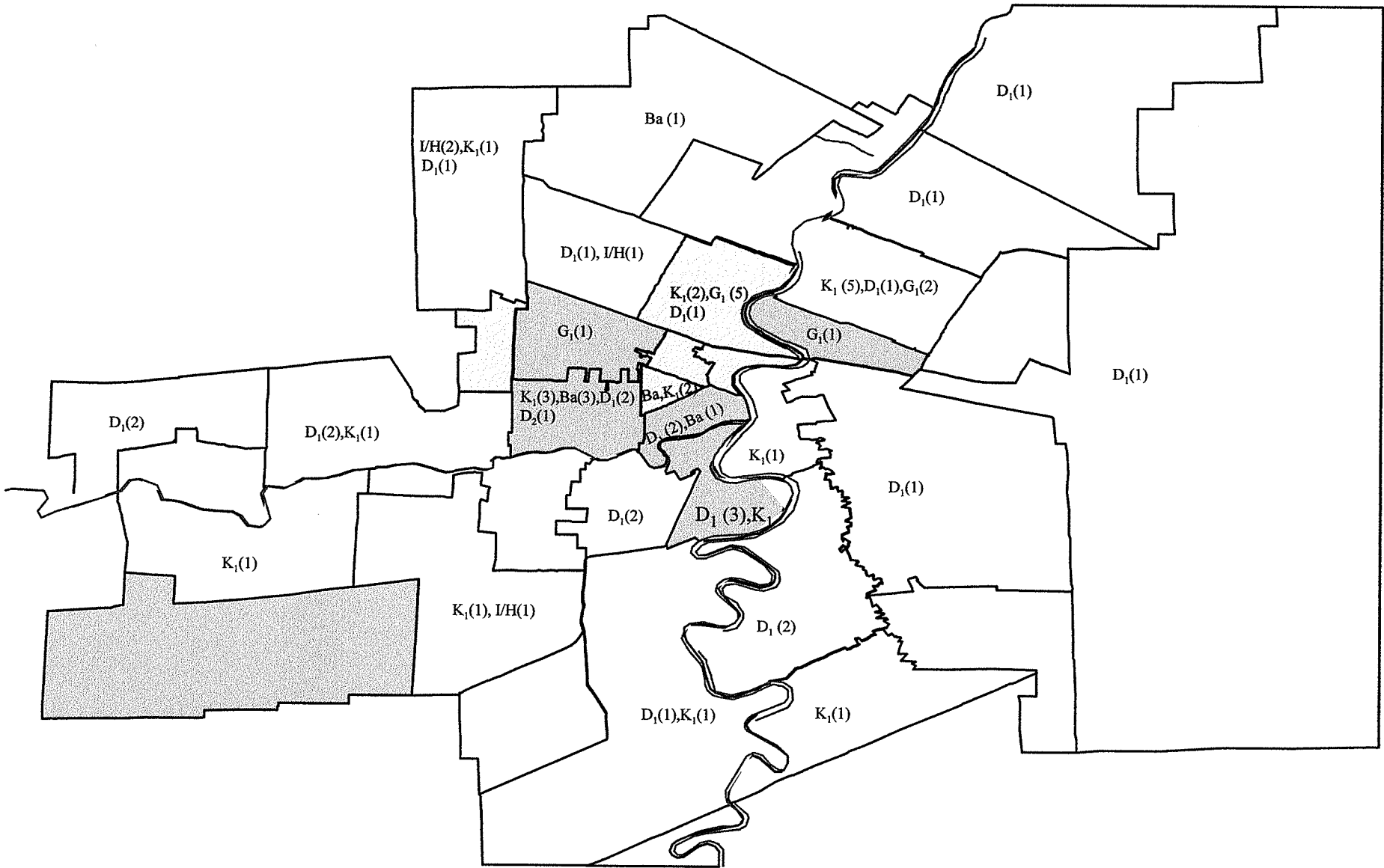


Figure 5.16: Geographic analysis of variant chlamydia genotypes within Winnipeg. Numbers in parentheses represent the number of genotypes identified. Symbols:  core  adjacent  peripheral sections of Winnipeg.

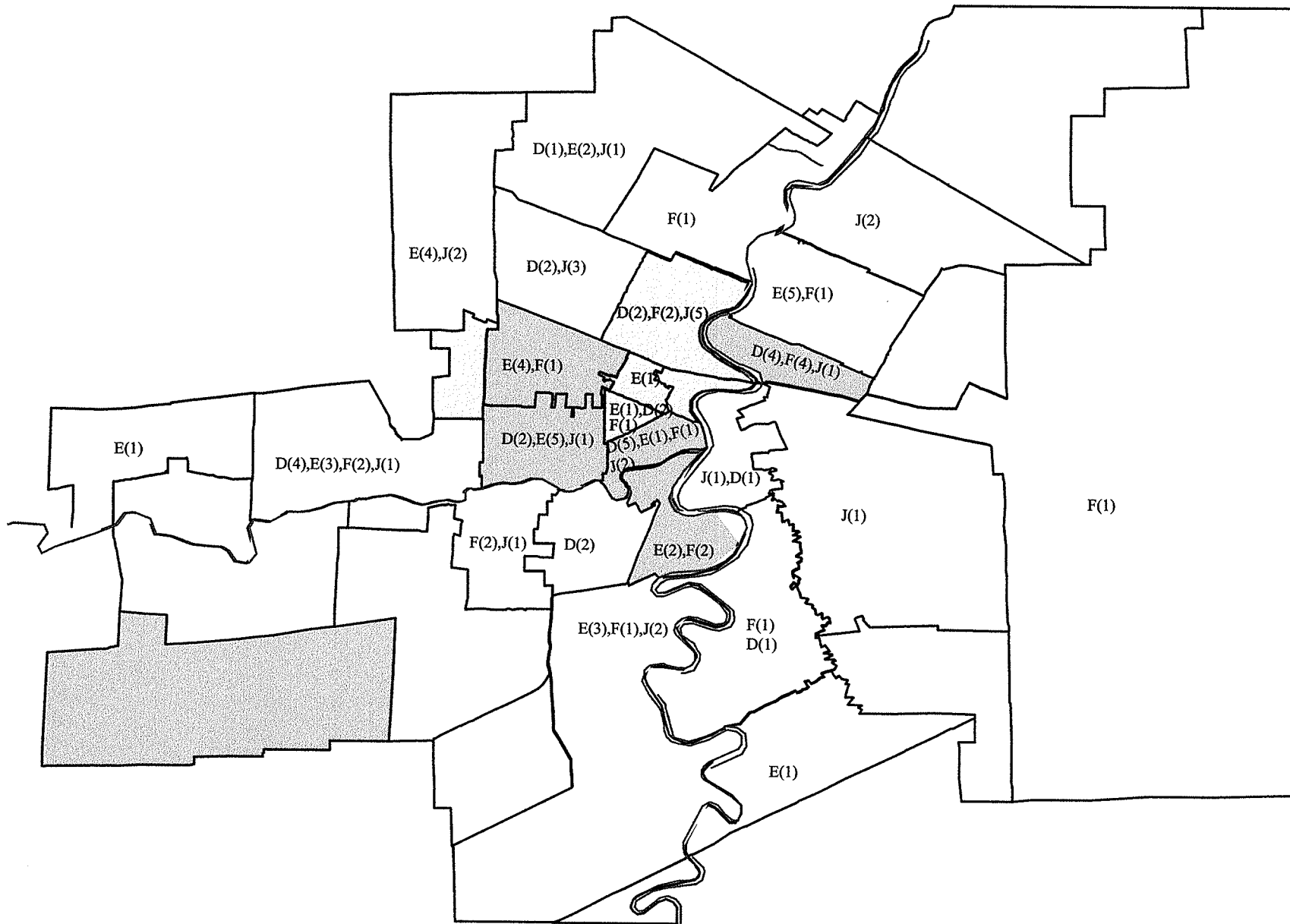


Figure 5.15: Geographic analysis of chlamydia genotypes D, E, F and G within Winnipeg. Numbers in parentheses represent the number of genotypes identified. Symbols:  core  adjacent  peripheral sections of Winnipeg.

This analysis has provided us with a larger overview of the distribution of specific chlamydia genotypes within several communities in the province, as well as potential sexual links between communities. D<sub>1</sub> is present in high risk rural areas, some low risk areas and most sections within Winnipeg. There are obvious sexual connections between these two areas allowing the isolation of D<sub>1</sub> variants. There is, potentially, a larger network within high risk rural areas and within all of Winnipeg, which allows for the maintenance of D<sub>1</sub>. Frequent connections between the two areas has allowed the continual transmission of D<sub>1</sub> across and into low risk rural areas. The same situation can be seen for K<sub>1</sub>, with their presence in Winnipeg and only one within high risk rural areas. Direct connections were not identified based on our data, however, this suggests that K<sub>1</sub>, through its constant movement in Winnipeg networks, has recently entered high risk areas. Ba and I/H isolates are more restricted within Winnipeg and may have only recently entered the high risk areas of Winnipeg.

These maps are useful in order to show foci of infection in the population. Identifying individual genotypes within communities helps to potentially show larger network patterns. Further research of this kind will help to cluster individual components together to reveal larger network patterns and identify which community connections identified by network analysis are associated with actual transmission.

## 5.7 Predicting positivity from component size vs k plex

Previously, Wylie and Jolly (2001) have demonstrated that network positivity is correlated with network structure. In examining large networks, between 10 and 82 people in size, they identified two structural types, radial and linear. Radial networks showed a hub and spoke structure with one central individual directly connected to many sexual partners. Linear networks showed a branching pattern consisting of many individuals, each of whom generally has 2 or 3 sex partners. Radial networks contained significantly less infected individuals than linear networks. We wished to determine whether this pattern might extend to smaller networks as well. We formalized our structural designation by using a network measure called k-plexes of size 3 (3-plex). A 3-plex consists of three people connected together in a line. A radial network of a given size, contains more possible 3-plexes than a similarly-sized linear network, therefore, the number of three plexes for a network reflects a structural characteristic. The 3-plex measure also incorporates some element of size as a larger network also contains more 3-plexes than a smaller network. We measured 3-plexes for each network in our dataset containing 3 or more people (networks of 2 contain no 3-plexes by definition). We then conducted a linear regression of 3-plex number against positivity and compared these results to a regression of network size against positivity. Our intention was to determine if network structure vs. network size was a better predictor of the number of infected people found in a network.

Figures 5.17 and 5.18 illustrate the number of positive cases regressed against either 3-plex number or component size. There is a strong positive correlation between positivity and both k-plex number and network size. Network size appears to provide a better fit to the data as it explains a greater amount of the variation in positivity ( $r^2 = 0.86$  for network size and 0.72 for k-plex number). Notably, there are five networks that are clear outliers on the 3-plex regression line (Figure 5.17). These are the largest radial components in our dataset and are the ones previously included in the analysis by Wylie and Jolly (2001).

## 5.8 Discussion

Patterns of STD rates have been highly correlated with geography. Populations infected with STDs may exhibit areas of high incidence and areas of low incidence. It is within high incidence areas where the risk of acquiring and transmitting infections are greatest, particularly as a function of core groups in those locations. Manitoba and the city of Winnipeg are also geographically segmented in terms of different levels of STD rates. Within Manitoba, north and northeastern sections are considered high risk areas. Winnipeg has the majority of chlamydia infected cases present within central areas of the city, or the core area. In order for a comprehensive understanding of the sexual connections occurring within the province, a network map of Manitoba was generated to visualize the extent to which connections occur between communities (Figure 5.1). The overall geographic distribution of sexual connections revealed that extensive bridging occurs between communities. Being the largest city of the province, the majority of connections directly link Winnipeg and rural Manitoba. Of interest are the numerous



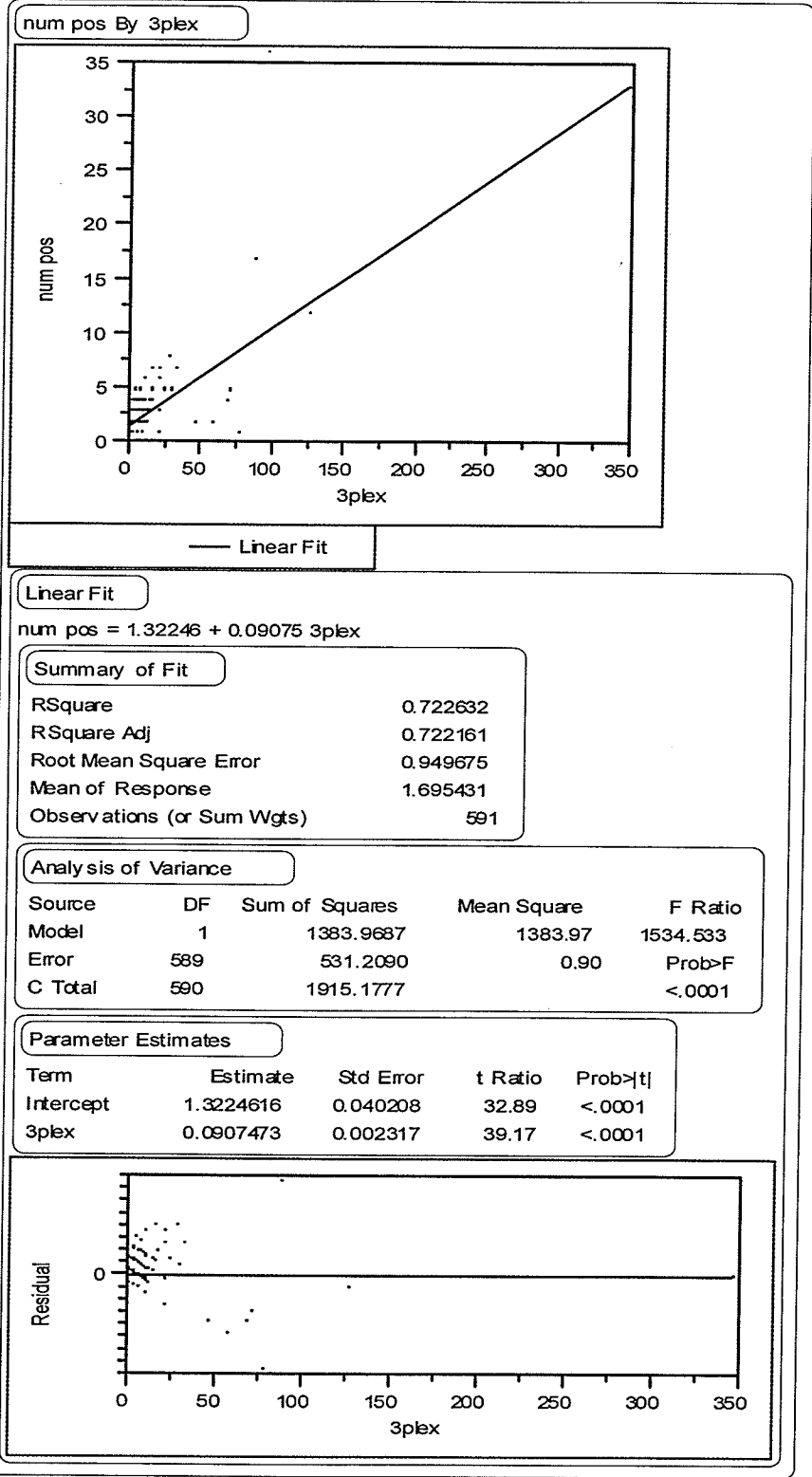


Figure 5.17: Linear regression data of the number of 3-plexes against number of positive chlamydia cases identified within networks. Dyads were not included as they do not consist of 3-plex structures.

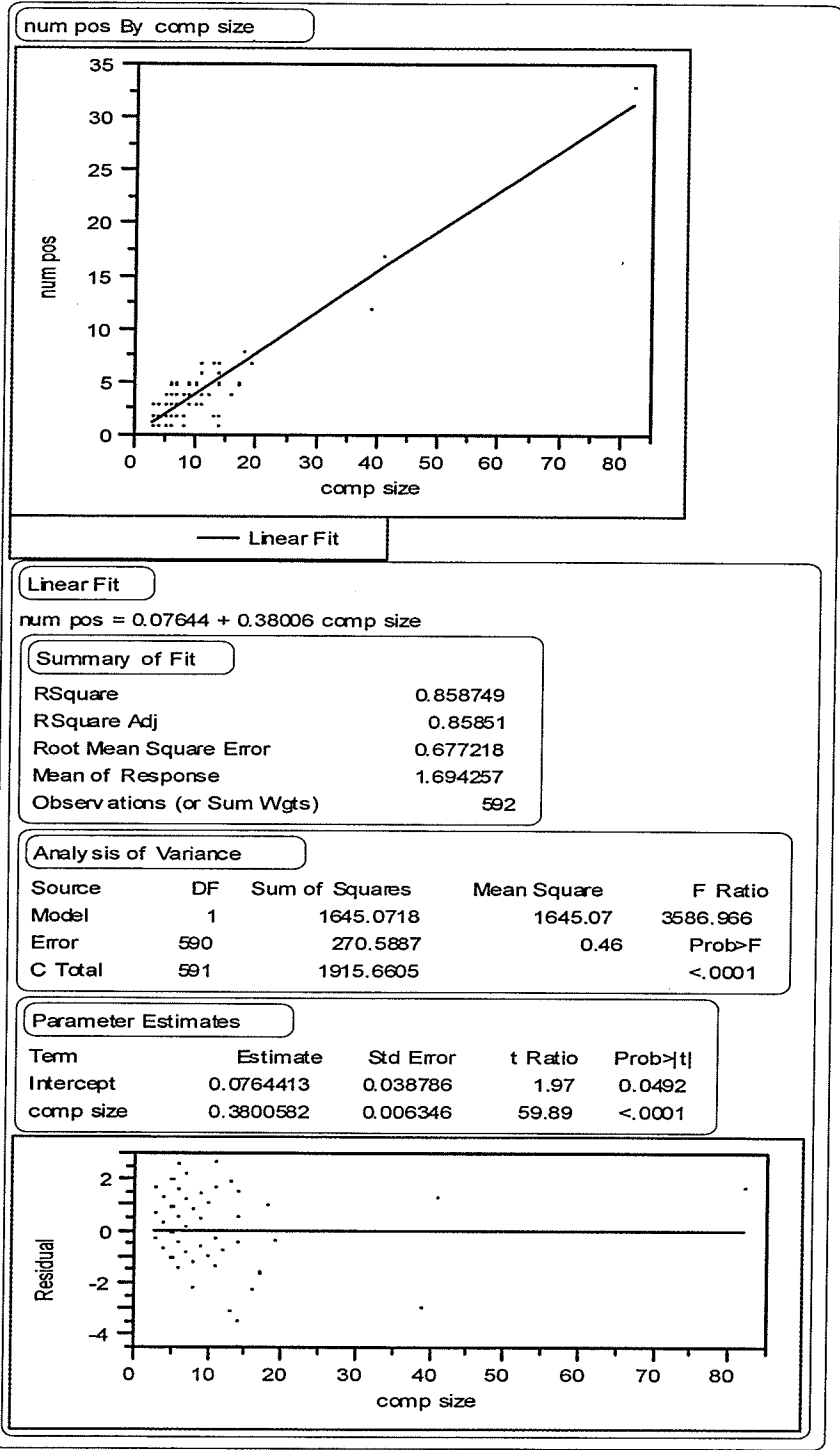


Figure 5.18: Linear regression data of component size and the number of positive cases identified within networks. For comparison with Fig 5.16, dyads were not included as they do not have 3-plex structures.

connections that occur between the high risk rural areas and Winnipeg. These high risk areas within Manitoba contain many isolated populations with limited access to the health care system and high population densities (Blanchard *et al.*, 1998). These characteristics are also risk markers for individuals involved in disease maintenance and spread, such as core groups. Our findings suggest that because there is frequent movement of individuals between rural areas and Winnipeg, there is the potential for rapid transmission of infections between the different areas.

To better understand the extent to which cases in Winnipeg choose their partners with respect to community location, we quantified the intra- and inter-area connections (Fig 5.2). Sixty one percent of all connections occurred between individuals in either high risk rural areas only or low risk Winnipeg locations only. This characteristic alone could result in endemic maintenance of chlamydia infections within those respective areas. Twenty four percent of all sexual connections occurred between all four risk areas, with most of these occurring between Winnipeg high risk areas and Winnipeg low risk areas. Overall, Manitoba has a high degree of sexual partner mixing within and across high and low risk areas. If in fact the high risk areas within the province are acting as reservoirs for STDs in the province, the number of bridging events occurring may be enough to effectively transmit infections. Targeting individuals who serve as the bridges between communities would confine infections within certain populations and could make it easier for public health officials to combat chlamydial infections within Manitoba.

Another risk marker associated with high risk areas is that of individuals coinfecting with STDs. In our dataset, information on the infectious status of all cases with chlamydia and gonorrhea were available. We mapped all networks containing at least one person who was coinfecting to study their distribution (Figure 5.3). Compared to the overall geographic map, coinfecting networks were largely restricted to high risk rural areas, such that most of the connections identified in western Manitoba are lost. This data is consistent with the idea that interrupting the flow of chlamydia between communities within core groups could help to reduce the incidence in these areas.

An additional geographic analysis was conducted by analyzing geographic patterns based on network size. Our data confirmed earlier findings that larger networks are most likely to be present within high risk areas. As the networks get smaller, larger numbers of connections are seen, which span the whole province. Therefore, further research into identifying patterns of larger networks would be helpful in targeting individuals responsible for endemic maintenance and spread of chlamydial infections.

Molecular genotyping of chlamydia isolates was conducted in our study. As with other studies (Ramirez, 2000), the genotypes most frequently isolated were D and E. These genotypes were classified as common, and the remainder of the genotypes were classified as rare. Reasons for the different proportions of genotypes represented within different populations are not known. We attempted to correlate the differences in prevalence with demographic and network variables (Table 5.1). There was no association with genotype prevalence and age, partner number, gender and risk area. There was a significant

difference with respect to partner location. Common genotypes were more frequently isolated from individuals who chose their partners from within the same community, while rare genotypes were more frequently isolated from individuals with partners either outside or both within and outside of their community. This pattern could reflect some large scale differences in the distribution of chlamydia present in different networks at the time of our study. It is also conceivable that common genotypes are better able to maintain endemic transmission within a community and hence would be more likely to be isolated from multiple members of that community (i.e. the "within" population). Rarer types may quickly die out of an immunologically experienced population and we may be more likely to observe transmission of rare types into different communities representing new immunologically inexperienced (with respect to that chlamydia genotype) hosts. Further research is required to determine if this relationship would consistently hold and its significance.

Geographic representation of the different chlamydia genotypes was also conducted. For this analysis, categories for genotypes were common (D and E), rare (F, G and J), and variant (Ba, I/H, K<sub>1</sub>, D<sub>1</sub>-D<sub>4</sub> and G<sub>1</sub>). As expected, based on the number of sexual connections across the province, genotypes D and E are distributed throughout both high and low risk areas in Manitoba (Figure 5.12). Both genotypes are common in northern and eastern Manitoba, while genotype E also circulates in several communities towards the southwest of Manitoba. Within Winnipeg, D and E was found largely in core and northern parts of the city. Genotypes D and E provide some information on larger scale network patterns. The prevalence of D along the eastern and northern sides of the

province suggest a linkage between many of the networks in these high risk areas. Genotype E also appears to have circulated in several networks connecting some Interlake communities. Additionally, several networks appear to interconnect throughout the core and northern portions of Winnipeg, with relatively few connections to other parts of the city.

In contrast to D and E, in Winnipeg, genotype K<sub>1</sub> appears to have been circulating largely in core and southern portions of the city, suggesting another distinct series of connected networks exist (Figure 5.16). With one exception, it was absent from rural Manitoba. Compared to K<sub>1</sub>, D<sub>1</sub> was isolated from many different locations in Winnipeg, suggesting either that it was circulating in a different set of connected networks, or it was present in both the connected networks defined by K<sub>1</sub> and those defined by D and E. The remaining rare variant genotypes have largely a random scattered distribution throughout Winnipeg and rural Manitoba.

This genotype distribution is only one snapshot in time. There is no previous work of this type in Manitoba to compare how the circulation patterns have changed over time. A study of gonorrhea strain distribution in Seattle provided a temporal look at the changes in auxotypes over time (Knapp *et al.*, 1987). These authors describe how emergence and persistence of specific strains in a community allows for a better understanding of how frequently new strains enter a population and how long a given strain remains in that particular population. These techniques also provide the means for improved

understanding of the circulation of STDs within specific subsets of individuals within a community.

Our data suggests that different chlamydia strains are generally widespread, reflecting the highly mobile and interconnected population in Manitoba. Before we can fully understand how certain chlamydia strains are transmitted in the community, it will be necessary to more closely link real-time chlamydia genotyping with newly diagnosed infections. This approach may provide further insights into the epidemiology of chlamydia and may facilitate development of innovative measures for control of this infection.

The last analysis we conducted was focused on predicting the number of chlamydia infected cases within components. Previous studies have suggested that the more k-plexes present in a network, the higher the transmission of infections throughout that network (Rothenberg *et al.*, 1998). We conducted a linear regression of 3-plex number or network size against positivity. For both variables, there was a positive linear relationship. The highest  $r^2$  was obtained with component size.

Observed on the 3 plex regression line were five outliers, which represented the largest radial networks (networks of size 13, 14, 16, and 17). These networks have fewer positive cases than expected given their k-plex number. Wylie and Jolly (2001) conducted an in-depth analysis of these and other large networks and found that large linear networks contained significantly more infected persons than radial networks. Given the emergence

of these networks as outliers, it appears that the relationship found by Wylie and Jolly (2001) may be unique to larger networks. Individuals in large radial networks, because of their numerous sexual partners, may have a low probability of transmitting an infection to most of their contacts. It is also conceivable that the five central individuals in the five large radial networks in our study recognized the risk associated with sexual behavior and used condoms on a regular basis, thus lowering the transmission probability within their sexual networks.

For most networks, it appears that size alone is a slightly better predictor of the number of positives. This does not rule out the possibility that the relationship with  $k$ -plex number is non-linear and more advanced model-fitting techniques may provide different answers. Notably, as shown in the previous chapter, there does appear to be differences in the network measures for closeness and betweenness in individuals in high and low risk areas. Different network measures may, therefore, also produce a better fit to the data than 3-plex measures.



## SUMMARY

The main objective of this research was to observe the extent to which molecular typing data for *Chlamydia trachomatis* would agree with the proposed transmission routes within the sexual networks constructed from contact tracing data. In addition, we described the demographic, geographic and behavioral patterns found within sexual networks in Manitoba to identify future research priorities and provide data necessary to develop new targeted STD control programs in Manitoba.

The majority of the constructed networks containing individuals linked either directly or indirectly by sexual contact were concordant with respect to chlamydia genotype (Chapter 3). This indicates that network analysis based on routinely collected data frequently does reflect the transmission routes of individual chlamydia genotypes. Considerable genetic diversity in the number of chlamydia genotypes circulating within the province of Manitoba was identified, where many of these appear to have been circulating in the province for at least six years. Relatively few chlamydia variants were identified. The lack of concordancy seen in some networks may reflect the identification of core groups or, alternatively, identify areas where STD control programs need to be changed or modified. The geographic connections identified by network analysis do frequently reflect transmission events as 59% of networks, containing individuals from more than one geographic location, were concordant with respect to chlamydia genotype present. Network analysis could therefore be used to identify different areas in a region where it would be useful to co-ordinate STD control efforts.

Characterization of chlamydia infected cases was also conducted and served to provide baseline data for Manitoba in terms of identifying persons involved in disease maintenance and spread (Chapter 4). Mixing patterns of infected individuals, measured by mixing matrices, were analyzed, as specific patterns affect the rate and degree of disease spread in a population. Assortative mixing primarily maintains infections in certain sub-populations while disassortative mixing primarily spreads infections to other sub-populations. Significant mixing patterns were observed for male and female cases present in high risk areas. These individuals had a tendency to choose partners who were within different age groups, and for males, to choose contacts with greater numbers of sexual partners, in comparison to cases in low risk areas. This highlights the extent to which control programs need to focus efforts specifically towards males and cases within high risk areas.

Demographic data was analyzed for cases within our dataset to characterize those populations most burdened with chlamydia infections. Several significant associations were found which contribute to high STD incidence rates. Cases young in age, male and located in high risk areas, are on average, present within large sexual networks. In addition, high risk area cases have more partners located within the same community, maintaining infections in that population. STD control efforts should focus on these groups, which in turn would lead to a decrease in the number of partners a given individual may have, interruption of chlamydia transmission and a decrease in network

sizes present in the population. Our results also highlight the need for further research on the role of mixing patterns and network size and structure on STD incidence patterns.

The ongoing discussion regarding the correlation between geography and STD rates lead us to a geographic analysis of sex partner formation (Chapter 5). We represented all geographic sexual connections at a provincial level, as well as those occurring within the city of Winnipeg. Overall, there is extensive sexual connections that span the province, Winnipeg and across both high and low risk areas. Of particular interest are the connections linked to high risk areas. Many of the identified connections linked Winnipeg and high risk areas. These high risk areas are mainly isolated populations with limited access to the health care system with high population densities. These factors are some risk markers associated with individuals involved in disease maintenance and spread, such as core groups. Individuals within Winnipeg also exhibit high connectivity between high and low risk urban areas. Frequent intra- and inter-area mixing enhance the potential for the spread of infections. Our findings suggest that there is frequent movement of individuals between rural areas and Winnipeg, resulting in rapid transmission of infections throughout the province.

In addition, geographic network connections for co-infected networks and different network sizes revealed that co-infected and large networks are mainly found within high risk areas. Therefore, further research into identifying patterns of larger networks would help in targeting individuals responsible for endemic maintenance and spread of chlamydial infections.

The molecular genotypes identified were used to correlate the differences in prevalence of certain genotypes with demographic and network variables. In addition, geographic distribution of genotypes was also analyzed. For the demographic analysis, the most frequently isolated genotypes were classified as common and the remaining genotypes were classified as rare. There were no significant associations for any of the variables except for partner location. In this case, common genotypes were isolated more frequently from cases whose partners were located within the same community, while rare genotypes were isolated from cases whose partners were most often located either outside, or within and outside of their community. Common types may be better able to maintain endemic transmission within a community while rare types may require more frequent exposure to immunologically inexperienced host. Further research is required to determine if this relationship would consistently hold and its significance.

Analysis of the geographic distribution of the different chlamydia genotypes was conducted. The emergence and persistence of certain strains in a community allows for a better understanding of how frequently new strains enter a population and how long a given strain remains in that particular population. There is no previous work of this type in Manitoba to compare how the circulation patterns have changed over time. Although there were different patterns of distribution for specific strains, further work needs to be conducted to more fully understand how certain chlamydia genotypes are transmitted in the community.

Possible associations between positivity (outcome variable) and component size and number of 3-plexes (explanatory variables) were tested using linear regression. For both explanatory variables, a positive linear relationship was identified, with the highest  $r^2$  obtained for component size. Therefore, size alone appears to be a better predictor of the number of positives found in a network. This does not rule out the possibility that the relationship with k-plex number is non-linear and more advanced model-fitting techniques may provide different answers. In addition, different network measures may produce a better fit to the data than 3-plex measures.

Overall, the use of social network analysis was very useful in studying the epidemiology of chlamydia in Manitoba. Population based characterization of infected individuals provides additional insight into behaviors over and above those gained by analyses based only on individuals. The use of contact tracing data in the construction of networks appears to be quite accurate as determined by the molecular epidemiology component of this research. Not only do contact tracing data and molecular data help to verify true sexual connections, they also allow analyses on demographic, behavioral and geographic patterns of chlamydia infected cases. This work, and additional work of this type, can provide useful information to aid control programs in targeting individuals involved in disease maintenance and spread, whether within certain age groups, geographic locations, or persons with high numbers of sex partners. In the end, it provides useful information for the development of appropriate prevention and control programs for chlamydia within Manitoba.

## REFERENCES

- Anderson RM. The transmission dynamics of sexually transmitted diseases: the behavioral component. In: Wasserheit JN, Aral SO, Holmes KK, Hitchcock PJ, editors. Research issues in human behavior and sexually transmitted diseases in the AIDS era. 1st ed. Washington, D.C.:American Society for Microbiology, 1991:38-60.
- Anderson RM. Some aspects of sexual behavior and the potential demographic impact of AIDS in developing countries. Soc Sci Med 1992; 34(3): 271-280.
- Anderson RM and Garnett GP. Mathematical models of the transmission and control of sexually transmitted diseases 2000;27(10):636-643.
- Anderson RM, Gupta S, Ng W. The significance of sexual partner contact networks for the transmission dynamics of HIV. J Acquir Immune Defic Syndr 1990;3:417-429.
- Anderson RM and May RM. Infectious diseases of humans: Dynamics and control. 1<sup>st</sup> ed. Oxford, England: Oxford University Press, 1992.
- Aral SO., et al. Demographic and societal factors influencing risk behaviors, *In*, Research issues in Human behavior and sexually transmitted diseases in the AIDS era. Wasserheit JN., et al. (eds). Washington DC, American Society for Microbiology, 1991.
- Aral SO. Sexual network patterns as determinants of STD rates: Paradigm shift in the behavioral epidemiology of STDs made visible. Sex Trans Dis 1999; 26(5): 262-265.
- Aral SO, Holmes KK. Epidemiology of sexual behavior and sexually transmitted diseases. In: Holmes KK, Mardh P, Sparling PF, Wiesner PJ, Cates W, Lemon SM, editors. Sexually transmitted diseases. 2<sup>nd</sup> ed. Toronto:McGraw-Hill Information Services Company, 1990:19-36.

Aral S, Holmes K. Social and behavioral determinants of the epidemiology of STDs: industrialized and developing countries. *In Sexually transmitted diseases*, 3<sup>rd</sup> ed., Holmes et al., 1999, pp. 39-76.

Auerbach DM, Darrow WW, Jaffe HW, Curran JW. Cluster of cases of the acquired immune deficiency syndrome: patients linked by sexual contact. *Am J Med* 1984;76:487-492.

Batagelj, V., and A. Mrvar. PAJEK - Program for large network analysis. Available at [http://vlado.fmf.uni\\_lj.si/pub/networks/pajek/](http://vlado.fmf.uni_lj.si/pub/networks/pajek/)

Bechtel GA., et al. Family, culture and health practices among migrant farm workers. *J Commun Health Nurs* 1995; 12:15.

Becker, KM, Glass GE, Brathwaite W, and Zenilman JM. 1998. Geographic epidemiology of gonorrhea in Baltimore, Maryland, using a geographic information system. *Am. J. Epidemiol.* 147:709-716.

Beller M, Middaugh J, Gellin B, Ingle D. The contribution of reinfection to gonorrhea incidence in Alaska, 1983 to 1987. *Sex Trans Dis* 1992;19:41-46

Berkowitz SD. *An introduction to structural analysis; the network approach to social research.* 1982; 1<sup>st</sup> ed., 1-24. Toronto, Canada: Butterworths.

Bickell NA., et al. Human papillomavirus, gonorrhea, syphilis and cervical dysplasia in jailed women. *Am J Public Health* 1991; 82:1318.

Blanchard JF, Moses S, Greenway C, Orr P, Hammond GW, and Brunham RC. The evolving epidemiology of chlamydial and gonococcal infections in response to control programs in Winnipeg, Canada. *Am J Public Health* 1998; 88:1496-1502.

Breakey *et al.* Health problems of homeless men and women in Baltimore. JAMA 1989; 262:1352.

Brewer T., *et al.* Migration, ethnicity and environment: HIV risk factors for women on the sugar cane plantations (bateyes) of the Dominican Republic. AIDS 1998; 12(14): 1879-87.

Brunham RC, Kimani J, Bwayo J, Maitha G, Maclean I, Yang C, Shen C, Roman S, Nagelkerke NJD, Cheang M, and Plummer FA. 1996. The epidemiology of Chlamydia trachomatis within a sexually transmitted diseases core group. J. Infect. Dis. 173:950-956.

Brunham, R., Yang C, Maclean I, Kimani J, Maitha G, Plummer F. 1994. Chlamydia trachomatis from individuals in a sexually transmitted disease core group exhibit frequent sequence variation in the major outer membrane protein (*omp1*) gene. J. Clin. Invest. 94:458-463.

Brunham R. Core group theory: A central concept In STD epidemiology. Venereology 1997; 10:34-39.

Brunham RC, Plummer FA. A general model of sexually transmitted disease epidemiology and its implications for control. Medical Clinics of North America 1990;74:1339-1352.

Bureau of communicable Disease Epidemiology. 1989 Canadian guidelines for screening for *Chlamydia trachomatis* infections. Can Dis Wkly Rep 1989; 15S5:1-11.

Caldwell JC and Caldwell P. The demographic evidence for the incidence and cause of abnormally low fertility in tropical Africa. World Health Stat Q 1983;36:2.



Calzavara LM, Bullock S, Myers T, Marshall VW, Cockerill R. Sexual partnering and risk of HIV/STD among aboriginals. *Can J Pub Health* 1999; 90(3): 186-191

Carswell JW, et al. Prevalence of HIV-1 in east African lorry drivers. *AIDS* 1989;3:759-761.

Catania JA, Binson D and Stone V. Relationships of sexual mixing across age and ethnic groups to herpes simplex virus-2 among unmarried heterosexual adults with multiple sexual partners. *Health Psych* 1996;15(5):632-370.

Cates W, Jr., Wasserheit JN. Genital chlamydial infections: epidemiology and reproductive sequelae. *Am J Obstet Gynecol* 1991;164:1771-1781.

Catlin BW. Nutritional profiles of *Neisseria gonorrhoea*, *Neisseria meningitidis* and *Neisseria lactamica* in chemically defined media and the use of growth requirements in gonococcal typing. *J infect Dis* 1973;128:178-194.

CDC. HIV infection, syphilis, and tuberculosis screening among migrant farm workers: Florida, 1992. *MMWR* 1992; 41:723.

CDC. Juvenile detention. Atlanta, 1996.

CDC. Syphilis screening among women arrestees at the Cook County jail: Chicago, 1996. *MMWR* 1998; 47:432-433.

Centers for Disease Control. Gang-related outbreak of penicillinase-producing *Neisseria gonorrhoea* and other sexually transmitted diseases: Colorado Springs, Colorado, 1989-1991. *MMWR* 1993; 42(2):25-28.

Centers for Disease Control and Prevention, Division of STD prevention. Sexually transmitted disease surveillance, 1995. Atlanta: U.S. department of health and human services, Public health service; September 1996.

Centers for Disease Control and Prevention, Division of STD prevention. Chlamydia in the U.S. Atlanta: U.S. department of health and human services, Public health service; April 2001.

Claman P, Toye B, Peeling RW, Jessamine P, Belcher J. Serologic evidence of chlamydia trachomatis infection and risk of preterm birth. *Can Med Assoc J* 1995; 153:259-262.

Cohen D., *et al.* The potential role of custody facilities in controlling sexually transmitted diseases. *Am J Public Health* 1992; 82:552-556.

Cohen I, Veille J, Calkins B. Improved pregnancy outcome following successful treatment of Chlamydial infections. *J Am Med Assoc* 1990; 263:3160-3163.

Cohen MS *et al.* Host defenses and the vaginal mucosa. *Scand J Nephrol* 1985;86:13-22.

Coste J, Laumon B, Bremond A, Collet P, Job Spira N. Sexually transmitted diseases as major cause of ectopic pregnancy: results from a large case-control study in France. *Fertil Steril* 1994; 62:289-295.

Cullen TA, Helgerson SD, LaRuffa T, Natividad B. *Chlamydia trachomatis* infection in native American women in a Southwestern tribe. *J Fam Pract* 1990;31:552-554.

D'Costa LJ, *et al.* Prostitutes are a major reservoir of sexually transmitted diseases in Nairobi, Kenya. *Sex trans Dis* 1985; 12:64.

Day S. Prostitute women and AIDS. *Anthropology* 1988; 2:421.

Day S. Do people lie but gene sequences tell the truth? Possible uses of molecular epidemiology in molecular anthropology. *N Med Anthropol Rev* 1994;2:20-26.

Elliot L, Blanchard J, Beaudoin C, Green C, Nowicki D, Matusko P. Geographic variations in the epidemiology of bacterial sexually transmitted diseases in Manitoba, Canada: Implications for phase-specific control strategies 1999. Unpublished data.

Ellen JM, Hessol NA, Kohn RP. An investigation of geographic clustering of repeat cases of gonorrhea and chlamydial infections in San Francisco, 1989-1993: evidence for core groups. *J Infect Dis* 1997; 175:1519-1522.

Fairley CK, Bowden FJ, Gay NJ, Paterson BA, Garland SM, Tabrizi SN. Sexually transmitted diseases in disadvantaged communities- too many partners, too little treatment? *JAMA* 1997;278:117-118.

Freeman, Everett and Borgatti. UCINET 5.0 for Windows: Software for Social Network Analysis 1999 © Analytic Technologies, inc. and the University of Greenwich.

Friedman SR, Neaguis A, Jose B, Curtis SR, Goldstein M, Ildefonso G, Rothenberg RB, Des Jarlais DC. Sociometric risk networks and risk for HIV infection. *Am J Pub Health* 1997; 87:1289-1296.

Garnett GP, Hughes JP, Anderson RM, et al. Sexual mixing patterns of patients attending sexually transmitted diseases clinics. *Sex Transm Dis* 1996; 23(3): 248-57.

Ghani, AC, Donnelly CA, and Garnett GP. 1998. Sampling biases and missing data in explorations of sexual partner networks for the spread of sexually transmitted diseases. *Stats Med.* 17:2079-2097.

Goeree R, Gully PR. The burden of chlamydial and gonococcal infection in Canada. In: Ronald AR, Peeling RW, editors. *Prevention of infertility*. Ottawa, Canada:Research

studies of the Royal Commission on New Reproductive Technologies. Minister of Supply and Services Canada, 1993:29-76.

Golub AL and Johnson BD. Crack's decline: Some surprises across U.S. cities. National Institute of justice, 1997.

Gorbach PM, Sopheab H, Phalla T, *et al.* Sexual bridging by Cambodian men: potential importance for general population spread of STD and HIV epidemics. Sex Trans Dis 2000; 27:320-6.

Guaschino S, De Seta F. Update on Chlamydia trachomatis. Ann NY Acad Sci 2000;293-300.

Gupta S, Anderson RM, May RM. Networks of sexual contacts: implications for the pattern of spread of HIV. AIDS 1989; 3:807-817.

Hamers FF, Peterman TA, Zaidi AA, Ranson RC, Wroten JE, Witte JJ. Syphilis and gonorrhea on Miami: Similar clustering, different trends. Am J Public Health 1995;85:1104-1108.

Han, Y., F. B. Coles, A. Muse, and S. Hipp. 1999. Assessment of a geographically targeted field intervention on gonorrhea in two New York state counties. Sex. Trans. Dis. 26:296-302.

Handsfield HH., *et al.* Localized outbreak of penicillinase-producing *Neisseria gonorrhea*: A paradigm for introduction and spread of gonorrhea in a community. JAMA 1989; 261:2357-2361.

Handsfield HH. Recent development in STD's: I Bacterial diseases. Hospital practice 1991;47-56.

Handsfield HH, Jasman LL, Roberts PL, Hanson VW, Kothenbeutel RL, Stamm WE. Criteria for selective screening for *Chlamydia trachomatis* infection in women attending family planning clinics. JAMA 1986;255(13):1730-1734.

Haraldsdottir S, Gupta S, Anderson RM. Preliminary studies of sexual networks in a male homosexual community in Iceland. J Acquir Immune Defic Syndr 1992;5:347-381.

Harary F, Norman RZ. Graph theory as a mathematical model in social science. Ann Arbor, Mich. Institute for social research, 1953.

Havanon N, Bennett A, Knodel J. Sexual networking in provincial Thailand. Stud Fam Plann 1993;24:1-17.

Health Canada, Bureau of HIV/AIDS, STD and TB, Population and Public health branch. Canadian Sexually Transmitted Disease surveillance report 1998/1999. CCDR; 26 (2000).

Hethcote HW and Yorke JA. Gonorrhoea transmission dynamics and control. Lecture notes in Biomathematics. Berlin:Springer-Verlag, 1984.

Hofferth et al. Premarital sexual activity among U.S. teenage women over the past three decades. Family Plann Perspect 1987;19:46.

Hoffman JP, Su SS, Pach A. Changes in network characteristics and HIV risk behavior among injection drug users. Drug and Alcohol Dependence 1997;46:41-51.

Holmes KK: Lower genital tract infection in women: urethral, cervical and vaginal infections, *In* Sexually transmitted diseases, 3<sup>rd</sup> Ed, Holmes KK *et al.* (eds) 1998. New York, McGraw-Hill, 527-546.

Holmes KK. Human ecology and behavior and sexually transmitted bacterial infections. Proc Natl Acad Sci USA 1994;91:2448-2455.

Hudson CP. Concurrent partnerships could cause AIDS epidemics. *Int J STD & AIDS* 1993;4:249-253.

Hunt CW. Migrant labor and sexually transmitted disease: AIDS in Africa. *J Health and Social Behav* 1989; 30:353-373.

Ison CA, Whitaker L, Renton A. Concordance of auxotype/serovar classes of *Neisseria gonorrhoea* between sexual contacts. *Epidemiol Infect* 1992;109:265-274.

Johnstone H., et al. Incidence of sexually transmitted diseases and Pap smear results in female homeless clients from the Chicago Health Outreach Project. *Health Care Women Int* 1993; 14(3):293-299.

Jolly A, Orr P, Hammond G Young TK. Risk factors for women undergoing testing for *Chlamydia trachomatis* and *Neisseria gonorrhoea* in Manitoba, Canada. *STD* 1995; 22(5):289-95.

Jolly AM, Wylie JL. Sampling individuals with large scale sexual networks: An evaluation of four approaches. *Sex Trans Dis* 2001;28:200-207.

Jones et al. HIV - related characteristics of migrant workers in rural South Carolina. *S Med J* 1991; 84:1088.

Judson FN. Assessing the number of genital chlamydial infections in U.S. *J Reprod Med* 1995;30(Suppl1):269-272.

Klovdahl AS. Social networks and the spread of infectious diseases: the AIDS example. *Soc Sci Med* 1985;21:1203-1216.

Klov Dahl AS, Potterat JJ, Woodhouse DE, Muth SQ, Darrow WW. Social networks and infectious disease: the Colorado Springs Study. *Soc Sci Med* 1994;38:79-88.

Knapp JS, Tam MR, Nowinski RC, Homes KK, Sandstrom EG. Serological classification of *Neisseria gonorrhoea* with use of monoclonal antibodies to gonococcal outer membrane protein I. *J Infect Dis* 1984;150:44-48.

Koenig ER. International prostitutes and transmission of HIV. *Lancet* 1989; 1:782-783.

Koopman JS and Lynch JW. Individual and causal models and population system models in epidemiology. *Am J Public Health* 1999; 89(8):1170-1174.

Lan J, Melgers I, Meijer CJ, Walboomers JM, Roosendaal R, Burger C, Bicker OP, van den Brule AJ. Prevalence and serovar distribution of asymptomatic cervical *chlamydia trachomatis* infections as determined by highly sensitive PCR. *J Clin Micro* 1995;33: 3194-3197.

Latkin C, Mandell W, Oziemkowska M, Vlahov D, Celentano D. The relationships between sexual behavior, alcohol use, and personal network characteristics among injecting drug users in Baltimore, Maryland. *Sex Trans Dis* 1994; 21(3): 161-167.

Laumann EO, et al. The social organization of sexuality: Sexual practices in the United States. Chicago and London, The University of Chicago press, 1994.

Lauman EO and Youm Y. Racial/Ethnic group differences in the prevalence of sexually transmitted diseases in the United States: A network explanation. *Sex Trans Dis* 1999; 26(5):250-261.

Lin JS, Donegan SP, Heeren TC, Greenberg M, Flaherty EE et al. Transmission of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* among men with urethritis and their female sex partners *J Infect Dis* 1998;178:1707-1712.

McCormack WM, Alpert S, McComb DE, Nichols RL, Semine DZ, Zinner SH. Fifteen month follow-up study of women infected with *Chlamydia trachomatis*. New England J Med 1979;300:123-125.

MacDonald N, Wells GA, Fisher WA, Warren WK, King MA, Doherty JA, Bowie WR. High risk STD/HIV behavior among college students. JAMA 1990;263:3155-3159.

Maticka-Tyndale E. Sexual scripts and AIDS prevention: Variation in adherence to safer sex guidelines by heterosexual adolescents. J Sex Research 1991;28:45-66.

May RM. The transmission and control of gonorrhea. Nature 1981; 291:376-377.

May RM, Anderson RM. Transmission dynamics of HIV infection. Nature 1987;326:137-142.

Morris M, Padhistis C, Wawer MJ, et al. Bridge populations in the spread of HIV/AIDS in Thailand. AIDS 1996;10:1265-71.

Nagelkerke N, Brunham R, Moses S, and Plummer F. Estimating the effective rate of sex partner change from individuals with sexually transmitted diseases. Sex Trans Dis 1994; 21(4):226-230.

National center for health statistics:vital and health statistics. Fertility, family planning, and women's health: New data from the 1995 national survey of family growth. Series 23: Data from the national survey of family growth, no. 19. Hyattsville, MD, February 1997.

Neaguis A, Freidman SR, Curtis R, Des Jarlais DC, Furst RT, Jose B et al. The relevance of drug injectors' social and risk networks for understanding and preventing HIV infection. Soc Sci Med 1993;38:67-78.



Neagu A, Freidman SR, Jose B, Goldstein MF, Curtis R, Ildefonso G et al. High risk personal networks and syringe sharing as risk factor for HIV infection among new drug injectors. *J Acquir Immune Defic Syndr Hum Retrovirol* 1996;11:499-509.

O'Rourke M, Ison CA, Renton AM, Spratt BG. Opa-typing: A high resolution tool for studying the epidemiology of gonorrhoea. *J Clin Microbiol* 1995;17:865-875.

Orr P, Sherman E, Blanchard J, Fast M, Hammond G, Brunham R. Epidemiology of infection due to *Chlamydia trachomatis* in Manitoba, Canada. *Clin Infect Dis* 1994; 19:876-883.

Over M, Piot P. Human immunodeficiency Virus infection and other sexually transmitted diseases in developing countries: public health importance and priorities for resource allocation. *J Infect Dis* 1996; 174: s162-175.

Padian NS. Prostitute women and AIDS. *Epidemiology* 1988; 2:413.

Plummer FA, Simonsen JN, Cameron DW, Ndinya Achola JO, Kreiss JK, Gakinya MN, et al. Co-factors in male-female sexual transmission of Human immunodeficiency virus type 1. *J Infect Dis* 1991; 163:233-239.

Potterat JL, Zimmerman-Rogers H, Muth SQ, Rothenberg RB, Green DL, Taylor JE, Bonney MS, and White HA. Chlamydia transmission: Concurrency, reproduction number, and the Epidemic trajectory. *Am J Epidemiol* 1999; 150(12): 1331-1339.

Potterat JJ. "Socio-geographic space" and sexually transmissible diseases in the 1990's. *Today's Life Sci* 1992;December:16-31

Quinn TC; Gaydos C; Shepherd M; Bobo L; Hook EW; Viscidi R; Rompalo A.  
Epidemiologic and microbiologic correlates of *Chlamydia trachomatis* infection in sexual partnerships. JAMA 1996;276(21):1737-1742.

Rettig PJ. Perinatal infections with *Chlamydia trachomatis*. Clinics in Perinatology 1998;  
15:321-250.

Rice PA, Schachter J. Pathogenesis of pelvic inflammatory disease; what are the questions? J Am Med Assoc 1991; 266:2587-2593.

Rosenberg D, Moseley K, Kahn R, Kissinger P, Rice J, Kendall C, Coughlin S, and Farley TA. Networks of persons with syphilis and at risk for syphilis in Louisiana: Evidence of core transmitters. Sex Trans Dis 1999; 26(2): 108-114.

Rothenberg RB. The Geography of gonorrhea. Am J Epidem 1983;117(6): 688-694.

Rothenberg RB, Potterat JJ, Woodhouse DE, Darrow WW, Muth SQ, Klovdahl AS. Choosing a centrality measure: epidemiologic correlates in the Colorado Springs study of social networks. Social Networks 1995;17:273-297.

Rothenberg RB, Sterk C, Toomey K, Potterat J, Johnson D, Schrader M, and Hatch S. Using social network and ethnographic tools to evaluate syphilis transmission. Sexually transmitted diseases 1998;25:154-161.

Rothenberg R, Woodhouse DE, Potterat N, Muth SQ, Darrow WW, Klovdahl AS. Social networks in disease transmission: the Colorado Springs study. NIDA Res Monogr 1995; 151:3-19.

Ryan CA, et al. Explosive spread of HIV-1 and sexually transmitted diseases in Cambodia. Lancet 1998; 351:1175.

Sanchez J, et al. Gender differences in sexual practices and seroprevalence of sexually transmitted infections. *Am J Public Health* 1996; 86:1098-1107.

Sattenspiel L. Population structure and the spread of disease. *Hum Biol* 1987;59:411-438.

Scott J. *Social network analysis: A handbook*. London:Sage publications, 1991.

Schachter J, et al. Screening for chlamydial infections in women attending family planning clinics. *West J Med* 1983; 1387:375.

Service SK and Blower SM. HIV transmission in sexual networks: an empirical analysis. *Proc R Soc Lond B* 1995; 260:237-244.

Shiboski S, Padian NS. Population- and Individual-based approaches to the design and analysis of epidemiologic studies of sexually transmitted disease transmission. *J Infect Dis* 1996;174(S2):S188-S200.

Stamm WE, Holmes KK. *Chlamydia trachomatis* infections of the adult. In: Holmes KK, Mardh P, Sparling PF, Wiesner PJ, Cates W, Lemon SM, et al, *Sexually transmitted diseases*. 2<sup>nd</sup> ed. Toronto: McGraw-Hill Information services company, 1990:181-194.

Steiner MJ, Cates W and Warner L. The real problem with male condoms is nonuse. *Sexually transmitted diseases* 1999; 26(8): 459-461.

Stephens RS, Sanchez-Pescador R, Wagar A, Inouye C, Urdes MS. Diversity of *Chlamydia trachomatis* major outer membrane protein genes. *J Bacteriol* 1987;169:3879-3785.

Sturm-Ramirez K, Brumblay H, Diop K, Gueye-ndiaye A, Sankale JL, Thior I, N'Doye I, Hsiesh CC, Mboup S and Kanki P. Molecular epidemiology of genital *chlamydia*

*trachomatis* infection in high-risk women in Senegal, West Africa. J Clin Micro 2000; 38(1): 138-145.

Syrjanen K, et al. Sexual behavior of women with human papillomavirus (HPV) lesions of the uterine cervix. Br J Vener Dis 1984; 60:243.

Tanfer K *et al.* Condom characteristics: The perceptions and preferences of men in the United States. Fam Plan Perspec 1993;25:67-73.

Wasserheit JN, Aral SO. The dynamic topology of sexually transmitted disease epidemics: implications for prevention strategies. J infect Dis 1996; 174 Suppl 2: S201-13.

Woodhouse DE, Rothenberg RB, Potterat JJ, Darrow WW, Muth SQ, Klovdahl AS, Zimmerman HP, Rogers HL, Maldonado TS, Muth JB and Reynolds JU. Mapping a social network of heterosexuals at high risk for HIV infection. AIDS 1994;8:1331-1336.

Wylie, J. L., and A. Jolly. 2001. Patterns of chlamydia and gonorrhea infection in sexual networks in Manitoba, Canada. Sex. Trans. Dis.28:14-24.

Yang, C. L., I. Maclean, and R. C. Brunham. 1993. DNA sequence polymorphism of the *Chlamydia trachomatis omp1* gene. J. Infect. Dis. 168:1225-1230.

Yuan Y, Zhang YX, Watkins NG, Caldwell HD. Nucleotide and deduced amino acid sequences for the four variable domains of the major outer membrane proteins of the 15 *Chlamydia trachomatis* serovars. Infect and Immun 1989;57:1040-1049.

Zenilman JM, Elish N, Fresia A, Glass G. The Geography of sexual partnerships in Baltimore: Applications of core theory dynamics using a geographic information system. Sex Transm Dis 1999;26(2):75-81.

Zimmerman HL, Potterat JJ, Dukes RL, et al. Epidemiologic differences between chlamydia and gonorrhea. *Am J Public Health* 1990;80:1338-1342.