

# Assessing Respondent Driven Sampling for Network Studies in Ethnographic Contexts

Kirk Dombrowski<sup>1\*</sup>, Bilal Khan<sup>1</sup>, Joshua Moses<sup>2</sup>, Emily Channell<sup>3</sup>, Evan Misshula<sup>3</sup>

<sup>1</sup>Social Network Research Group, John Jay College, City University of New York, New York, USA

<sup>2</sup>Culture and Mental Health Research Unit, Jewish General Hospital and McGill University, Montreal, Canada

<sup>3</sup>City University of New York Graduate Center, New York, USA

Email: \*kdombrowski@jjay.cuny.edu

Received December 19<sup>th</sup>, 2012; revised January 21<sup>st</sup>, 2013; accepted January 29<sup>th</sup>, 2013

Respondent Driven Sampling (RDS) is generally considered a methodology for recruiting “hard-to-reach” populations for social science research. More recently, Wejnert has argued that RDS analysis can be used for general social network analysis as well (where he labels it, RDS-SN). In this article, we assess the value of Wejnert’s RDS-SN for use in more traditional ethnographic contexts. We employed RDS as part of a larger social network research project to recruit  $n = 330$  community residents (over 17 years of age) in Nain, a predominantly (92%) aboriginal community in northern Labrador, Canada, for social network interviews about food sharing, housing, public health, and community traditions. The peer referral chains resulted in a sample that was then analyzed for its representativeness by two means—a comparison with the Statistics Canada 2006 Census of the same community, and with house-by-house demographic surveys carried out in the community as part of our research. The results show a close fit with available community statistics and our own survey. As such, we argue that the RDS sampling used in Nain was able to provide a useful and near-representative sample of the community. To demonstrate the usefulness of the results, the referral chains are also analyzed here for patterns in intragroup and intergroup relationships that were apparent only in the aggregate.

**Keywords:** Respondent Driven Sampling; Labrador Inuit; Ethnographic Methods; Network Sampling; Arctic Social Science

## Introduction

Respondent Driven Sampling (RDS) is generally considered a methodology for recruiting “hard-to-reach” populations for social research. It was pioneered in the mid-1990s by sociologist Douglas Heckathorn (1997) and modified and extended in the decade since (Heckathorn & Jeffri, 2001; Heckathorn, 2002a, 2002b; Heckathorn et al., 2002; Salganik & Heckathorn, 2004; Salganik, 2006; Heckathorn, 2007; see Wejnert & Heckathorn, 2010). More recently, Wejnert has argued that RDS can be used for general social network analysis as well (where he labels it, RDS-SN; see Wejnert, 2010). In this paper, we assess the value of Wejnert’s RDS-SN for use in more traditional ethnographic contexts.

As part of a larger social network research project, we used Respondent Driven Sampling (RDS; Heckathorn, 2002a) to recruit  $n = 330$  community residents (over 17 years of age) in Nain, a predominantly (92%) aboriginal community in northern Labrador, Canada, for social network interviews about food sharing, housing, public health, and community traditions. Per protocol, the RDS system employed in Nain involved the use of numbered referral coupons, which in turn allowed researchers to track referral chains. As described below, the peer referral chains resulted in a sample that was then analyzed for its representativeness by two means—a comparison with the Statistics Canada 2006 census of the same community, and with house-by-house demographic surveys carried out in the community as part of our research. To demonstrate the usefulness of the re-

sults, the referral chains are also analyzed here for patterns in intragroup and intergroup relationships that were apparent only in the aggregate.

While far from foolproof, RDS referrals have been shown to reliably recruit broad samples of otherwise hard-to-reach populations and, given sufficient referral depth and adequate sample-size, to achieve sampling equilibrium and independence from referral starting points (Heckathorn, 2007; Salganik & Heckathorn, 2004). Questions about the ability of RDS methods to produce anticipated results have recently been raised in formal terms by Gile and Handcock (2010), and Goel and Salganik (2009, 2010). As shown below, however, the sample of respondents recruited in Nain conformed closely to Statistics Canada’s published results for the community, including proportional distributions of ages, genders, ethnic identities, education levels, and employment statistics. Given this close fit with known community statistics, we argue that the RDS sample recruited in Nain was able to overcome these challenges, and provide a useful and near-representative sample of the community.

In employing RDS in this way, we join Wejnert (2010) in arguing that RDS methodologies provide a basis for network analysis and community description, as well as for recruitment and bias estimation. At another level, the current approach differs from past uses of broad survey forms in the Arctic by examining social relationships both in the aggregate and directly, and thereby forming conclusions about internal social networks based on relational data rather than statistical inference drawn from individual characteristics (as when, for example, data

\*Corresponding author.

trends in one individual variable, say age, are correlated with another individual variable, say income, to argue for the presence of age discrimination). The method presented here—of directly observing social relationships—has been present in Arctic small-scale ethnography for many years, but is seldom undertaken systematically or on such a large scale as made possible by RDS methods.

### The Nain Networks Project

The data discussed below were collected as part of a larger project aimed at understanding the informal networks that residents of Nain use to access housing, food, health related counseling, traditional knowledge, and other factors (Dombrowski et al., 2012). This research took place in Nain, Labrador, from January through June, 2010. Nain is a predominantly Inuit community, and the capital of the newly formed indigenous autonomous area of Nunatsiavut. The community was formed by Moravian missionaries in the late 18th century, and is currently composed of approximately 1200 people, roughly 60% of whom are age 18 or over.

Because social networks were a main focus of the overall study—and also because of high levels of residential mobility and lack of phone service made other forms of random sampling difficult or impossible—incentivized peer-to-peer recruitment was thought the most reliable way to obtain a large sample of respondents for one-on-one interviews (all of which were conducted by project ethnographers Dombrowski and Moses). Each interview lasted roughly one hour, and in total, approximately 16,000 network connections were documented among the 773 adult residents and formal social institutions of Nain<sup>1</sup> on issues of food sharing, housing assistance, domestic violence assistance, hunting/fishing partnerships, kinship, jobs assistance, sources of traditional knowledge, and youth support. The researchers spent a little more than 5.5 continuous months in the community conducting interviews and performing ethnographic observation.

The network connections at the center of this project were obtained by asking for the names of various exchange partners and those who the respondent saw as potential sources of personal help. These dyadic data were then coded and amalgamated into a series of specific networks. The project was carried out with the approval and advice of the Nunatsiavut Research Committee and informed consent of all participants. The specific networks of interest to the project were the result of prior discussions with Nain community members, including three focus groups convened to discuss social issues of interest to the community. All of the network data on the specific network topics above were derived independently of RDS recruitment, and will be analyzed separately. To be clear, the analysis that follows is based solely on the social connections revealed by the recruitment process, and the demographic data collected during the intake interview.<sup>2</sup>

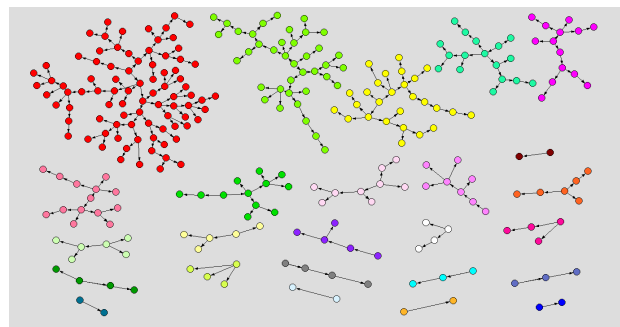
<sup>1</sup>In all, we documented 749 permanent adult residents in Nain. The remaining 24 named alters in the network interviews were either general social institutions in Nain (the hospital, the women's shelter, the Royal Canadian Mounted Police) or were individuals who resided outside the community.

<sup>2</sup>This is largely because all of the other networks were designed to discover specific aspects of social interaction. The recruitment network was the only network data collected that did not specify the exchange of a particular item, idea, or relationship. In the RDS portion of the interview participants were free to recruit from their entire social circle. And thus, while homophily and affiliation patterns are certainly found in the other networks, the analysis here is intended to serve as a baseline against which the patterns of connection in other networks can be gaged.

Starting with a small number of seeds, RDS methods begin by providing payment for a research interview (in our case, \$30 for a ~1 hour interview), from which both individual and social network data are obtained. Following this, the researcher offers recruiting bonuses for participants who bring other qualified participants into the project. To do this, each participant is given three individually numbered coupons that they may distribute to those in their personal social network who fit the project profile. For each new recruit who comes in to the research project with a recruitment coupon, the recruiter who gave him/her that coupon receives the additional recruitment bonus (in our case \$10). In turn, each new recruit is paid for participating in the research interview, and then given three recruiting coupons of his/her own. All recruitments are tracked, and steps are taken to ensure that duplicate participation is prevented (in our case the names of respondents were tracked and over the course of 5+ months, the researchers came to know almost all adult residents of the community by name). In general the recruitment payment is intentionally set low enough to avoid encouraging coercion on the part of recruiters, yet its presence provides the recruiter with an incentive to choose among his/her associates those individuals with whom he/she has enough influence such that the receiver will participate in the study, and the recruiter can be assured of getting the recruitment fee.

In Nain, we sought to recruit any adult resident of the community over the age of 17, and as per our agreement with the Nunatsiavut Research Committee, ongoing open enrollment took place during the entire 6 months of the project. In all over 800 referral coupons were distributed to respondents, and 288 total recruitments (out of 330 total interviews) took place (see **Figure 1**). The remaining 42 interviews were either initial seeds (16) or “walk-ins” (26, i.e. individuals who heard about the project indirectly and scheduled an interview without a recruitment coupon after the first month of research). The instructions for referral coupon distribution were intended to allow maximum enrollment/inclusion of community adults. Respondents were told that they would receive the recruitment bonus if they gave their coupons to anyone 18 years old or older who currently lived in the community and whom they knew by name, if/when that person completed the research interview. In Nain, the small, close-knit nature of the community meant that, without exaggeration, every participant knew all other potential participants by name, and thus the potential number of respondents for all participants was, in theory, equal to the field of all possible respondents.

Recruitment did not happen at random, however. Nearly 80%



**Figure 1.**  
RDS sampling trees.

(227/288 total recruitments) of those recruited via referral coupon indicated that their recruiter was either a “close relative”, “distant relative”, “close friend”, or “friend”.

Given that all recruiters had, in theory, access to all eligible respondents, researchers sought a better way to gage the size of the “likely” pool of possible recruits for each respondent. To do this, the intake interview asked each participant to give the approximate number of “people you are close to here in Nain—people who you trust and feel you can count on for help.” This, we felt, was a better indicator of the field of possible recruitments than the standard RDS protocol of individuals simply known by name. This latter figure (people you can count on) ranged from two to 150 (see **Figure 2**), with an arithmetic mean of 17.5, and a standard deviation of 15.7, and was taken to be an indication of each person’s self-estimated network size (i.e. degree, for purposes of RDS analysis, see Heckathorn 2002a).

To test that the sampling procedure remained accurate to the actual demographic distribution of the community, we performed two separate assessments of the RDS sample. In the first, the RDS-generated estimates of population proportions derived from the Nain recruitment data are compared with Statistics Canada census sources for Nain (see **Table 1**). Because the analysis that follows tests the for social boundaries across 7 variables—age, gender, ethnic identity, place of birth, relocation factor, education level, and work status—those variables make up the basis of the comparison that follows. A subsequent point of comparison, income, is given in only aggregate form in the Census data, and as such no comparison of income distributions between the sample and the Census is possible.

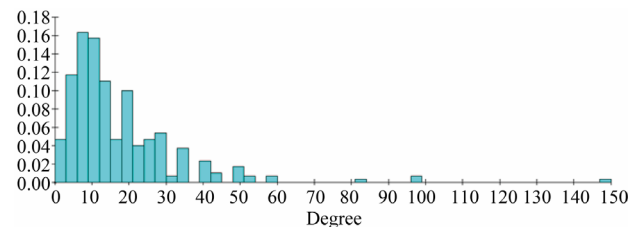
As a second test, we performed three independent house-by-house surveys of the community using data from three key respondents. In a small indigenous community where all individuals are well known to each other, information on residents, their employment status and approximate income, and their list of dependents can be drawn from nearly any adult on nearly any other. By obtaining multiple estimates of each household, we were able to aggregate the three surveys, deal with conflicting accounts, and come to an estimate based on eventual con-

sensus from the three key respondents.

### Assessment 1—A Comparison with the Census Canada Data

A key feature of RDS methodology is its ability to use information from the recruitment process to estimate the errors and sources of bias found in most peer-recruitment/snowball sampling techniques. Key among these are potential biases caused by the over-recruit of groups with systematically higher network degrees (i.e. higher number of social contacts). Where sub-groups of the larger population demonstrate higher average degree, over-recruitment will occur simply from the fact that individuals in this group are more likely to know someone participating in the study, and are thus more likely to be recruited to the study, due simply to their greater number of contacts. Thus in a situation where, say, women have a higher average number of social contacts than men, peer referral would be likely to recruit more women than men—even if all recruiters chose randomly from their list of contacts and the number of men and women in the community were equal.

A second key problem associated with peer referral recruitment is the issue of homophily within a subgroup of the population. Homophily is the tendency of a respondent to recruit someone “like” him/herself according to any number of important local distinctions. Extending the example above: where men tend to recruit other men disproportionately from a what



**Figure 2.**  
Degree distribution.

**Table 1.**

A comparison of RDS population distribution estimates with the 2006 Statistics Canada Census. The upper and lower bounds represent estimates at  $p = 0.05$  confidence, based on analysis by RDSAT 6.0.2 (2007).

CENSUS CATEGORY	RDS Comparisons					
	Sub-category	Census Value	Census Percentage of Population	RDS Estimated Percentage of Population	RDS Upper Bound	RDS Lower Bound
Age characteristics	20 to 29 years	155	0.26	0.282	0.225	0.349
Age characteristics	30 to 39 years	145	0.24	0.228	0.174	0.294
Age characteristics	40 to 49 years	135	0.22	0.191	0.150	0.245
Age characteristics	50 to 59 years	105	0.17	0.084	0.050	0.12
Age characteristics	60 and Over	65	0.10	0.098	0.038	0.115
Gender	Men over 20	325	0.52	0.52	0.458	0.586
Gender	Women over 20	300	0.48	0.48	0.414	0.543
Aboriginal population	Aboriginal identity	950	0.92	0.916	0.863	0.953
Aboriginal population	Non-Aboriginal identity	85	0.08	0.084	0.04	0.093
Educational attainment	No HS certificate	395	0.58	0.744	0.662	0.792
Labour force activity	Employed	310	0.41	0.509	0.442	0.586

we might expect in a random sample of their social contacts, we would characterize this tendency as homophily and note its potential to bias the sample toward men, especially if the same bias to recruit within group was not present among women. Such questions make starting conditions particularly important, as once the sample (by chance) tilts toward a group that shows high or even moderate levels of homophily, the remaining recruitments will be affected by this, and skew the sample toward the homophilous group.

While such issues were once thought to be significant obstacles to the use of peer-driven sampling for the recruitment of representative populations, RDS theorists have proposed means by which these biases and their effects can be estimated, and weighting factors developed to remove such biases from the sample (Wejnert et al., 2008; Wejnert, 2010). By tracking the recruiting trends of respondents throughout the sampling process, and by ascertaining their network degrees, these two forms of data can be used to provide bounded estimates of population proportions. To accomplish this, the effects of homophily and degree-based biasing are calculated for any sub-grouping which might be defined for a population, regardless of whether such potential division actually results in a self-conscious group as such in the community. For example, researchers may investigate the effects of locally significant categories, such as the categories of “Inuit”, “Kablunângajuk”, and “white” that residents use to distinguish quasi-ethnic-racial distinctions in Nain. Or, researchers can use the sampling data to estimate the influence of age by dividing the population into artificial age groups or “bins” and compare the interrelation of older and younger members of the community, regardless of whether actual “age groups” exist in the community. And finally, because of the formal nature of the estimation process, researchers can look at the combined effects of several categories, local and researcher-defined, in multivariate forms of estimation.

Toward this end, we note that RDS homophily estimation necessarily involves several steps to “disentangle” intra/inter-group affiliation from degree based affiliations (see Wejnert et al., 2008; Wejnert, 2010). In the RDSAT analysis package, estimates are given for Hd (degree homophily), Ha (affiliation homophily—without regard for degree, sometimes called assortative mixing), and Hx which isolates the value of Ha when Hd is accounted for. To avoid unnecessary confusion, in the latter part of this paper, where we turn to the social network analysis of the community, Hx shall be referred to simply as H as there our interest lies primarily in the social boundaries between groups. For purposes of estimation, however, it is the combination of Ha and Hd as these can be used to correct for biases in the recruitment of research subjects that are our primary concern. In **Table 1**, we used the data from our sample in the RDSAT analysis software to produce a series of bias estimations for our data, and to use the corrected sample to provide estimates of population proportions for the community as a whole. More specifically, our RDS sample was analyzed for each available category of the most recent Statistics Canada Census data from Nain, using the double corrections for recruitment biases caused by degree and affiliation homophily.

In all, the RDS estimates derived from the recruitment process are very similar to the Census figures in most areas. Underlined cells in Table 1 indicate where Census numbers do not fall within the RDS estimated boundaries (set at  $p < 0.05$  confidence-level). These include the percentage of the population between the ages 50 and 59 where our RDS estimate is lower

than the Census count. This difference likely indicates a bias in our recruitment method, with an under-sampling of residents within this age bracket. Likewise, differences in Employment Status/Labor Force Activity (where the RDS estimates of employment level are higher) may also indicate a sampling bias. But here we note that our estimates may have been influenced by events taking place after 2006. In the intervening years between the 2006 Census and our research, Nain has undergone a series of changes in economic climate which have accompanied the implementation of the Labrador Inuit Land Claims Agreement of 2006 (and subsequent building boom and increase in public employment in Nain, and the opening of the Voisey’s Bay Nickel Mine nearby). The importance of these economic changes on the success of the sampling is discussed in more detail below. For the moment, we would argue that the difference seen here between the Census level and our estimates from the RDS sample indicate a change in local employment conditions, rather than a sample that was biased towards those who are employed.

Highlighted as well is that the level of non-completion of High School estimated by the RDS (from 66% - 79%) is higher than the local Census Data for Nain (58%). This too may indicate a skewing of our sample. It is worth considering, though, that the RDS estimate conforms well to the High School completion/non-completion rates from the predominantly Inuit region of Nunavut, where High School completion rates are estimated by Statistics Canada to be 29.6% in 2006-2007 (and thus a non-completion rate of 70.4%, which falls within the RDS estimated boundaries; Statistics Canada 2006). Again, more recent and detailed Census data for Nain are necessary to determine the source of the discrepancy.

Given these considerations, we feel that the data from Table 1 would indicate that RDS produced a close-to representative sample of the population of Nain as a whole, especially on issues of aboriginal status, gender, and age. But as above, recent economic activities require greater examination, especially given the fact that researchers paid respondents for interviews and referrals. Knowing this, the question of how these payments may have skewed the resulting sample is of genuine concern.

## Assessment 2—Comparison with the Community Survey

Toward the end of the interview process, we performed three distinct house-by-house tours with three of our project advisors (all from Nain, ages 22 - 48, who had worked with us for the previous 5+ months of interviews), and asked each to provide us with the name, age, gender, employment status, and approximate income of each household resident above 17. These surveys were then checked against one another for agreement and disagreement, and a subsequent meeting was held with all three advisors to discuss the results. In places where there was disagreement, discussion most often ended in consensus, with one or another of the advisors providing more up-to-date information to the researchers and other advisors. Where differences remained, it was most often about employment status and individual income, and in these cases the latter was coded as “missing” and we recorded only the name, age and gender of the adult residents. The total amount of missing data was low (less than 2%).

We initially used these data as a check on our interview data and to assign network IDs to those individuals named in network interviews but who did not complete an interview. In the process, however we collected sufficient data on those Nain residents whom we didn't interview to perform an analysis on the representativeness of our sample in areas where differences with the 2006 Census were suspected—especially those that included household income and employment status (where the Census gives little detail). We note that this evaluation is based, obviously, on reported (i.e. second hand) data. Because our survey included estimates by our advisors of the incomes and employment status of individuals who had completed our interviews, our first task was to verify the accuracy of their reporting ability. The results were a close match between the estimates of our advisors and our interview data for those individuals who did participate in the interviews. Even on the topic of income, which is generally considered private, the use of bins allowed our advisors to accurately identify the income and employment status of our interviewees more than 94% of the time.<sup>3</sup> With this knowledge, we felt that these same surveys represented highly reliable data on those who did not participate in our interviews.

A comparison of the employment status, age, and income of our sample versus those adult residents of Nain that we did not interview is available in **Table 2**. Here we see that our interview pool was younger, slightly less employed, and of somewhat lower household income (though their personal income was roughly the same) as those we did not interview. The differences are far from stark, however. And when one examines the average age of those under 65 in both categories, the numbers are even more similar—which implies that we didn't so much oversample young people as we did under-sample the elderly. This was something we were cognizant of at the time. In response, we performed several house visits with elderly Nain residents. Usually, however, those over 65 were uninterested in the network questions we were asking in the main interviews. Instead, they told us stories about the community and its history, and we listened and learned a great deal of information that complemented the ethnographic aims of the project. But these interviews didn't produce survey data and so they weren't counted as part of the sample of 330 network interviews or included in the referral network discussed in this article.

In comparing the results of Analysis 1 and Analysis 2, the results of the latter send a mixed message. The lower average age of the sample seen in **Table 2** would seem to reinforce the conclusion that our sample contained an insufficient number of those between age 50 and 59, even after correction by the RDS estimates. However, Analysis 2 would seem to indicate that we sampled a greater proportion of unemployed or underemployed

**Table 2.**

A comparison of the RDS sample with the population not interviewed during the study. Data for the latter was obtained through three independent house-by-house estimates by local residents who had worked closely with the study team.

Comparison Category	Adults in Nain		
	Interviewed (n = 330)	Not Interviewed (n = 419)	Total (n = 749)
Average Age	36.5	41.7	39.4
Average Age (18 - 64)	34.9	37.8	36.5
Mean Working Status <sup>4</sup>	1.73	1.98	1.87
<u>Mean Income</u>			
Individual	\$227/week	\$238/week	\$233/week
Household	\$486/week	\$554/week	\$524/week

persons, while Analysis 1 seemed to indicate the opposite. Given the economic changes in the community over the last several years, we put more faith in the results of Analysis 2. Yet here the differences between the sample group and those who did not complete the interview are actually quite small. Individual income differences are not pronounced, and a more significant difference is found in the levels of household income. This may have been the result of lower income households sending in a number of un-recruited individuals (of which there were in total 26) in order to maximize household benefit from the interview fees. Normal recruiting forbid intra-household recruiting, and as such, the results in Analysis 2 may help us better understand the high numbers of “walk-ins” we experienced during the interview phase of the research. The result, however, is that we can see some evidence of bias in the sample, though we note that these biases appear to be small (our sample was on average 7% younger than the average adult, and only 2.6% below the average individual income).

In light of these findings, we feel that using RDS for sample recruitment in ethnographic contexts, at least in so far as those contexts are similar to those found in Nain, provides a strong alternative to the use of conventional sampling strategies, which are often frustrated by incomplete rolls and high personal mobility. Obviously, caution must remain, as the remaining biases represent concerns. For us, these were compensated for by the fact that RDS offers several distinct advantages over other forms of random sampling in ethnographic contexts. In the first place, after the first wave of recruitments, the researchers played no role in recruiting local residents to the study. This allowed us to respect the privacy of those who did not want to be involved. This is a critical feature for ethnographic research. The second benefit of the RDS method is the greater transparency it provides prospective interview participants. Again, after the first wave, all of those recruited to the study had the opportunity to speak privately with someone who had already been through the interview process, and who could thus vouch for the good faith and open-intentions of the researchers. And finally, the process resulted in the rapid recruitment of a large proportion of the adult members of the community, placing the researchers and their interests on familiar terms with a large

<sup>3</sup>We note as well that our RDS sample contained at least 1 adult from each of 218 households in the community, meaning that we already had estimates for two of our questions (household income and numbers/distributions of household residents) from a resident of that household. In this way, our three key respondents were, in the end, responsible only for estimates of the income, age, and employment status of house-mates of our interviewees who did not themselves participate in the interview, and for households for which no adult had participated in the interview. The Census lists 271 residences in the community, though a number of these are apartment buildings that provide housing for teachers and other seasonal residents mostly associated with the school. Our estimate of locally available housing was 254 dwellings at the time of our research. Using the latter, our survey included at least one adult from 86% of the households in Nain.

<sup>4</sup>In order to gain a better picture of the current employment situation in Nain we asked interview participants to estimate their weekly income (individual and household) and current employment status. The latter responses were categorized as unemployed (0), occasional (1), seasonal (2), part time (3), or full time (4). The surveys of the community also used these same categories.

number of adults, and members of nearly every household. All of these, we felt, significantly enhanced the ethnographic portion of the study, rather than hindered it (as formal surveys in small communities sometimes do). Given the results of the assessment above, the fact that all of these things could be accomplished even while gaining a close to representative sample of the community seems to us to support Wejnert's claim that RDS holds much potential for use outside of communities normally considered "hidden".

### Use of RDS Estimates for Understanding Social Networks

One further advantage of the RDS recruitment method is that the estimates of homophily used to correct for sampling bias can also be used to understand some of the network tendencies of the population from which the sample is derived. Inter- and intragroup tendencies to association are of particular interest to anthropology, particularly in times of social and cultural change. Where RDS can be used to successfully sample from a community, it can also provide us with the means to examine these tendencies in a form in which they can be measured for relative strength and compared for relative importance.

To do this, RDSAT measures homophily within groups on a scale from  $-1$  to  $1$  (Heckathorn, 2002), with a score of  $H = 0$  indicating no preference for in-group association;  $H = 1$  indicating the highest possible preference for in-group association (implied, for example, if all men recruited to the project in turn recruited only other men), and a score of  $H = -1$  indicating the highest possible preference to connect with those outside of the group (implied, again in a situation where all of the men recruited to the project in turn recruited only women). The same scale can also be used to measure the level of association between groups (see Wejnert, 2010), labeled here as "affiliation" (A). Thus while "homophily" tracks the tendency of a group to connect only with others in the same group, "affiliation" (A), tracks the tendency of members of one group to connect with those of a specified other group (again, at a rate higher than that predicted by a random mixing of ties within the overall population). Like homophily, affiliation is scored on a scale of  $-1$  to  $1$ , with a positive score indicating a tendency for intergroup association, and a negative score indicating intergroup disassociation (Heckathorn, 2002a).

By way of example, an affiliation score between women and men of  $A = 0.355$  would indicate a tendency of those in the former group (i.e. women) toward association with those in the other group (men). A negative score of the same absolute value ( $A = -0.355$ ), on the other hand, would indicate a tendency for disassociation by women to men. Importantly, such scores need not be symmetrical. That is, in a situation where association/disassociation levels between more than two groups is being measured, it is possible and perhaps even likely that the preference for association from group A to group C will be different than group C's preference for affiliation with group A. Such social asymmetries are an important feature in understanding the dynamics of local social boundaries discussed below.

For purposes of comparison, a homophily or affiliation score of  $H \geq 0.3$ ;  $A \geq 0.3$  will be considered here an indicator of important in-group (in the case of H) or between-group (in the case of A) preference, and a score of  $H \leq -0.3$ ;  $A \leq -0.3$  or lower would indicate an important level of in-group/between-group avoidance. Levels closer to zero may still indicate im-

portant boundaries, but to simplify the interpretation of a large number of table cells, some cutoff is necessary.

The results of analyses of the social boundaries formed by gender, employment status, education level, age, ethnic identity, place of birth, and income are presented in **Tables 3-9**.

**Table 3.**

Gender. The tendency of same and cross-gender affiliation using RDSAT 6.0.2 (2007).

Gender	Affiliation Matrix	
	Men	Women
Men	-0.063	0.063
Women	0.04	-0.04

**Table 4.**

Employment. The tendency of affiliation for same and cross-status employment using RDSAT 6.0.2 (2007).

Employment Status	Affiliation Matrix	
	Not Working	Working Full or Part Time
Not Working	0.102	-0.102
Working Full or Part Time	-0.174	0.174

**Table 5.**

Education. The tendency of affiliation for same and cross-education status using RDSAT 6.0.2 (2007).

Education Level	Affiliation Matrix	
	Did Not Finish High School	Finished High School
Did Not Finish High School	0.089	-0.089
Finished High School	-0.263	0.263

**Table 6.**

Age. The tendency of members of an age bin to affiliate with others of the same/different bins, using RDSAT 6.0.2 (2007).

Age	Affiliation Matrix				
	18 - 29	30 - 39	40 - 49	50 - 59	Over 60
18 - 29	0.079	-0.179	0.042	-0.031	-0.421
30 - 39	-0.257	0.123	0.028	0.004	-0.169
40 - 49	-0.026	0.006	0.015	0.032	-0.395
50 - 59	-0.238	-0.092	0.062	0.123	-0.493
Over 60	-0.383	-0.022	-0.213	-0.314	0.253

**Table 7.**

Place of birth. The tendency of affiliation by place of birth using RDSAT 6.0.2 (2007).

Place of Birth	Affiliation Matrix	
	Born in Nain	Not Born in Nain
Born in Nain	0.355	-0.355
Not Born in Nain	-0.295	0.295 <sup>5</sup>

<sup>5</sup>While the homophily/affiliation score of 0.295 does not meet the 0.3 criteria it is so close as to deserve note, particularly where the other group demonstrates similar levels of inclusion/exclusion.



**Table 8.**

Ethnicity. The tendency for members of ethnic groups to affiliate, using RDSAT 6.0.2 (2007).

Ethnic Self-Ascription	Affiliation Matrix		
	Inuit	Mixed <sup>6</sup>	White/Other
Inuit	0.2	-0.042	-0.517
Mixed	-0.3	0.219	0.069
White/Other	-0.315	0.335	-1.0 <sup>7</sup>

**Table 9.**

(a) Individual Income. The tendency of members of an income category to affiliate with others of the same/different category, using RDSAT 6.0.2(2007). (b) Household Income. Similarly, the tendency of individuals for a given household income categories to affiliate with others of the same/different category, using RDSAT 6.0.2 (2007).

(a)

Weekly	Affiliation Matrix			
	Under \$200	\$200 - \$300	\$300 - \$500	Over \$500
Under \$200	0.048	0.036	-0.13	-0.295
\$200 - \$300	-0.104	0.092	0.023	-0.322
\$300 - \$500	-0.284	0.051	0.154	-0.137
Over \$500	-0.2	-0.013	0.029	0.101

(b)

Weekly	Affiliation Matrix			
	Under \$300	\$300 - \$500	\$500 - \$750	Over \$750
Under \$300	0.134	0.001	-0.174	-0.265
\$300 - \$500	0.013	0.228	-0.326	-0.308
\$500 - \$750	0.011	-0.191	0.11	-0.126
Over \$750	-0.271	-0.355	-0.322	0.324

The full interpretation of these findings goes beyond the scope of the current paper. However, from these data we can draw some preliminary observations. In light of the above discussion on income/employment status and recent transformations of the local economy, we can see from **Table 9(b)** that those in the highest income bracket (by household) demonstrate higher levels of insularity ( $H \geq 0.3$ ;  $A \leq -0.3$ ) than other income brackets, and a more pronounced sense of social separation than when income is examined at the individual level. This is perhaps not surprising, but the method employed here allows us to examine the relative importance of household income in determining ones social connections in comparison to say, gender (which demonstrates little influence), education level (which shows low but non-zero tendencies to boundedness), and place of birth (which, like household income at the highest level, shows marked tendencies to social network influence).

A simple rank ordering of these various factors is not possi-

<sup>6</sup>In Nain, unlike elsewhere in Labrador, the terms “mixed” or “metis” are not used. The local Inuit term is “*Kablunāngajuk*”, an older (and often pejorative) term meaning, roughly “trying to be a white kind of person”.

<sup>7</sup>The number of self-identified white/other respondents was very low (6 of 330), and none were recruited by each other. Because of their small numbers their recruiting tendencies are subject to random factors much more than the recruiting patterns of either self-identified Inuit or Mixed/Metis.

ble, given the different levels of homophily and affiliation shown by the sub-categories within any given metric (i.e. those in lower household income brackets do not show the same pattern of inward-looking social network formation as do those of higher income). But these data do allow for some general conclusions. From the above we can see that ethnic affiliation (**Table 8**), wide differences in age (**Table 6**), place of birth (**Table 7**), and household income (**Table 9(b)**) seem the most influential factors in determining the social connections of adult residents of Nain. Gender (**Table 3**), educational attainment (**Table 5**), individual income (**Table 9(a)**), and employment status (**Table 4**) seem less influential.

Taking this one step further, we can also examine the combined influence of several of these factors, which can help us discover ways that several of these factors interact in determining the patterns of interconnection found in the community. **Table 10** shows a multivariate analysis of place of birth (in Nain or elsewhere) and household income (below/above \$500 per week). Here we can see that the influence of place of birth is mediated in important ways by household income in its influence on the social connections of community residents. Strong rates of homophily are found among higher income individuals born in Nain, and lower income individuals born outside the community. From these data, it is clear that both of these groups are more inward looking than lower income individuals born in Nain or higher income individuals born elsewhere.

Intergroup rates of affiliation are just as revealing. Reading across each row, we can see that:

- (Row 1) Individuals of lower income who are born in Nain have significant disaffiliation with individuals born outside of the community who live in higher income households. There are also indications of disaffiliation with higher income individuals born in Nain, but these do not meet our threshold level of  $>0.3$  used throughout the paper.
- (Row 2) As above, individuals born in Nain who live in households with higher income have considerable in-group homophily, and very high rates of disaffiliation with individuals not born in Nain. This is clearly a group with firm social boundaries.
- (Row 3) Individuals not born in Nain and who live in low income houses likewise show considerable in-group connections, and very high rates of disaffiliation with individuals born in Nain who live in high income houses. They show expected (by random mixing) levels of affiliation with

**Table 10.**

Bivariate PoB and household income. Using RDSAT 6.0.2 (2007).

	Affiliation Matrix			
	Born in Nain & Under \$500	Born in Nain & Over \$500	Not Born in Nain & Under \$500	Not Born in Nain & Over \$500
Born in Nain & Under \$500	0.164	-0.213	0.03	-0.351
Born in Nain & over \$500	-0.222	0.312	-0.705	-0.747
Not Born in Nain & under \$500	0.152	-0.801	0.326	-0.053
Not Born in Nain & over \$500	0.175	-0.506	0.063	0.095

individuals of lower income households who were born in Nain, and with those individuals born outside of Nain (like themselves) but who live in high income households.

- (row 4) Individuals not born in Nain who live in households with high incomes show expected levels of affiliation with all groups except those individuals born in Nain who live in high income households.

Taken together, these results point to the fact that different means for social differentiation interact in ways one might not expect when viewed individually. What emerges here is the relative social boundedness of those individuals born in Nain who live in higher income households, notable in both their inward-looking tendencies and in their high rates of disaffiliation with those outside their group. Findings like these lend considerable support to past research in Nain (Brantenberg, 1977a, 1977b) and in the nearby community of Makkovik (see Ben-Dor, 1977; Kennedy, 1977, 1982). In each of these cases, questions of ethnic divisions and the influence of economic and political factors on emerging social boundaries were noted. Forty years later, such differences still exist, though many have been redefined by the Land Claims process (see Dombrowski et al., 2012; Dombrowski et al., forthcoming). These findings also lend (local) clarity to questions about the impact of the cash economy and wage paying jobs on the social fabric of Northern communities.<sup>8</sup>

In closing, we note that the ability of RDS-SN to both capture a representative sample and reveal community organizational trends seems well borne out in Nain. It is our feeling that RDS represents a new and important tool for research in ethnographic contexts where researchers seek to understand the social dynamics of a small community while respecting individual privacy and maximizing community exposure to the research. Importantly, these techniques are capable of revealing trends that may not be apparent to local residents, or which may be known in only a general way—no one in Nain indicated to us that the combination of place of birth and household income would allow us to discover important social divisions, though many did comment informally that some of the well-off families in the community did keep to themselves, and there is a separate and widespread notion that some individuals in the community practiced a measure of gate-keeping when it came to the allocation of new jobs and sources of income. As such, the methods used here may be useful in formalizing local no-

tions and testing their relative importance, singly and in combination. We emphasize that this sort of analysis remains decidedly experimental in ethnographic analysis. We know of no other large scale RDS implemented project in a conventional ethnographic context. But the results of our own efforts to implement these procedures would indicate that researchers may wish to employ these techniques in contexts where sampling remains problematic, and research rapport with the local community remains of paramount importance. We look forward to further evaluations when this does take place.

## Acknowledgements

This project was supported by a grant from the US National Science Foundation, Office of Polar Programs, Division of Arctic Social Sciences, GR ARC 0908155, with the approval of the Nunatsiavut Research Committee. All of the material contained here was obtained with the informed consent of all participants. The analysis of the data and all conclusions and recommendations are the responsibility of the Principal Investigator/Lead Author and do not represent the opinions of either the US National Science Foundation, The Nunatsiavut Government, the Nunatsiavut Research Committee, or the City University of New York. Special thanks are due to Fran Williams, Jane Dicker, Toby Pijogge and Eva Lampe in Nain. Supporting modeling work was performed at the Social Network Research Group labs at John Jay College, [www.snrg-nyc.org](http://www.snrg-nyc.org).

## REFERENCES

- Ben-Dor, S. (1977). Inuit-settler relations in Makkovik, 1962. In R. Paine (Ed.), *The White Arctic: Anthropological essays on tutelage and ethnicity*, (pp. 306-325). St. Johns: Memorial University of Newfoundland.
- Bernier, R. (1997). The dimensions of wage inequality among aboriginal peoples. *Business and Labour Market Analysis*, 109. Ottawa.
- Brantenberg, T. (1977a). Ethnic values and ethnic recruitment in Nain. In R. Paine (Ed.), *The White Arctic: Anthropological essays on tutelage and ethnicity* (pp. 326-343). St. Johns: Memorial University of Newfoundland.
- Brantenberg, T. (1977b). Ethnic commitments and local government in Nain, 1969-76. In R. Paine (Ed.), *The White Arctic: Anthropological essays on tutelage and ethnicity* (pp. 376-410). St. Johns: Memorial University of Newfoundland.
- Dombrowski, K. (2002). The praxis of indigenism. *American Anthropologist*, 104, 1062-1073. doi:10.1525/aa.2002.104.4.1062
- Dombrowski, K. (2007). Lifestyle and livelihood: Culture politics and Alaska native subsistence. *Anthropologica: The Journal of the Canadian Anthropology Society*, 49, 211-230.
- Dombrowski, K., Khan, B., Moses, J., Channell, E., & Curtis, R. (2012). The Nain networks project. *A Research Brief to the Nunatsiavut Research Committee*. [http://snrg-nyc.org/wp-content/uploads/2012/08/Nain\\_Brief\\_final.pdf](http://snrg-nyc.org/wp-content/uploads/2012/08/Nain_Brief_final.pdf)
- Dombrowski, K., Khan, B., Moses, J., Channell, E., & Curtis, R. (forthcoming). Sampling Social Divisions in a Rural Inuit Community. *Identities*.
- Gile, K. J., & Handcock, M. S. (2010). Respondent-driven sampling: An assessment of current methodology. *Sociological Methodology*, 40, 285-327. doi:10.1111/j.1467-9531.2010.01223.x
- Goel, S., & Salganik, M. J. (2009). Respondent-driven sampling as Markov chain Monte Carlo. *Statistics in Medicine*, 28, 2202-2229. doi:10.1002/sim.3613
- Goel, S., & Salganik, M. J. (2010). Assessing respondent-driven sampling. *Proceedings of the National Academy of Sciences*, 107, 6743-6747. doi:10.1073/pnas.1000261107
- Heckathorn, D. (1997). Respondent-driven sampling: A new approach

<sup>8</sup>Bernier (1997), and more recently Maxim, Beavon, and Whitehead (2001) discuss income differentials rather than class in their analyses of aboriginal stratification. The latter defend studying income differentials within aboriginal groups (in addition to the much more commonly studied effects of income inequality between aboriginal groups and neighboring Euro-Canadians) because a high income differential within a group (or "polarization", see Bernier, 1997: p. 4; Maxim et al., 2001: p. 476) is often associated with high mortality statistics. Along these lines, Bernier notes that income differentials between the income of the highest earning 20% and lowest earning 20% of the aboriginal population in Canada were highest among the Inuit (1997: pp. 13-14). He notes, however that this comparison also showed that the lowest earners among the Inuit work substantially fewer hours a year than other Inuit. The reasons for this difference, however, are not pursued. Marybelle Mitchell (1996) is one of the few authors to attempt a class analysis among Canadian Inuit, though Prattis and Chartrand (1990) earlier compared the Inuit situation (amid resource development) to the sort of power differentials that arise under colonial regimes. Dombrowski (2002, 2007) has sought to raise such issue in the context of post-land claims indigenous politics in the North more generally, but in contrast with both Mitchell, and Prattis and Chartrand, argues that the wider class dynamics of the surrounding Euro-American society must be taken into account to understand local class dynamics.



- to the study of hidden populations. *Social Problems*, 44, 174-199. doi:10.2307/3096941
- Heckathorn, D. (2002a). Respondent-driven sampling II: Deriving valid population estimates from chain-referral samples of hidden populations. *Social Problems*, 39, 11-34. doi:10.1525/sp.2002.49.1.11
- Heckathorn, D. (2002b). Development of a theory of collective action: From the emergence of norms to AIDS prevention and the analysis of social structure. In J. Berger, & M. Zelditch (Eds.), *New directions in sociological theory: Growth of contemporary theories*. New York: Rowman and Littlefield.
- Heckathorn, D. (2007). Extensions of respondent-driven sampling: Analyzing continuous variables and controlling for differential recruitment. *Sociological Methodology*, 37, 151-208. doi:10.1111/j.1467-9531.2007.00188.x
- Heckathorn, D., & Jeffri, J. (2001). Finding the beat: Using respondent-driven sampling to study jazz musicians. *Poetics*, 28, 307-329. doi:10.1016/S0304-422X(01)80006-1
- Heckathorn, D., Salaam S., Broadhead, R. S., & Hughes, J. J. (2002). Extensions of respondent-driven sampling: A new approach to the study of injection drug users aged 18-25. *AIDS and Behavior*, 13, 55-67. doi:10.1023/A:1014528612685
- Hicks, J. (2004). On the application of theories of "internal colonialism" to Inuit societies. *Annual Conference of the Canadian Political Science Association*, Winnipeg, 3-5 June 2004.
- Kennedy, J. C. (1977). Local government and ethnic boundaries in Makkovik, 1972. In R. Paine (Ed.), *The White Arctic: Anthropological essays on tutelage and ethnicity*, (pp. 359-375). St. Johns: Memorial University of Newfoundland.
- Kennedy, J. C. (1982). Holding the line: Ethnic boundaries in a northern Labrador community. *Social and Economic Studies*, 27. St John's: Memorial University Institute of Social and Economic Research.
- Maxim, P. S., Beavon, D., & Whitehead, P. C., (2001). Dispersion and polarization of income among aboriginal and non-aboriginal Canadians. *Canadian Review of Sociology*, 389, 465-476.
- Prattis, J. I., & Chartrand, J.-P., (1990). The cultural division of labour in the Canadian north. *The Canadian Review of Sociology and Anthropology*, 27, 49-73. doi:10.1111/j.1755-618X.1990.tb00444.x
- RDSAT 6.0.1 (2007). <http://www.respondentdrivensampling.org/>
- Salganik, M. J., & Heckathorn, D. (2004). Sampling and estimation in hidden populations using respondent-driven sampling. *Sociological Methodology*, 35, 193-238. doi:10.1111/j.0081-1750.2004.00152.x
- Salganik, M. J. (2006). Confidence intervals, design effects, and sample size calculation for respondent-driven sampling. *Journal of Urban Health*, 83, 98-111. doi:10.1007/s11524-006-9106-x
- Semaan, S., Jennifer, L., & Liebman, J. (2002). Street and network sampling in evaluation studies of HIV risk-reduction interventions. *AIDS Review*, 4, 213-223.
- Statistics Canada (2006). <http://www.statcan.gc.ca/pub/81-595-m/81-595-m2009078-eng.htm>
- Tanner, A., & Henderson, S. (1992). Aboriginal land claims in the Atlantic provinces. In Coates, K. (Ed.), *Aboriginal land claims in Canada* (pp. 131-165). Toronto: Copp Clark Pitman Ltd.
- Wejnert, C., Heckathorn, D., Ramirez-Valles, J., & Diaz, R. M. (2010). The dual homophily model: Disentangling the effects of affiliation preference and differential degree on homophily. *Annual Meeting of the American Sociological Association Annual Meeting*, Sheraton Boston and the Boston Marriott Copley Place, Boston, 30 May 2011. [http://www.allacademic.com/meta/p241120\\_index.html](http://www.allacademic.com/meta/p241120_index.html)
- Wejnert, C. (2010). Social network analysis with respondent-driven sampling data: A study of racial integration on campus. *Social Networks*, 32, 112-124. doi:10.1016/j.socnet.2009.09.002