# Using Social Network Analysis to Identify Social Topology in Communities of Vulnerable People

James Tompkins

MGIS Capstone Project

The Pennsylvania State University

April 30th 2014

## Abstract

**Background:** Risk-based behaviors like injected drug use (IDU) and participation in the commercial sex trade can lead to heightened potential for transmission of bloodborne pathogens and sexually transmitted infections. To understand transmission patterns, social network analysis and spatial analysis methods have been used independently of one another. However, little consideration has been made to the role place has in social networks and how specific locations impact transmission risk.

**Methods:** A sample of vulnerable people was obtained through respondent driven sampling in Winnipeg, Canada from January to December, 2009. Spatial data were generated from intersections or site names respondents named in questions regarding their activities. Pajek network analysis software was used to assess the network data for the study sample, with participation locations linked to respondents, and betweenness centrality was calculated. The spatial network was mapped in ArcGIS software. A kernel density estimation (KDE) function was run on the count of points created from study responses. A KDE surface was then generated using betweenness centrality as a weight for the individuals and activity sites in each component to generate a social topology.

**Results:** Approximately half of respondents said they had ever used injected drugs. 101 unique sites were named by the 600 respondents as sites where they participated in risk-related behaviours. Analysis of the network revealed a giant component comprising 55% of the network, and key activity spaces in social topologies were observed. Spatial analysis showed a mean distance of less than 4 km existed between individuals and the people and places they were linked to.

**Conclusions:** This research demonstrates the importance of considering risk-related activity sites in assessing social networks for transmission risk. Inclusion of locations as social network nodes is shown to highlight a cohesive network among an RDS sample and identifies social topologies optimal for targeted intervention and prevention resources.

## Introduction

Sexually transmitted infections (STIs; such as chlamydia, gonorrhea, and infectious syphilis) and bloodborne pathogens (BBPs; such as HIV and Hepatitis C) are treatable diseases that have increased at an alarming rate over the last decade (World Health Organization, 2011; Centers for Disease Control and Prevention, 2013; Remis, 2007). Each year, 448 million new cases of curable STIs occur in adults throughout the world and many live unaware that they are infected (WHO, 2011; Public Health Agency of Canada, 2011). STIs and BBPs require close contact for transmission (PHAC, 2010). They are diseases with a significant basis in behavioural patterns; their spread through a population does not occur in an identifiably uniform or random distribution (Wylie *et al.*, 2005). Rather, it is often a function of complex social interaction between two individuals or a network of people. Epidemiologically, emphasis is placed on populations inclined to engage in risk-based behaviours like unprotected sex, heightened sexual promiscuity and use of injected drugs through needle sharing (Wylie *et al.*, 2005; De *et al.*, 2007).

To understand the transmission of these diseases, graph theory is applied to social networks. Through social network analysis, networks of people are modeled using graphs to understand personal connections to other people and places. In practice, researchers treat these interactions as potential paths of transmission, examining behaviours that make an individual susceptible to infectious diseases. Numerous studies (De *et al.*, 2004; De *et al.*, 2007; Bansal *et al.*, 2007) include measures of social aggregation in defining sociosexual networks with risk prevalence.

In order to identify larger behavioural patterns of STI epidemiology, these network components must be connected either to each other or other, larger groupings through undisclosed or unidentified links. Infection and recurrent infection in these social groups contributes to the idea that proximity is a significant factor in the endemic propagation of these pathogens. This concept has been explored several times by plotting case incidents in areal geographic units, often observing a ‘core population’ (geographic units with the top 50% of cases) (Rothenberg, 1983; Potterat *et al*., 1984; Becker *et al*., 1998). While geographic clustering of events is likely and often unsurprising in analysis of STI incidents, the fact that their spread is dependent on interaction means that social clustering (even where it involves significant distance between involved parties’ home locations) is as important as geographic clustering. Mapping linear connections between individuals has reinforced the idea of the ‘core population’ in STI/BBP incidence among sexual contacts and suggested that small distances are a determinant in risk behaviour participation among social contacts (Rothenberg et al, 2005; Shane, 2013).

However, individuals in the same sociosexual network may be located in geographically distant communities, seemingly isolated from each other but connected through the personal timelines and travel of others. De *et al*. (2004) linked STI surveillance records of individuals to the social venues contacts mentioned. Their results demonstrated that a gonorrhea outbreak in northern Albertan communities could be traced back to eighteen sociosexual network members who would not know each other normally but who met a sexual partner at the same bar; the bar itself created many links in a network that may otherwise not have formed.

The *place* that individuals engage in risk-related behaviours may be as much a part of the network of activity as the people they connect with. This notion was explored by Cummins *et al.* (2007), who suggested that application of place to health solutions has relied too much on the conventional, Euclidean view of space in the past. Instead a relational perspective to place and distance is suggested for handling concepts of location where individuals are concerned. For example, two people may continually live in very distant locations but their relational activity space – attending the same school for a period of time, visiting the same location while traveling – can bring them very close together. Embracing the idea that geographic position of resources or actors is not the only determinant, and that social networks and social power also impact transmission, can help realize relativistic distance in health solutions. Thus, the place that vulnerable individuals engage in risk-related behaviours may be as much a part of the network of activity as the people they associate with. This project continues the exploration of locations by examining whether mapping social behaviours is reflected in the prevalence of specific risk-related social venues.

## Methods

To reach these objectives, data was obtained from a 2009 study commissioned by the Canadian Institute of Health Research (CIHR) entitled “The behavioural, social and cultural factors affecting the epidemiology of sexually transmitted and bloodborne pathogens in high risk populations: Determining risk space in Canada’s populations”. Information from 600 individuals in Winnipeg, Manitoba was collected by a respondent-driven sampling (RDS) method as described by Jolly and Wylie (2013). This data did not include identifying information; records identified an individual only by the unique code on the RDS coupon they obtained. Participants were questioned by community health workers regarding their behaviours and individual risk factors. Data for each individual included gender, infection types (e.g. HIV, Hepatitis C, chlamydia, gonorrhea, and infectious syphilis), laboratory-confirmed infection status (e.g. positive, negative and inconclusive), and geographic information of current residence and activity sites (e.g. IDU, sex client pickup, sex worker pickup) reported as the name of the place or nearest intersection.

Geographic locations were manually geocoded in QGIS 1.8 ([www.qgis.com](http://www.qgis.com)) using Open Street Map ([www.openstreetmap.org](http://www.openstreetmap.org)) as a basemap. Location names were populated in the format “Street 1 / Street 2” and validated so the reciprocal combination did not appear in the processed data. Thus, each geographic location contained a latitude and longitude to represent either a street intersection or place name. A total of 719 locations were geocoded; 101 of these were unique location names among risk-related activity sites (IDU or sex trade transactions) and will be referred to as R1 to R101 where R denotes its role in risk behaviours.

Problems were encountered while geocoding due to naming conventions used by participants. For cases where the same place was referenced using two different names (e.g. Higgins Park versus the intersection known as Higgins Ave / Main Street), clarification was obtained from the health care worker who conducted the survey. The location was confirmed as green space at the corner of Higgins Avenue and Main Street (M. Ormand, personal communication, January 23, 2014). Therefore, mentions of Higgins Park were recoded as “Higgins / Main”. High geolocation rates were achieved for all site types, both residence (93.2%) and “hangouts” (IDU (85.1%), sites a sex client was met (84.8%) and sites a sex worker was met (87.5%)). Less than 9% of the locations could not be geocoded. Sites were not geolocated where guessing would have led to estimated coordinates. Sites were omitted spatially if misspelled street names were unidentifiable, if generic areas were named (e.g. “Main Street”, “all over”, “downtown”, or “riverbank”), and if two streets were parallel.

### Social Network Analysis

Network theory is applied to social relationships to measure social “distance” between individuals, the structure of a community of connected people, and to identify who the central figures are in that social structure. By including sites mentioned for risk-related activities as nodes in the social network, the significance of individuals and locations in STI/BBP exposure risk can be measured together. To this end, potential social connections that may not otherwise appear in coupon referrals are inferred from shared social behaviours practiced in the same locations.

Each individual’s participation in the “hangout” activity space was analysed using social network analysis methods. The data were imported into Pajek (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>) (Batagelj and Mrvar, 2011) and used to determine the relative importance of individuals and locations in the activity space within the network. In network theory, a node’s degree is often used to measure its relative importance by counting the number of edges (connections) linking it to other nodes. A network partition operation was run in Pajek to separate the study responses into connected components – subsets in which any two nodes are linked, and to which no additional nodes from other subsets can be connected – where each represented a network of associated people and places. Since the data collected could only link any individual with just one other respondent and a limited number of sites, another metric was needed to estimate social influence.

We chose betweenness centrality to represent the social importance of each node for its ability to represent how much of a network can be reached by established contacts with each person or location. Betweenness, succinctly represented as:

$$C\_{B}\left(υ\right)= \sum\_{s\ne υ\ne t\in V}^{}\frac{σ\_{st}(υ)}{σ\_{st}}$$

gives the proportion of times any node υ acts as a bridge along a path in the network between two other nodes, *s* and *t* (Brandes, 2001). Essentially, this allows us to observe the likelihood that any person or place allows opportunities for the risk of transmission through the network to additional people who may otherwise not be connected. Pajek betweenness values were exported and used to perform the spatial analysis.

### Spatial Analysis

The integration of social network analysis with spatial analysis was used to illustrate the social topology present among respondent networks. The outputs were used to understand the spatial concentration of the social networks and the distribution of respondents with respect to where the social topology was strongest.

The spatial coverage of components and their relation to risk-related activities were analyzed using several spatial analysis tools in ArcGIS 10.2 ([www.esri.com](http://www.esri.com)) (ESRI, Inc., Redlands, CA). The geographic extent of each component was determined by mapping the minimum bounding geometry of all respondents that were identified through the SNA above. An “XY to Line” operation was used to draw the edges of each network graph in geographic space, a visualization technique used successfully in a previous analysis of social contacts of HIV-positive persons (Rothenberg *et al*., 2005). Straight lines on the map represent Euclidean connections between the residences of linked individuals, as well as connections between an individual’s residence and the places that individual claimed to engage in risk-related behaviours.

Point pattern analysis methods such as kernel density estimation (KDE) have been successfully used in a number of studies (e.g. analyze agricultural impact of sinkholes (Aurit *et al*., 2013); crime patterns (Gerber, 2014); and cyclone frequency (Joyner and Rohli, 2010)) and was used in this study to highlight the geographic concentration of potential risk in each component. The KDE function fits a surface over each point in the dataset, using values of surrounding points within a defined radius to estimate values on unknown points in the area. We performed the operation twice for each connected component. The first used raw point counts (p = 1) to calculate the point values in the surface estimation. The second assigned the betweenness score of each point, aimed at highlighting locations of greatest importance within the sociospatial model. In each case, we used the ArcGIS default output cell size and search radius calculated by the KDE tool, defined as the shorter of the height or width of the data set’s spatial extents divided by factors of 250 and 30, respectively, for the two measurements.

## Results

The study population of 600 individuals was over half (53%) male and nearly three-quarters (73.3%) belong to aboriginal groups (First Nations (53.8%), Métis (19.2%) and Inuit (0.3%)). Over half of respondents stated they had injected drugs (50.5%) but only 16.2% provided a geographic location in which they injected drugs. Only 2 injected drug users said they had not injected in the last six months at all. Under half (47.2%) said they had sexual contact with someone outside of their close social network, while just over 10% said they had slept with a sex client or sex worker in the same period. Only 51 respondents who had sexual contact with a sex client or sex worker provided a response as to where they most frequently met those partners; 8 individuals provided unidentifiable spatial information. Roughly 12% refused serological testing for STIs and BBPs. Of those who consented to tests, laboratory results showed that injected drug use was practiced among 81.2% infected with HIV and 91.1% infected with Hepatitis C (HCV); just over 97% of those who were co-infected with HIV and (HCV) injected drugs.

Figure 1. Number of social networks by size (in number of nodes) (A) without geography (n = 600, N = 147) and (B) with geographic place included as connections (n = 701, N = 57)

|  |
| --- |
|  |
|  |

Fifty-seven components of size two or greater were identified with the largest containing 366 nodes (Figure 1b, Figure 2a). Overall, 44% of the components contained two nodes (dyads); 19% contained three nodes (triads) and 5 components were greater than size 10. 23 components (40.4%) included at least one geographic location. The largest component contained over two-thirds (68%) of the risk-behaviour activity locations. Only one of the five largest components (Component 7) did not have a single risk-related activity among its nodes (Figures 2a-e). 28% of dyads were comprised of one respondent linked to a location but no other people, while 36.4% of triads included one geographic place and two coded respondents. Absent of risk behaviour spatial information, the network was segmented into 147 networks (Figure 1a), the largest of which was only 46 nodes (Wylie and Jolly, 2013). In the previous analysis, the giant component accounted for less than 8% of the study sample; with locations this group included 55% of all nodes. While social connections could not be established for 128 of the 600 respondents (21.3%), inclusion of the risk behaviour activity sites demonstrated a much more cohesive network.

The five largest components (Figure 2) account for 72.9 percent of the nodes in the analysis. Among these groups, the mean distance between any two connected respondents or a respondent and a site for risk-related activities was 3.45 kilometers. This is comparable to the mean distance in Component 1 alone (3.7 km), where the site R101 was as far as 37.4 kilometers from some of its connections. Twenty-two percent of connections were less than 100 metres, and forty-four percent were under a kilometer apart.

Figure 2. Components larger than size 10

|  |  |  |  |
| --- | --- | --- | --- |
| Component | Social Network | **Geographic Extent** | **Highest Betweenness in Component** |
| (A)Component 1 (n = 366) |  |  |

|  |  |
| --- | --- |
| Node | Betweenness |
| R101 | 0.52459732 |
| 84 | 0.487359125 |
| 115 | 0.458573 |
| 239 | 0.45347 |
| 171 | 0.453319 |
| R55 | 0.442058809 |
| 592 | 0.429493 |
| 107 | 0.425018 |
| R45 | 0.406562 |
| 95 | 0.313019 |

 |
| (B)Component 2(n = 45) |  |  |

|  |  |
| --- | --- |
| Node | Betweenness |
| 17 | 0.618393 |
| 29 | 0.544397 |
| 2 | 0.531712 |
| 10 | 0.502114 |
| 59 | 0.502114 |
| 76 | 0.491543 |
| 50 | 0.459831 |
| 104 | 0.332981 |
| 109 | 0.30444 |
| 119 | 0.285412 |

 |
|  |  |  |  |
|  |  |  |  |
| (C)**Component 3****(n = 36)** |  |  |

|  |  |
| --- | --- |
| Node | Betweenness |
| 12 | 0.695798 |
| 34 | 0.583193 |
| 1 | 0.426891 |
| R5 | 0.420168 |
| 269 | 0.393277 |
| 204 | 0.391597 |
| 28 | 0.258824 |
| 65 | 0.216807 |
| 162 | 0.213445 |
| 56 | 0.164706 |

 |
| (D)**Component 7****(n = 21)** |  |  |

|  |  |
| --- | --- |
| Node | Betweenness |
| 395 | 0.65789474 |
| 271 | 0.56315790 |
| 330 | 0.52631579 |
| 197 | 0.51052632 |
| 138 | 0.41578947 |
| 507 | 0.284211 |
| 198 | 0.27894737 |
| 460 | 0.194737 |
| 323 | 0.1 |
| 459 | 0.1 |

 |
| (E)**Component 29****(n = 13)** |   |  |

|  |  |
| --- | --- |
| Node | Betweenness |
| 446 | 0.7424242 |
| 470 | 0.53030303 |
| 436 | 0.4393939 |
| 550 | 0.3030303 |
| R31 | 0.1666667 |
| 449 | 0.1666667 |
| 435 | 0.1666667 |

 |

Component 1 reflected the demographic splits of the full dataset, with males making up 55.4% of the network and First Nations and Metis people representing 75.9% of the 298 respondents. Of those that consented to testing for STI/BBP, 31.2% were infected with HCV and 6.7% were HIV-positive; 5% were co-infected with HCV and HIV. Over half of injected drug users in Component 1 were infected with HCV and 9% were co-infected (summarized in Table 1). The table indicates a high “unknown” rate across the five components for participation in risk behaviours. This is the proportion of the study sample who did not give a location for this activities – it includes both nonparticipants and those who did not disclose this information.

**Table 1:** Characteristics of components 1, 2, 3, 7 and 29.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Component | Sex*% Male**% Female**% Other* | Activity*% IDU**% Sex Trade**% Both**% Unknown* | Disease*% with STI**% with BBP**% Both**% Refused* | Ethnicity*% Caucasian**% First Nations**%Métis**% Other* | Mean Distance / Component Area |
| 1 | 55.4%42.3%2.0% | 21.5%5.7%3.4%69.5% | 4.7%32.6%13.1%0.7% | 20.8%56.4%19.5%3.0% | 3.7 km296 km2 |
| 2 | 55.8%44.2%0.0% | 4.7%0.0%0.0%95.3% | 2.4%41.9%11.6%0.0% | 44.2%27.9%18.6%9.3% | 2.4 km46 km2 |
| 3 | 46.9%50.0%3.1% | 15.6%0.0%0.0%84.4% | 3.1%28.1%3.1%3.1% | 56.3%28.1%15.6%0.0% | 2.3 km51 km2 |
| 7 | 61.9%38.1%0.0% | 0.0%0.0%0.0%100% | 0.0%9.5%9.5%0.0% | 23.8%52.4%9.5%14.3% | 3.2 km23 km2 |
| 29 | 54.5%45.5%0.0% | 9.1%18.2%0.0%72.7% | 9.1%27.3%9.1%0.0% | 27.3%36.4%18.2%18.2% | 2.7 km6 km2 |

All five of the largest components had a mean distance less than four kilometres between any two connected respondents or a respondent and a site for risk-related activities. This reflects the compact spatial nature of the majority of the social connections. Several pockets of activity occurred throughout the city, with individual sites (e.g. R101, at the western extent, R45 and R55 in downtown Winnipeg) that stand out as foci of activity. The densest areas of social activity fall within the overlap of all five components, coinciding with the core of Winnipeg (Figure 3a). R101, with the highest betweenness score in Component 1, stands out as a point of social influence among Winnipeg’s vulnerable population despite its position outside of the city’s limits. The two other sites of highest betweenness in Component 1 (R45 and R55) together form the area of strongest social influence in both the full activity space (Figure 3a) and the betweenness-weighted density surface for Component 1 (Figure 3b).

Figure 3. (A) Spatial Extents of Components 1, 2, 3, 7 and 29 with density of all activity locations. (B) shows the density of Component 1 betweenness values for the core of Winnipeg.



Using the betweenness values, activity spaces for each of the major components were identified. The key activity spaces within each are highlighted in red and yellow, indicating their importance to their own social group (Figure 4). In these social networks, only two risk-related activity sites have betweenness values higher than zero: R5 (Component 3) and R31 (Component 29).

Figure 4. The betweenness-weighted density surface for (A) Component 2, (B) Component 3, (C) Component 7 and (D) Component 29, in the same geographic area as Figure 3b.



## Discussion

Our study focuses on the social importance of specific locations used for illicit activities among a network of vulnerable people in Winnipeg, Canada. We demonstrate that spatial analysis using social network metrics can be used to better understand the social topology that enables STI/BBP transmission among IDUs and commercial sex participants. As has been shown in other cities, the activities of vulnerable people was concentrated in certain neighbourhoods (Brouwer *et al.*, 2013).

To observe the role of place in a community of vulnerable persons, we focus on Component 1 (Figure 3b). The strength of these social activity sites is evident in their relative betweenness scores, several of which rank among the highest in the network (Figure 2). Without the strong influence of several sites (R45, R55 and R101 in particular), the network would split into more groups, until it matched the size and number of components as the original study by Jolly and Wylie (2013). Using social geography to establish network connections in this way demonstrates a larger connected community of people within the RDS sample not seen previously. This is supported by the density surface for the other large components (Figure 4). Only two of these networks include sites in the social topology and in all four instances, the surface focuses on individual points. These smaller network components lack any suggestion of a larger community with a focus of social activity like that exhibited in Component 1.

The presence of over two-thirds of risk-related sites in Component 1 demonstrates the importance of gathering information on locations used in risk-related behaviours. As shown, a topology generated by minimal sites of activity suggests neighbourhoods or locations which improve the understanding of potential disease transmission. Figure 5b stresses this overlap, where a concentration of people in the southwest and northeast areas of the downtown core indicate overlap of four (yellow) or all five (orange) of the largest components. These regions coincide with the neighbourhoods highlighted in the maps of social density. Most research involving spatial analysis of infectious disease has relied on temporally fixed residential addresses, but this strongly supports the need for comprehensive collection of activity locations in epidemiological studies (Martinez *et al.*, 2013; Brouwer *et al.*, 2013). Figure 5b and the position of high-betweenness sites in Figure 5c suggest that additional information from participants could potentially link other components to Component 1.

Figure 5. Spatial co-incidence of components for (A) the entire study area. The downtown core highlights (B) the number of components at each set of coordinates and (C) the position of high-betweenness nodes across all five large components.



### Limitations

There are limitations to this approach in the nature of the RDS sample. Foremost, connections between a respondent and more than two people will only occur if multiple coupon recipients submitted to the study, leading to an underestimation of one-degree links and overestimation of connections made through multiple degrees (Rudolph *et al.*, 2013). Likewise, survey questions only asked what site or sites were used most frequently for risk-related behaviours in the previous six months. Although the study was conducted by a healthcare worker trusted by many participants, the potential for bias through self-reporting of personal risk behaviours exists.

The number of answers permitted or encouraged may also present opportunities for bias. Only 14% of respondents that named a location identified more than one. Nine people answered “all over” to questions about location and another four refused to name any sites. While it has been shown that vulnerable populations achieve a perception of minimal risk by having smaller activity spaces (Martinez *et al.*, 2013), assembling a social topology should endeavour to capture as much information on individuals’ participation in activity space as possible. Even with generalized geographic descriptions like nearest intersections, more complete personal activity histories will identify overlap in sites where risk is shared.

As noted in previous research into the risk spaces of vulnerable populations, future studies should collect more detailed information from participants about the geographic locations of risk behaviours (Martinez *et al*., 2013).The context of timing and how frequently participants visit locations is a dimension of the social topology of vulnerable peoples that needs further investigation. Furthermore, it should be noted that while the use of space in social network data established otherwise missing connections, participants may not make habitual visits to the same venues. People who do visit the same location may not do so simultaneously or together. This exploratory approach uses generalized spatial information to represent connections through place, and demonstrates a cohesive community at work in the RDS. More detailed records of participants’ activity space could be used to verify this finding.

The spatial precision of the model is also affected by the inexact nature of the response formats. Some respondents provided the closest pairing of major cross streets to where they lived, injected drugs, or looked for commercial sex; others provided names of smaller local, or ‘side’, streets. It is impossible to determine exactly what degree of error is introduced by open-ended questions that allow identification through the “nearest intersection”. This could be overcome by adopting the interactive method of identifying sites through landmarks used by Brouwer *et al.* (2013). Future methodologies could address both issues of spatial context and precision, and assess sites individuals have prolonged, direct contact with, by leveraging personal activity journals or smartphone GPS. Prefaced with the value of identifying areas of highest social risk through which diseases may pass, individuals may be willing to better identify sites.

## Conclusions

Previous work on the geography of STI and BBP epidemiology has used a combination of line symbology and chloropleths (maps that assign graded coloring schemes to differentiate between characteristics of areal units) to depict patterns in disease (Rothenberg et al, 2005; Rothenberg, 1983; Potterat et al, 1985). While the visualization of social network edges illustrates how social distance is covered geographically, lost in the density of links is which locations are most active. We expanded this thinking to identify areas of risk by including locations where risk behaviours are in evidence as part of the social network data. Our findings support other work that has identified short distances or small activity spaces among communities of vulnerable people (Rothenberg *et al.*, 2005; Shane, 2013; Martinez *et al.*, 2013). However, we also show the first indication that a large, cohesive community is represented in the RDS sample.

It should be noted that it is impossible, based on the level of detail in the spatial information, to discern sites that are truly shared. The relationship these individuals have with each other through the locations they visit for risk-related behaviours is estimated based on their perception of what is near where they live, inject drugs, and meet with sex trade partners. But the number of mentions some intersections receive suggest that there are small neighbourhoods or regions which may influence the social topology (Figure 4a).

Figure 6. (A) Sites and respondents displayed by risk activity type, and compared with presence of (B) Hepatitis C and (C) HIV infection. Symbol size is determined by betweenness score of the node at that location.



Observed alongside the high BBP infection rate in the study cohort (Figure 6), these patterns suggest that a geography akin to the ‘core population’ identified by Rothenberg *et al* (1983, 2005) is present in the activity space of vulnerable people. Although caution must be used in interpretation of the social topology generated, this approach demonstrates potential for using the geographic foci of risk-related behaviours to model vectors of disease spread. Strong evidence is shown that the RDS sampled a community of connected people, reflecting findings by Rudolph *et al.* (2013) that networks and neighbourhoods of drug and sex risk behaviours will reach more individuals at risk.

Using a spatial approach to the epidemiology of infectious diseases can have several implications on resource use. Identification of neighbourhoods of highest risk-related activity rates can allow the most efficient placement of prevention and intervention programs and resources, like syringe exchanges, disposal and health treatment facilities. It is also applicable to other highly infectious diseases that can spread through social contact. Locations of contact and social topology could be used to model transmission of other diseases, such as norovirus, measles, mumps, rubella and tuberculosis – all of which can spread through shared food, close contact or pronounced periods of shared airspace.

## Acknowledgements

Dr. Ann Jolly and Dr. Justine Blanford provided invaluable guidance, support, time, effort, and feedback on the analyses as well as direction integrating social and spatial aspects of the model. Not to mention patience. I would like to recognize the contributions of Dr. John Wylie for the laboratory data and insight into the information on infection status. The authors would also like to thank the Canadian Institute of Health Research, and Margaret Ormand for her effort and data collection work in Winnipeg, Manitoba, Canada.

## References

Aurit, M. D., Peterson, R. O., & Blanford, J. I. (2013). A GIS analysis of the relationship between sinkholes, dry-well complaints and groundwater pumping for frost-freeze protection of winter strawberry production in Florida. *PloS One*, *8*(1), e53832. doi:10.1371/journal.pone.0053832

Bansal, S., Grenfell, B. T., & Meyers, L. A. (2007). When individual behaviour matters: homogeneous and network models in epidemiology. *Journal of the Royal Society, Interface / the Royal Society*, *4*(16), 879–91. doi:10.1098/rsif.2007.1100

Batagelj V and Mrvar A (2011). Pajek [Computer software]. Available from <http://pajek.imfm.si/doku.php?id=download>.

Becker, K. M., Glass, G. E., Brathwaite, W., & Zenilman, J. M. (1998). Geographic epidemiology of gonorrhea in Baltimore, Maryland, using a geographic information system. *American Journal of Epidemiology*, *147*(7), 709–16. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9554611

Brandes, Ulrik (2001). ["A faster algorithm for betweenness centrality"](http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.11.2024) (PDF). *Journal of Mathematical Sociology* **25**: 163–177. Retrieved 24.04.2014.

Brouwer, K. C., Rusch, M. L., Weeks, J. R., Lozada, R., Magis-rodríguez, C., & Strathdee, S. A. (2013). NIH Public Access, *102*(5), 1190–1199. doi:10.1080/00045608.2012.674896.Spatial

Centers for Disease Control and Prevention (2013). CDC Fact Sheet: Incidence, Prevalence, and Cost of Sexually Transmitted Infections in the United States. <http://www.cdc.gov/std/stats/STI-Estimates-Fact-Sheet-Feb-2013.pdf>

Cummins, S., Curtis, S., Diez-Roux, A. V, & Macintyre, S. (2007). Understanding and representing “place” in health research: a relational approach. *Social Science & Medicine (1982)*, *65*(9), 1825–38. doi:10.1016/j.socscimed.2007.05.036

De, P., Singh, A. E., Wong, T., Yacoub, W., & Jolly, A. M. (2004). Sexual network analysis of a gonorrhoea outbreak. *Sexually Transmitted Infections*, *80*(4), 280–5. doi:10.1136/sti.2003.007187

De, P., Cox, J., Boivin, J.-F., Platt, R. W., & Jolly, A. M. (2007). The importance of social networks in their association to drug equipment sharing among injection drug users: a review. *Addiction (Abingdon, England)*, *102*(11), 1730–9. doi:10.1111/j.1360-0443.2007.01936.x

Gerber, M. S. (2014). Predicting crime using Twitter and kernel density estimation. *Decision Support Systems*, *61*, 115–125. doi:10.1016/j.dss.2014.02.003

Heckathorn, D. D. (1997). Respondent-Driven Sampling : A New Approach to the Study of Hidden Populations \*. *Social Problems*, *44*(2), 174–199.

Jolly AM and Wylie JL (2013). Sexual Networks and Sexually Transmitted Infections; “The Strength of Weak (Long Distance) Ties”. *The New Public Health and STD/HIV Prevention: Personal, Public and Health Systems Approaches*, 77-109.

Jolly, A.M. (2012) Email regarding “Proposed geographic analysis of risk interactions of marginalised people in Ottawa and Winnipeg”.

Joyner, T. a. (2010). Kernel Density Estimation of Tropical Cyclone Frequencies in the North Atlantic Basin. *International Journal of Geosciences*, *01*(03), 121–129. doi:10.4236/ijg.2010.13016

Manitoba Centre for Health Policy. (2006). Winnipeg Area Definitions: Neighbourhood Clusters. <http://mchp-appserv.cpe.umanitoba.ca/viewConcept.php?conceptID=1311>

Martinez, A. N., Lorvick, J., & Kral, A. H. (2013). Activity spaces among injection drug users in San Francisco. *The International Journal on Drug Policy*. doi:10.1016/j.drugpo.2013.11.008

Nakaya, T., & Yano, K. (2010). Visualising Crime Clusters in a Space-time Cube: An Exploratory Data-analysis Approach Using Space-time Kernel Density Estimation and Scan Statistics. *Transactions in GIS*, *14*(3), 223–239. doi:10.1111/j.1467-9671.2010.01194.x

Pfeffer J (2004). Createpajek [computer software]. Available from <http://vlado.fmf.uni-lj.si/pub/networks/pajek/howto/excel2Pajek.htm>.

Potterat, J. J., Rothenberg, R. B., Woodhouse, D. E., Muth, J. B., Pratts, C. I., & Fogle, J. S. (1984). Gonorrhea as a Social Disease. *Sexually Transmitted Diseases*, *12*(1), 25 – 32.

Public Health Agency of Canada. (2010). Canadian Guidelines on Sexually Transmitted Infections. *PHAC website*. Retrieved April 19, 2013, from [http://www.phac-aspc.gc.ca/std-mts/sti-its/cgsti-ldcits/index-eng.php](http://www.phac-aspc.gc.ca/std-mts/sti-its/cgsti-ldcits/index-eng.php#toc)

Public Health Agency of Canada. (2011). Summary: Estimates of HIV Prevalence and Incidence in Canada, 2011. *PHAC website*. Retrieved May 02, 2013, from [www.phac-aspc.gc.ca/aids-sida/publication/survreport/estimat2011-eng.php](http://www.phac-aspc.gc.ca/aids-sida/publication/survreport/estimat2011-eng.php)

Public Health Agency of Canada. (2012). HIV/AIDS: M-Track Survey: 2012. *PHAC website*. Retrieved April 15, 2013, from [www.phac-aspc.gc.ca/aids-sida/about/mtrack-eng.php](http://www.phac-aspc.gc.ca/aids-sida/about/mtrack-eng.php)

Remis, R. S. (2007). *MODELLING THE INCIDENCE AND PREVALENCE OF HEPATITIS C INFECTION AND ITS SEQUELAE IN CANADA , 2007* (p. 52). Retrieved from [www.phac-aspc.gc.ca/sti-its-surv-epi/model/pdf/model07-eng.pdf](http://www.phac-aspc.gc.ca/sti-its-surv-epi/model/pdf/model07-eng.pdf)

Rothenberg, R. B. (1983). The geography of gonorrhea. *American Journal of Epidemiology*, *117*(6), 688–694.

Rothenberg, R., Muth, S. Q., Malone, S., Potterat, J. J., & Woodhouse, D. E. (2005). Social and Geographic Distance in HIV Risk. *Sexually Transmitted Diseases*, *32*(8), 506–512. doi:10.1097/01.olq.0000161191.12026.ca

Rudolph, A. E., Crawford, N. D., Latkin, C., Fowler, J. H., & Fuller, C. M. (2013). Individual and neighborhood correlates of membership in drug using networks with a higher prevalence of HIV in New York City (2006-2009). *Annals of Epidemiology*, *23*(5), 267–74. doi:10.1016/j.annepidem.2013.02.006

Shane, A. (2013). *Defining intervention location from social network geographic data of people who inject drugs in Winnipeg, Canada*. University of Ottawa. Retrieved from https://www.ruor.uottawa.ca/en/bitstream/handle/10393/24394/Shane\_Amanda\_2013\_thesis.pdf?sequence=1

World Health Organization (2011). Sexually Transmitted Infections (Fact Sheet No. 110). <http://www.who.int/mediacentre/factsheets/fs110/en/index.html>

Wylie, J. L., Cabral, T., & Jolly, A. M. (2005). Identification of Networks of Sexually Transmitted Infection: A Molecular, Geographic, and Social Network Analysis. *Journal of Infectious Diseases*, *191*(6), 899–906.

Wylie, J. L., & Jolly, A. M. (2013). Understanding recruitment: outcomes associated with alternate methods for seed selection in respondent driven sampling. *Biomed Central Medical Research Methodology*, *13*(93). doi:10.1186/1471-2288-13-93