Introduction to HIV/AIDS and sexually transmitted infection surveillance

# **MODULE 4**

SUPPLEMENT A guide to using RDS Analyst and NetDraw





# Introduction to HIV/AIDS and sexually transmitted infection surveillance

# **MODULE 4**

SUPPLEMENT A guide to using RDS Analyst and NetDraw



Regional Office for the Eastern Mediterranean



#### **WHO Library Cataloguing in Publication Data**

World Health Organization. Regional Office for the Eastern Mediterranean

Introduction to HIV/AIDS and sexually transmitted infection surveillance: Module 4. Supplement: a guide to using RDS analyst and NetDraw / World Health Organization. Regional Office for the Eastern Mediterranean, Joint United Nations Programme on HIV/AIDS

p.

who-em/std/160/e

1. Acquired Immunodeficiency Syndrome – epidemiology 2. HIV Infections – epidemiology 3. Sexually Transmitted Diseases – epidemiology 4. Data Interpretation, Statistical I. Title II. Regional Office for the Eastern Mediterranean III. Joint United Nations Programme on HIV/AIDS

(NLM Classification: wc 503.4)

#### © World Health Organization 2014

#### All rights reserved.

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the World Health Organization in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by the World Health Organization to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall the World Health Organization be liable for damages arising from its use.

Publications of the World Health Organization can be obtained from Knowledge Sharing and Production, World Health Organization, Regional Office for the Eastern Mediterranean, PO Box 7608, Nasr City, Cairo 11371, Egypt (tel: +202 2670 2535, fax: +202 2670 2492; email: emrgoksp@who.int). Requests for permission to reproduce, in part or in whole, or to translate publications of WHO Regional Office for the Eastern Mediterranean – whether for sale or for noncommercial distribution – should be addressed to WHO Regional Office for the Eastern Mediterranean, at the above address: email: emrgogap@who.int.

Design by Permanent Impressions.

# Contents

L	Step-by-step guide to using RDS Analyst	5
	Introduction	7
	Installing RDS-A	8
	Console and data viewer	10
	Working with data	21
	Analysis of the sample	22
	Analysis of the population	27
	Plots to visually explore RDS data	33
	Conclusion	38
	References for the estimator table	39

2	Using NetDraw for interpreting RDS data and for problem solving	41
	Getting started with NetDraw	43
	Prepare data for NetDraw	43
	Generating recruitment chains	46

# Step-by-step guide to using RDS Analyst

# I. Step-by-step guide to using RDS Analyst

# Introduction

# Why RDS Analyst (RDS-A)?

RDS-A is a software package for the analysis of respondent-driven sampling (RDS) data. RDS-A was conceived and developed by the Hard-to-Reach Population Methods Research Group (HPMRG) as an easier and more useful alternative to other available software for the analysis of data collected through RDS methods. There are several reasons for using RDS-A, including that:

- it is free
- it is intuitive and easy to use with point and click commands
- it allows for programming in the computer package R at the batch/command line
- code is reusable and can be copied from the batch/command line
- it can analyse multiple data sets at once
- it can analyse multiple variables at once
- it easily reads multiple files (SPSS Statistics, STATA, SAS, Microsoft Excel, text in RDSAT format) without having to modify them
- it has the latest estimators
- it allows for many types of analyses including population and sample estimates, population confidence intervals, cross tabs, means, medians, percentiles, standard deviations, design effect, and proportional analyses (bi-variate with confidence intervals)
- it produces diagnostic plots
- it transforms and recodes variables
- data can be viewed and edited in a Microsoft Excel-like spreadsheet.

# **Overview of the software**

In order to provide a free and easy-to-use software alternative to proprietary data analysis software there are several layers to RDS-A.

- The base layer of RDS-A is R, a free, open-source and powerful computer package that can be downloaded at: http://www.r-project.org/. If you know R statistical language, you can make your own commands at the batch/command line.
- The next layer is JGR (Java graphical user interface [GUI]) designed for use with R (for more information see: http://rforge.net/JGR/).

- The next layer is Deducer (for more information see: http://www.deducer.org/pmwiki/ index.php?n=Main.DeducerManual). This layer provides an intuitive GUI for R which allows non-technical users to learn and perform analyses by pointing and clicking on commands.
- The final layer is the RDS-A point-and-click GUI.

# **Installing RDS-A**

# **HPMRG** web site

To download RDS-A, go to the HPMRG web site (www.hpmrg.org) (see Figure 1.1).

× Google	nrg.org/	
	Amster 🚺 Suggested Sites 🔻 🔘 Calorie Count Account H M converting kilograms and 🍘 Hidden Population Metho 🖉 Get more Add-ons 👻 🎒 Instant Virtual Extranet	
No clear coursening i	page discussion View source history	in in
STATISTICS UCLA	Hard-to-Reach Population Methods Research Group (Redirected from Main Page) The Hard-to-Reach Population Methods Research Group (HPMRG) focuses on developing statistical methodology to help improve understanding of hard-to-reach or otherwise "hidden"	
navigation = Main Page = About us Software = RDS Analyst Software = RDS Analyst RDS Analyst = RDS Analyst Workshop (UCSF, June 15-16, 2010) = Development = Recent changes = Help search	populations. These populations are characterized by the difficulty in survey sampling from them using standard probability methods. Typically, a sampling frame for the target population is not available, and its members are rare or stigmatized in the larger population so that it is prohibitively expensive to contact them through the available frames. Examples in an economic setting include unequilated workers and the self-employed. Hard-to-reach populations in the US and elsewhere are under-served by current sampling methodologies mainly due to the lack of practical alternatives to address these methodological difficulties. The Hard-to-Reach Population Methods Research Group is an collaborative interdisciplinary group of researchers from several universities: Dr. Krista J. Glie is Assistant Professor of Statistics in the Department of Mathematics and Statistics at the University of Massachusetts - Amherst. Her research focuses on developing statistical methodology for social and behavioral account concentrativen sampling. In particular, her dissertation and recent work focus on understanding the implications of acrism from several parametricular, her dissertation and recent work focus on understanding the implications of accument threes assombiolal sampling (RDS) methodology, and on introducing improved expressions of URC add and the second for the sample of RDS data. For details see her web page Ø.	AN AN ANALY ANALY
Co Search toolbox What links here Related changes Special pages Printable version Permanent link	Sampling (N2S) memodology, and of minodoculing improved estimations stategies for KoS state. For details series web page 2. Dr. Mark S. Handcock is Professor of Statistics in the Department of Statistics at the University of California – Los Angeles. His research involves methodological development, and is ba largely on molivation from questions in the social and epidemiological sciences. He has published extensively on survey sampling, network inference, and network sampling methods. He recently moved to UCLA from the University of Washington. He teaches <i>Statistical Analysis of Networks of an Sample Survey Techniques 2</i> . For details see his web page 2. Dr. Lisa G. Johnston is an epidemiologist, applied researcher and RDS consultant. Dr. Johnston has six years of experience providing supervision and training on using RDS methods, HIV/STT biological-behavioral surveillance survey planning and implementation, and the RDS Analysis Tool (RDSAT). She has provided RDS technical assistance in over 30 countries and done extensive consulting for the Center for Disease Control and Prevention (CDC) and many other institutions including Family Health International (FHI), United Nations Development Program, and UNAIDS. She is currently adjunct professor at Tulane University, School of Public Health and Tropical Medicine, and a Senior Analyst at the University of California, San Francisco, Global Health Sciences. For details see har web page Q.	

## Figure I.I Example of the HPMRG web site for downloading RDS-A

At the left hand side is a navigation bar. Click on "RDS Analyst Software". This will take you to the page shown in Figure 1.2.<sup>1</sup>

STATISTICS	page discussion edit history RDS Analyst Install	move watch
navigation = Main Page = About us = RDS Analyst Manual = RDS Analyst Install = RDS Analyst users	Contents [hide] 1 Introduction 2 Basic facts 3 Installation 3.1 Installation on an Windows PC 3.2 Installation on an Apple Macintosh 4 Getting started - The Manual	
group © RDS Analyst Workshop (Hanoi, Sept 23-25, 2013) © Recent changes © Help	Introduction <u>RDS Analyst</u> ( <b>RDS-A</b> ) is a software packag <u>RDS Analyst</u> has an easy-to-use graphical comprehensive framework for working with Example capabilities are an easy format for	user interface to the powerful and s RDS data, including tools for samp

Figure 1.2 Page for downloading RDS-A

 $<sup>\</sup>overline{}^{1}$  You may have to log in to access the software package.

#### **Download RDS Analyst**

The current development form of RDS-A is for Microsoft Windows and Apple Macintosh. Click on "Installation on a Windows PC" or "Installation on an Apple Macintosh" (see Figure 1.2) depending on your needs. Once you click on either of the installations in the contents, you can click on the installers directly to download RDS-A (see Figure 1.3).

Installation on an Windows PC	[ed
The installer is at:	
 → http://neolab.stat.ucla.edu/cranstats/RdsAnalystSetup.0.35.exe  @	
Download the install and double-click on it to install the software.	
This can install all programs and utilities needed. If you already have some elements installed you can deselect (or cancel) during the installs. It is recon you install this all the first time. This installer is over 110Mb in size and will take time to download.	nmended tha
A reboot is not required. You do not need to uninstall any components to update (This includes R and Java). However the <u>RDS Analyst</u> application or th application must not be running when you update.	ne R
You need the Java Runtime Environment to use <u>RDS Analyst</u> . You can check to see if you have java installed at http://javatester.org/version.html @ If ye message A JRE has been found. Do you want to install another one anyway?, it means that Java is already installed. In this case, click No so as to not	•
If you do not have Administrator privileges and the Java Runtime Environment is not installed then you will not be able to run RDS Analyst.	
Note for experienced users: This creates a private version of R for <u>RDS Analyst</u> to use and ensures <u>RDS Analyst</u> has the right version of R available f you already have R installed separately, the two versions will peacefully coexist and you can use the other version of R just as you were originally.	for its use. If
	for its use. If
you already have R installed separately, the two versions will peacefully coexist and you can use the other version of R just as you were originally.	for its use. If [ed
you already have R installed separately, the two versions will peacefully coexist and you can use the other version of R just as you were originally. Finally, be sure to sign up for the RDS Analyst Users Group.	
you already have R installed separately, the two versions will peacefully coexist and you can use the other version of R just as you were originally. Finally, be sure to sign up for the RDS Analyst Users Group. Installation on an Apple Macintosh	
 you already have R installed separately, the two versions will peacefully coexist and you can use the other version of R just as you were originally. Finally, be sure to sign up for the RDS Analyst Users Group. Installation on an Apple Macintosh There is a version for Apple Macintosh computers. They must have Intel CPUs (i.e., be purchased post-2006). To install:	
 you already have R installed separately, the two versions will peacefully coexist and you can use the other version of R just as you were originally. Finally, be sure to sign up for the RDS Analyst Users Group. Installation on an Apple Macintosh There is a version for Apple Macintosh computers. They must have Intel CPUs (i.e., be purchased post-2006). To install: Download and install R-3.0.1 from here: http://hpmrg.org/software/R-3.0.1.pkg @	
 you already have R installed separately, the two versions will peacefully coexist and you can use the other version of R just as you were originally. Finally, be sure to sign up for the RDS Analyst Users Group. Installation on an Apple Macintosh There is a version for Apple Macintosh computers. They must have Intel CPUs (i.e., be purchased post-2006). To install: Download and install R-3.0.1 from here: http://hpmrg.org/software/R-3.0.1.pkg @ Download the <u>RDS Analyst</u> Installer: http://hpmrg.org/software/RDSAnalystInstaller.0.2.dmg @	[ed

#### Figure I.3 RDS-A installation

Below are detailed descriptions of downloading RDS-A for Microsoft Windows or Apple Macintosh.

#### **RDS-A for Microsoft Windows**

The Windows installer is at: http://hpmrg.org/software/RDsAnalystSetup.0.1.exe

Download the installer and double-click on it to install the software. To ensure your installation is the latest version of the packages, use the updater at:

http://hpmrg.org/software/RDsAnalystUpdater.0.1.exe

and for the latest and secure version of Java:

http://hpmrg.org/software/jre-7ull-windows-i586.exe

This will install all programs and utilities needed. If you already have some elements installed you can deselect (or cancel) during the installation. It is recommended that you install the entire package the first time you install RDS-A.<sup>1</sup> This installer is over 110Mb in size and will take time to download. A reboot is not required. You do not need to uninstall any components in order to update (this includes R and Java). However, the RDS-A application or the R application must not be running when you update.

You need the Java Runtime Environment to use RDS-A. If you get the messages "a JRE has

<sup>&</sup>lt;sup>1</sup> Note for experienced users: This creates a private version of R for RDS Analyst to use and ensures RDS-A has the correct version of R available for its use. If you already have R installed separately, the two versions will peacefully coexist and you can still use the other version of R.

been found" and "Do you want to install another one anyway?", it means that Java is already installed. In this case, click "No" if you do not want to reinstall it. After you install, you should ensure your installation is the latest version of the packages by downloading the updater: http://hpmrg.org/software/RDSAnalystUpdater.0.1.exe.

#### **RDS-A for Apple Macintosh**

For Apple Macintosh users, it is necessary to have Intel central processing units purchased after 2006. Download and install R-2.15.2 from: http://hpmrg.org/software/R-2.15.2.pkg. Then download the RDS-A installer at: http://hpmrg.org/software/RDSAnalystInstaller.0.1.dmg. It should load as a disk-image. Double-click on the installer in it (i.e. "RDSAnalystInstaller") to install the software. A reboot is not required. You do not need to uninstall any components to update (this includes R and Java). However, the RDS-A application or the R application must not be running when you update. RDS-A uses Java to work. <sup>1</sup> The RDS-A application and R will be in your applications folder. To run RDS-A, double-click on it in the applications folder.

After you install, use the updater to ensure your installation is the latest version of the packages. Download the RDS-A updater at: http://hpmrg.org/software/RDSAnalystUpdater.0.1.dmg. It should load as a disk-image. Double-click on the installer in it (i.e. "RDSAnalystUpdater") to install the software. This just installs the core packages (that is, anything that has changed since the full installation was made). It will typically be a few Mb in size.

#### Desk icon

Once you download RDS-A, you will have a desk icon that looks like Figure 1.4:



#### Figure 1.4 Desk icon

Click this icon to launch RDS-A. Alternatively, if you did not install this desktop icon, you will need to go to programs to launch RDS-A.

# **Console and data viewer**

#### Introduction

RDS-A has two main windows, the "Console" (Figure 1.5) and the "Data viewer" (Figure 1.6), which should open when you download the program and/or when you launch the program using either the desktop icon or from the programs folder.

<sup>&</sup>lt;sup>1</sup> To use RDS-A, you need Java installed on your Apple Macintosh. If you are using Mac OS X 10.6 and below, Java comes pre-installed. If you are using Mac OS X 10.7 (Lion) and above, then Java is not pre-installed. You can check to see if you have Java installed at http://javatester.org/version.html.To install Java version 6, open the "Java Preferences.app" located in the Applications > Utilities folder on your Apple Macintosh. It will ask if you want to install Java if it is not already there. Accept its invitation.

# Console

The primary window is the Console. The menus at the top allow you to open, save and edit files and to do statistical analysis, plots and diagnostics. Each of the menu items are discussed in more detail below. The panel along the left side is a navigation panel. You can use it to quickly go from one statistical output to another. You can also remove output that you do not want by selecting it in the navigation panel and pressing the "Remove" button at the bottom of the navigation panel. The panel along the bottom is a command console. You can type in commands in the R language here.

The panel in the upper right, which takes up most of the screen, is an output window. By default, RDS-A commands that are submitted for execution show up as red text and the generated output shows up as black text. There are two tabs at the top of the output panel. The "Console View" shows all the output, like one long ream of paper. The "Element View" shows you just a single output element at a time, without any red RDS-A commands. You can click between these to get a view of the history of the commands and output. The Console also records a log of all the commands and output produced from them.

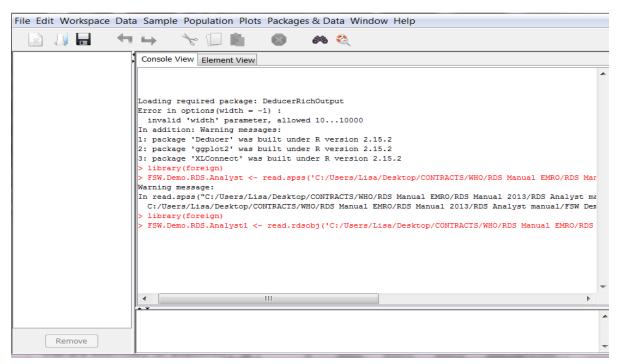


Figure 1.5 RDS-A Console

#### Learning the tabs on the Console

The Console has ten tabs: File, Edit, Workspace, Data, Sample, Population, Plots, Packages and Data, Window and Help. The tab features identified in bold in Table 1.1 are featured in more detail in this guide. Many of the tab features are intuitive and do not need further instruction.

TAB NAME	FEATURES UNDER TAB	MAIN PURPOSE OF TAB
File	<b>Open Data</b> , <b>Save Data</b> , New Document, Open Document, <b>Save</b> , Set Working Directory, Quit	Open and save data sets, set default location for data and save output using the Set Working Directory.
Edit	Undo, Redo, Cut, Copy, Copy Special, Paste, Delete, Select All, Find, Find Next, Clear Console, Increase or Decrease Font Size, Preferences	Edit features. "Preferences" allows you to change defaults such as output styles.
Workspace	Open, Save, Save As, Clear All	Objects created in RDS-A are held in computer memory as a workspace and are not saved on disk unless told to do so. To save objects when RDS-A is closed, save your workspace when exiting. Save and open (previously saved) workspaces from this menu. For multiple projects, save entire workspace for each project in a separate file. Then open them from this menu. "Clear All" empties workspace (i.e. removes all objects).
Data	Edit Factor, <b>Recode</b> Variables, Transform, Compute Weights, Reset Row Names, Sort, Edit Meta Data, Convert to RDS (Coupon Format, Recruiter ID Format)	Edit Factor: Adds or subtracts values of a categorical variable. Recode Variables: Recodes variables into existing or new variables. Transform: Makes complicated changes to variables (see transformation panel). Edit Meta Data: Specifies characteristics of RDS data (see Figure 1.11). Convert to RDS: Direct and manual way to (re)convert a data set into an RDS data set. It is rarely used (as the program does this automatically), but is useful if the automatic method missed something.
Sample	Frequencies, Descriptives, Contingency Tables, Recruitment Homophily	Unadjusted exploratory analysis of continuous, categorical and descriptive data. Frequencies: Tables of one or more variables, possibly stratified by others. Descriptives: Unadjusted summary statistics. Contingency Tables or Crosstabs: Includes test statistics. Recruitment Homophily: Statistic of differential recruitment in the sample.

# Table I.I RDS-A Console

TAB NAME	FEATURES UNDER TAB	MAIN PURPOSE OF TAB
Population	Frequency Estimates, Descriptive Estimates,	Adjusted statistics of continuous, categorical and descriptive data to describe the population sampled.
	Population Crosstabs, Test Difference in Proportions,	Frequency Estimates: Tables of one or more variables, possibly stratified by others.
	<b>Population Homophily,</b> Differential Activity	Descriptive Estimates: Adjusted summary statistics.
		Population Crosstabs: Cross tabulations of categorical variables.
		Population Test Differences in Proportions: Tests hypothesis that two population proportions are equal.
		Population Homophily: Statistic of differential recruitment in the population.
		Differential Activity: Ratio of the mean network size for those with the outcome to the mean network size of those without it.
Plots	Plot Recruitment Tree,	Plots to visually explore RDS data.
	<b>Diagnostics,</b> Plot Builder, Import Template, Open Plot, Quick, Interactive	Plot Recruitment Tree: Graphics plot of the recruitment tree.
		Recruitment Diagnostics: Bar plot of number of recruits by wave, scatter plot of network size verses wave, bar chart of number of recruits from each seed, histogram of number of recruits for each respondent, boxplots by wave, seed, etc.
		Plot Builder: Interface to create simple and sophisticated plots, including pie charts, histograms, bar plots, scatter plots, bubble plots, etc. The Plot Builder, allows the import of an existing template (Import Template), opening an existing plot (Open Plot), or working interactively with existing plot primary forms (Quick and Interactive).
Packages & Data	Object Browser, Data Viewer, GUI Add-ons, Package Manager, Package Installer, Example: Faux, Example: FauxMadrona, Example: Fauxsycamore	To re-open the Console and Data Viewer, install additional packages for underlying R engine and look at packages currently loaded, edit and view any "objects" in workspace, such as RDS data sets (spreadsheets with the prefix: rds), spreadsheets (data frames), functions, etc. There are three example data sets.
Window	Close Window, Close Same Type, Close All, Console, Data Viewer	Use this to go between the Console and Data Viewer windows, to choose a graphics window. Also, lists currently open windows to choose from and to choose a window to bring to the front to work on.
Help	R Help, About, Deducer Help, RDS Analyst Introduction Manual, RDS Analyst Reference Manual, Citation Information	Help for R, Deducer and RDS-A, and suggested citation for publications.

# Data viewer

The other window is titled "Data Viewer". This is where data are displayed and can be edited. The Data Viewer provides an easy-to-use, spreadsheet-like environment to view and edit data. Copy and pasting is supported, and is compatible with Microsoft Excel 2003/2007, so data can be moved from Microsoft Excel by simply copying it to the Data Viewer. Contextual menus can also be used to insert, delete and copy rows and columns.

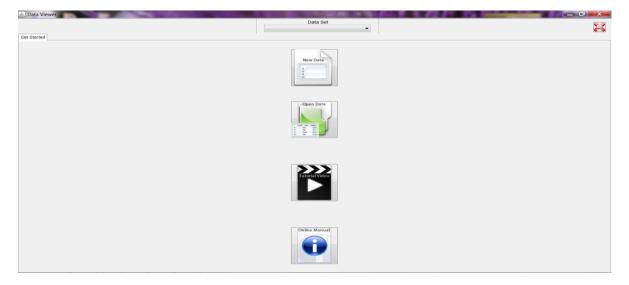


Figure I.6 RDS-A Data Viewer

# **Open data**

There are two ways to open data:

I) Go to the Data Viewer and select the "Open Data" icon (see Figure 1.7)



# Figure 1.7 Open Data icon

2) Go to the top left-hand corner of the Console, go to File |Open Data| (Figure 1.8) and select the file you want to load.

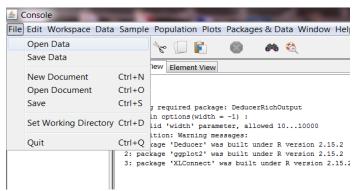


Figure I.8 Open Data from Console

Once you select the file you want to use, click on it. The example below (Figure 1.9) is loading a SPSS Statistics file titled FSW Demo RDS Analyst.

Organize 👻 🧐 Open 🔹 Share with 👻 Burn New folder										
☆ Favorites	Name	Date modified	Туре	Size						
🗼 Downloads	闦 RDS Analyst step by step	4/4/2013 10:44 AM	Microsoft Word Document	2,357 KB						
👢 Dropbox	🕞 FSW Demo RDS Analyst	3/27/2013 1:10 PM	SPSS Statistics Data Document	118 KB						
😹 Recent Places	🕙 FSW Demo RDS Analyst	3/27/2013 9:30 AM	Microsoft Excel Worksheet	247 KB						
📃 Desktop										
Documents										

Figure 1.9 Selecting a file to open

## Loading data

Once you click on your file, the "Load RDS Data" window will open (Figure 1.10). In this window the variables are located in the panel on the left-hand side. Your data should represent valid RDS data which includes a row for each respondent (i.e. case), and columns for each survey response variable. Most data will be in Data Format|Coupon. Required columns of data include unique IDs and network size data for each participant. In addition, the file should have coupon numbers formatted such that there is one column for the participant's coupon number and additional columns for each of the coupon numbers connected to that participant.

oad RDS Data	-	1000	100.2	×
	Data Format Coupon Recuiter II			
Variables net	A Recurter II		ject ID	
coupon c1 c2 c3		Net	work Size	
chestionar City F8 F9		Sub	ject's Coupon	
PS1 PS2 PS7 PS9_3	-	Cou	pons	
Optional Max # of Coupon	s:			
Population Size Esimate	Low	Mid	High	
	Run		Reset	Cancel

Figure 1.10 Load RDS Data window

The following pieces of information<sup>1</sup> are entered so that RDS-A can properly read the data file:

- I. "Subject ID": These are unique ID numbers which can be serial numbers starting from I to your final sample size or some other unique ID number (do not use the coupon numbers as your unique ID number). In the example below (Figure 1.11), the unique ID number is named "chestionar", which is each participant's unique questionnaire number and is the 6th variable in the data set (see Figure 1.10).
- 2. "Network Size": The number of people each participant reported as being part of their personal social network. This variable is named "net" (Figure 1.11.) and was the first variable in the data set (see Figure 1.10).
- 3. "Subject's Coupon": This is the coupon that each participant used to enrol in the survey. This variable is named "coupon" in the data set (Figure 1.11).
- 4. "Coupons": These are the coupons given to each participant for them to use in recruiting their peers into the survey. These variables are named C1, C2, and C3 (there were a maximum of three coupons used in this study) in the data set (Figure 1.11).
- 5. "Max # of Coupons" (optional): This is the maximum number of coupons used in the survey. For instance, if during the beginning of the survey four coupons were provided to each participant to use in recruiting others and this was reduced to three coupons in the middle of the survey and then two near the end of the survey, the maximum number of coupons would be four. There were a maximum of three coupons used in this survey (Figure 1.11)
- 6. "Population Size Estimate" (optional if you use the default): The mid-range population size estimate for the survey below is 1500 (Figure 1.11). You only need to put in the mid range, but can also put in the low and high range if you know that. This is used in some of the computations. An accurate population size estimation is often unknown, so it is all right to make an approximate guess which should not be close to the sample size. The defaulted low estimate of the population size is 500, the mid estimate is 1000 and the high estimate is 1500.

Once you completed these steps, click "Run".

	Data Format Coupon Recuiter	
Variables City F8		Subject ID chestionar
F9 PS1 PS2 PS7		Network Size
PS9_3 PS9_5 PSS7 PSP7		Subject's Coupon coupon
PSP9 PSC5 PSC6		Coupons c1 c2
Optional		c3
Max #	of Coupons: 3	

# Figure I.II Loading the RDS data window: partial view

<sup>1</sup> Note that the pieces of information entered do not need to be in any special order in your data set (for instance, the variable "net" could have been at the end of the data set rather than the first variable of the data set).

### **Data Viewer**

When data are loaded the "Data Viewer" window will open. There are two tabs in the "Data Viewer" window: "Data View" and "Variable View". The "Data View" looks similar to a Microsoft Excel worksheet (Figure 1.12). At the top, centre, is the name of the data set: FSW.Demo.RDS.Analyst.

🛓 Da	ita Viewer	-			-		10-	2.0	<b>1</b>					uff. 1	
File E	dit Help														
											Data Se	et			
									(d	f) FSW.Dem	.RDS.Analys		•		
Data \	/iew Variab	le View													
	net	coupon	c1	c2	c3	chestionar	City	F8	F9	PS1	PS2	PS7	PS9_3	PS9_S	5 PSS
1	20	201	2011	2012	2013	B0100000		7	1	1	5	16	1	1	49
2	10	202	2021	2022	2023	B0200000		7	1	1	5	17	1	NA	50
3	20	203	2031	2032	2033	B0300000		7	1	1	5	18	1	NA	NA
4	11	L 204	2041	2042	2043	B0400000		7	0	0	8	16	0	4	1
5	15					B0500000		7	1	1	1	15	1	2	NA
6	2	2 2011	20111	20112	20113	B0110000		7	1	1	5	15	1	1	NA
7	8	3 2012	20121	20122	20123	B0120000		7	1	1	8	17	1	5	NA
8	30	2013	20131	20132	20133	B0130000		7	1	1	7	16	1	6	NA
9	3	3 2021	20211	20212	20213	B0210000		7	0	0	5	18	1	NA	4
10	10	2032	20321	20322	20323	B0320000		7	1	0	4	16	NA	5	NA
11	4	1 2041	20411	20412	20413	B0410000		7	0	0	9	14	1	1	NA
12	15					B0430000		7	0	0	9	19	0	1	5
13	3					B0510000		7	1	1	5	17	1	1	NA
14	15		20521	20522		B0520000		7	1	0	4	15	1	NA	NA
15	6		20531	20532		B0530000		7	1	0	4	20	1	NA	NA
16	3					B0131000		7	0	0	5	12	1	NA	NA
17	12					B0132000		7	1	1	3	18	1	NA	3
18	20			201332		B0133000		7	1	0	4	15	1	NA	NA
19	3	3 20321	203211	203212	203213	B0321000		7	0	0	4	16	NA	NA	NA

#### Figure 1.12 Data View

More than one data set can be loaded into the "Data Viewer" and accessed during an RDS-A session. To identify the data set you want to work with, use the down arrow button to the right of the file name and a drop down box of all the files loaded during the same work session will be displayed. You can select a specific file and it will appear in the "Data Viewer". For instance, in Figure 1.13, there are three files from which to choose during this session.

	Da	ata S	et			_				
(df) MAU	(df) MAURITIUS_RDS_FSW_Final_Dataset ▼									
	.Demo.RDS.A									
(df) MAU	RITIUS_RDS_	FSW	_Final_[	Dataset_	23.se	er –				
DA (rds) FSV	V.Demo.RDS.	Analy	st1			Q1				
13500086	1		32	Black Ri	ve	Yes				
13500086	1		42	Pample	mo	Yes				

#### Figure 1.13 Selecting from multiple files

Take a minute to look at the data under the "Data View" tab in Figure 1.12 and Figure 1.14. Some things to note are the following:

- Missing data are automatically transformed into "NA"
- Text can remain in the file (only shown in Figure 1.14)
- The seeds do not need to be identified as missing (RDS-A can recognize the seeds since they were not recruited by anyone).

File E	dit Help												
										Data	a Set		
									(rds) FSW	.Demo.RDS.A	nalyst1		
D-t- 1													
Data V	view Variable	View											
	. Q111wor	Q112Sup	Q113Sup	Q201age	Q202part	Q203payi	Q203NO	Q204last	Q301mo	Q302payi	Q303payi	Q304pay	Q305
1	3	Yes	3	16	10	9	1	Yes	One mont	3	2200	Yes	My pa
2	3	Yes	2	14		3	1	No	One week	1	1000	Yes	Mysel
3	NA	Yes	2	16	28	25	3	Yes	Yesterday	3	800	Yes	Mysel
4	NA	No	NA	18	2	2	0	Yes	One week	2	5000	Yes	Mysel
5	NA	Yes	2	14	7	6	1	Yes	One week	1	500	Yes	Mysel
6	NA	Yes	6	14	10	6	4	Yes	Yesterday	5	2000	Yes	Mysel
7	NA	Yes	3	18	24	24	0	No	Yesterday	3	2500	No	
8	NA	No	NA	15		7	0	Yes	One week	4	2500	Yes	Mysel
9	NA	Yes	2	15		14	1	Yes	Yesterday	8	500	Yes	Myself
10	NA	Yes	4	12	20	14	6	No	Yesterday	3	500	No	
11	NA	Yes	1	15	30	30	0	Yes	Yesterday	2	1000	Yes	Myself
12	NA	Yes	1	17	16	15	1	No	One week	5	700	Yes	Myself
13	NA	Yes	2	15	15	11	4	No	Yesterday	3	1500	No	
14	NA	Yes	1	15	11	10	1	Yes	One week	3	1000	Yes	Myself
15	NA	No	NA	11	25	25	0	Yes	Yesterday	2	2000	Yes	Myself
16	NA	Yes	2	16	3	3	0	No	One week	2	1500	No	
17	NA	No	NA	13	13	12	1	Yes	Yesterday	2	200	Yes	My pa
18	2	Yes	2	15	5	4	1	Yes	One mont	2	2000	Yes	Myself
19	1	Yes	3	20	50	50	0	Yes	Yesterday	2	500	Yes	Joint c

Figure 1.14 Data view of another file during the same work session

## Identification of seeds and waves in Data Viewer

Under the "Data View" tab, at the far right, are three additional columns: The recruiter ID, seed and wave. The "recruiter.ID" column identifies the seeds and the "wave" column identifies the number of waves in the sample. In Figure 1.15, the last column shows five seeds (wave 0), 10 participants who were in wave one, 13 participants who were in wave two and so on.

	dit						
						Data S	et
					(r	ds) FSW.Dem	io.R 🔻
Data \	lew	Variable V	/iew RDS				
		stiwt	W9Migrant	W9Natio	recruiter.id	seed	wave
1	1	0.14716	1	4	0	B0100000	0
2	1	0.29432	0	2		B0200000	0
3	2	0.14716	0	3	0	B0300000	0
4	1	0.26756	0	3	0	B0400000	0
5	1	0.19621	0	3	0	B0500000	0
6	1	1.47159	1	4	B0100000	B0100000	1
7	2	0.36790	0	3	B0100000	B0100000	1
8	2	0.09811	0	2	B0100000	B0100000	1
9	1	0.98106	0	3	B0200000	B0200000	1
10	2	0.38496	0	2	B0300000	B0300000	1
11	1	0.96240	1	4	B0400000	B0400000	1
12	2	0.19621	NA	NA	B0400000	B0400000	1
13	1	0.98106	0	2	B0500000	B0500000	1
14	0	0.25664	0	1	B0500000	B0500000	1
15	1	0.49053	0	2	B0500000	B0500000	1
16	1	0.98106	1	4	B0130000	B0100000	2
17	2	0.32080	0	2	B0130000	B0100000	2
18	1	0.19248	0	2	B0130000	B0100000	2
19	2	0.98106	0	2	B0320000	B0300000	2
20	2	1.28321	0	2	B0320000	B0300000	2
21	0	0.76992	1	4	B0410000	B0400000	2
22	2	0.38496	0	2	B0410000	B0400000	2
23	1	0.64160	1	4	B0410000	B0400000	2
24	1	0.73579	1	4	B0430000	B0400000	2
25	1	1.92481	0	3	B0510000	B0500000	2
26	0	0.96240	1	4	B0510000	B0500000	2
27	2	0.25664	1	4	B0510000	B0500000	2
28	0	0.98106	0	3	B0520000	B0500000	2

Figure 1.15 Identification of seeds and waves in Data Viewer

# **Variable View**

Under the "Variable View" tab (Figure 1.16), normally three columns of data are displayed in addition to the far left column of serial numbers. The first column is "Variable", which displays the variable name. The second column is "Type". There are eight possible types of data under the "Type" column (see Figure 1.17): "Character" or string, "Factor" (the term "factor" in the underlying R engine designates a categorical variable; factors are nominal but they are used to represent both ordinal and nominal variables), "Double" (non-integers), "Integer", "Logical" (true, false), "Date", "Time", and "Other". Sometimes if the program is not reading your data properly, you will need to change the type of data assigned to your variables.

The third column is "Factor Levels". In the data set in Figure 1.16, all of the factor (or category) labels from the original SPSS Statistics file were retained in RDS-A.

🛓 Data Viewe	er		and the last
ile Edit Hel	p		
	iable View		
Data View Va			
Variable		Туре	Factor Levels
16 0107ma	tial	Factor	(1) Single; (2) Married; (3) Divorce/separated; (4
	naryPartner	Factor	(1) Alone; (2) Husband; (3) Male partner; (4) Par
	FirstMoney	Double	(2) Money (2) Maccana, (b) Maic particly (1) Par
	neyOtherWork	Factor	(1) Yes; (2) No;
20 Q111wor		Double	(-/ / (-/ /
	portAnyone	Factor	(1) Yes; (2) No;
	portHowMany	Double	(-) ·, (-) ····,
23 Q201age		Double	
	nershowMany	Double	
	ingPartners	Double	
- 1/	VpayingPartner	Double	
	SexCondom	Factor	(1) Yes; (2) No;
	neyLastTime	Factor	(1) Today; (2) Yesterday; (3) One week ago; (4)
	ingClientsNum	Double	
	ingLastAmount	Double	
	nigUseCONDOM	Factor	(1) Yes; (2) No;
	ingSuggestCONDOM	Factor	(1) Myself; (2) My partner; (3) Joint decision;
	ingClientNotPay	Factor	(1) Yes; (2) No;
	ingNoAvailableCondom	Factor	(1) Yes; (2) No;
	ingTooExpensive	Factor	(1) Yes; (2) No;
36 Q306pay	ingNoPleasure	Factor	(1) Yes; (2) No;
37 Q306pay	ingNoPleasurePartner	Factor	(1) Yes; (2) No;
38 Q306pay	ingDidntThink	Factor	(1) Yes; (2) No;
39 Q306pay	ingTrustPartner	Factor	(1) Yes; (2) No;
40 Q306pay	ingDrunk	Factor	(1) Yes; (2) No;
	ingAllergic	Factor	(1) Yes; (2) No;
42 Q306pay	ingSafePartner	Factor	(1) Yes; (2) No;
43 Q307pay	ingNoCondomReason	Factor	(1) Not available; (2) Too expensive; (3) Not plea
44 Q307oth	er	Factor	(1)
45 Q308pay	ingFrequencyCondom	Factor	(1) Always; (2) Most times; (3) About half the tin

Figure 1.16 Variable View

🍝 D	ata Viewer		-						
File	File Edit Help								
Data	View Variable View								
	Variable	Туре	Facto						
16	Q107martial	Factor	<b>T</b> (1) Si						
17	Q108primaryPartner	Character	(1) Al						
18	Q109ageFirstMoney	Factor							
19	Q110moneyOtherWork	Double	(1) Ye						
20	Q111work_cat	Integer							
21	Q112SupportAnyone	Logical	(1) Ye						
22	Q113SupportHowMany								
23	Q201ageFirstSEX	Date							
24	Q202partnershowMany	Time							
25	Q203payingPartners	Other							
26	Q203NONpayingPartner	Double							
27	Q204lastSexCondom	Factor	(1) Ye						
28	Q301moneyLastTime	Factor	(1) T(						
29	Q302payingClientsNum	Double							
30	Q303payingLastAmount	Double							
31	Q304paynigUseCONDOM	Factor	(1) Ye						
32	Q305payingSuggestCONDOM	Factor	(1) M						

Figure 1.17 Data types

## Save data

To avoid having to re-enter the data parameters again, you can save your file as a RDSOBJ file which is easily recognizable by RDS-A as an RDS file. To do this, go to "Console" and select File|Save Data. Once you do this the following window will open and you will select "OK" to save the file (Figure 1.18). Once you click "OK", you will need to tell it where to save the file and give it a name. Once you select a file name (the name selected below is the same name used for the original sPSS file), click "Save".

	Save F	RDS Data Set RDS Data FSW.Demo K		cel		
Save as .rds	-	rst manual		• 🗊 🖾	•	X
Recent Items	FSW Den	no RDS Analy no RDS Analy lyst step by st	st			
Desktop						
My Documents	File name: Files of type:	FSW Demo R All Files	DS Analyst	•		Save Cancel

# Figure 1.18 Saving RDS dataset from RDS-A

In Figure 1.19, you can see the saved RDSOBJ file, highlighted. This action will also automatically convert your file to a DL file and a vna file to be used in recruitment graphics (see the step-by-step guide to NetDraw in Section 2 of this Supplement).

Organize ▼ _ ) Open Share with ▼ Burn New folder								
☆ Favorites	Name	Date modified	Туре	Size				
🔈 Downloads	FSW Demo RDS Analyst.DL	4/4/2013 3:51 PM	DL File	21 KB				
👢 Dropbox 🗏	FSW Demo RDS Analyst.rdsobj	4/4/2013 3:51 PM	RDSOBJ File	320 KB				
lace 😓 Recent Place	E FSW Demo RDS Analyst	4/4/2013 3:51 PM	vna Document	0 KB				
💻 Desktop	🕙 RDS Analyst step by step	4/4/2013 3:14 PM	Microsoft Word Document	5,182 KB				
Documents	🖷 FSW Demo RDS Analyst	3/27/2013 1:10 PM	SPSS Statistics Data Document	118 KB				
📜 Libraries	🖺 FSW Demo RDS Analyst	3/27/2013 9:30 AM	Microsoft Excel Worksheet	247 KB				

Figure 1.19 Saved RDS-A datasets from RDS-A

# Working with data

# **Recode variables**

To recode a variable in the data set, go to Data|Recode Variables. Select a variable from the left-hand side panel of the "Recode Variables" window and move it into the panel labeled "Variables to Recode" to the right by using the blue arrow keys. If you want to recode your variable into another variable name, you can change the new variable name by clicking "Target" and entering the new name. Select "Define Code". In the example in Figure 1.20, the variable selected is "lab\_HIV", which are results from the HIV test (see window on left, "Recode Variables").

After selecting "Define Code", the "Set Variable Codings" window will open (Figure 1.20, window on right). "Lab\_HIV" is currently coded as 1 for yes and 0 for no. For example, if you want to change 0 to 2 so that 2 is the code for no, type the code you want to change (0) into the "Value" panel, under "Code". The options for values are "=", "≥" and "≤" (the later two options are for ranges) or "Else". Under the 0, put 2 into the panel to the right of "into" and then select "Add". The new coding will appear in the bottom panel, "Recodings". Click "OK".

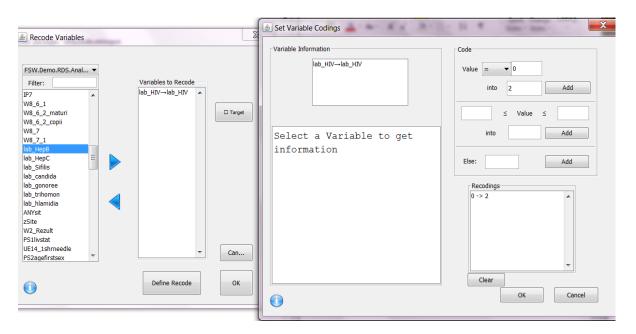


Figure 1.20 Recoding variables and setting variable codes

# **Compute weights**<sup>1</sup>

To compute weights used in analysis of RDS data select Data|Compute Weights|Run. The "Compute Weights" window will open (Figure 1.21). Here you can compute weights for a specific variable using a specific "Type" (estimator method) by typing the variable name into the "Group Variable" panel. The Gile's Successive Sampling (Gile's ss) method and RDS-II do not need a group variable, however for RDS-I a group variable is needed.

<sup>&</sup>lt;sup>1</sup> Weights are often computed and exported for running regression analyses in other software packages. However, regression analyses for network data are complicated and the use of exported weights will not ensure that your regression output is representative of anything and should be noted as a limitation of your analysis. Furthermore, exporting weights to another program to run descriptive analyses will not eliminate biases in the sample.

Scompute Weights	X
RDS Data FSW.Demo.RDS.Analyst	New Variable weights
Type Gile's SS RDS-I RDS-I/DS RDS-II Arithmetic Mean	Population Size Estimate Number of Simulated Sample 1000 Group Variable
Run	Fixed Population Size           Reset         Cancel

Figure 1.21 Compute Weights

# Analysis of the sample

To analyse the sample (unweighted analysis), use the "Sample" tab in the "Console" window.

# **Sample frequencies**

To conduct frequencies of the sample click Sample|Frequencies. A window named "Run Frequencies" will open. From the left-side panel, select the variable(s) for analysis (you can select more than one variable at a time) by highlighting the variable and putting it/ them in the right-side panel using the blue arrows. In the example in Figure 1.22, "lab\_ HIV", "lab\_HepB" and "lab\_HepC" are selected. Click "OK".

length Stranger Stran	-		X
FSW.Demo.RDS.Analyst         Filter:         W8_7         W8_7_1         Iab_Sifilis         Iab_candida         Iab_gonoree         Iab_trihomon         Iab_hlamidia		Run Frequencies On: lab_HIV lab_HepB lab_HepC	
ANYsit zSite W2_Rezult	- -	ОК	Cancel

Figure 1.22 Selecting variables: Sample frequency

Note: Most windows have an information icon 🔍 which you can click to learn more about the feature you are using. Upon clicking this icon, you will be taken to the manual at http://www.deducer. org/. You will need to be connected to Internet to access this. This online manual is being updated frequently so any updates to the software can be found here.

The output appears in the "Console View" panel. In Figure 1.23, The "Frequencies" for lab\_HIV show the values 0 (HIV negatives) and 1 (HIV positives). The "Console View" displays the number of cases (0 = 551, 1 = 107), the percentage (0 = 83.7%, 1 = 16.3%) and the cumulative percentage. If your data have labels, the labels will show up in the far left-hand column. Below the "Frequencies" table is a "Case Summary" table with information about the number of valid responses, missing data, total and percentage missing.

差 Console	-	-	and a state	-	-	
File Edit Workspa	ce Data	Sample P	opulation Plot	s Packages	& Data Window	v Help
2 🎝 🖥	4	→ ``	s 🔲 💼	•	<i>8</i> % 🍳	
rownames(FSW.Der [Frequencies] lab F	Console	View Eleme	nt View			
[Frequencies] lab_H [Frequencies] lab_H		luencies luencies	(lab_HIV)			
			# of			
		Value	Cases	%	Cumulative	%
	1	0	551	83.70	83.	70
	2	1	107	16.30	100.	00
	Cas	e Summa	ry ( lab_HI\	V)		
					%	
		Valid	Missing	Total	Missing	
	1	658.00	0.00	658.00	0.00	

Figure 1.23 Sample frequency output of HIV

# Descriptives

Click Sample|Descriptives from the "Console" window. The "Descriptives" window will open (Figure 1.24). From the left-side panel, select the variable(s) for analysis (you can select more than one variable at a time) by highlighting it/them and putting it/them in the right-side panel using the blue arrows. To stratify variables, put them into the "Stratify By" panel. The variable selected for "Descriptives of" is PS1 (number of sexual partners in the past three months [continuous variable]). Click "Run".

Descriptives			X
FSW.Demo.RDS.Analyst Filter:	-	Descriptives of: PS1	
net coupon c1 c2 c3 chestionar City			
F8 F9 PS2 PS7 PS9_3 PS9_5	-	Stratify By:	
0	Continue	Reset	Cancel

Figure 1.24 Selecting a variable: Descriptive analysis

Once you select the variable(s) you want to analyse, another window, also named "Descriptives", will open (Figure 1.25). This window will allow you to select from the "Functions" panel the descriptive analyses you want by moving the listed functions to the "Run Descriptives" panel. The functions selected for this analysis are "Mean", "St. Deviation", "Valid N", "Maximum", "Minimum" and "Median". Click "Run".

Secriptives		X
Functions 25th Percentile 75th Percentile Skew Kurtosis		Run Descriptives Mean St. Deviation Valid N Maximum Minimum Median
		Custom
0	Run	Reset Cancel

#### Figure 1.25 Selecting functions: Descriptive analysis

Figure 1.26 shows the output of the descriptive analysis of sexual partners in the past three months. In this output, there is an average (mean) of 5.48 sexual partners in the past three months with a standard deviation of 2. The valid N is 658, the maximum is nine partners and the minimum is one partner. The median is five partners.

#### **Descriptive Statistics**

	Mean	St. Deviation	Valid N	Maximum	Minimum	Median
PS1	5.48	2.00	658	9.00	1.00	5.00

## Figure 1.26 Output: Descriptive analysis

## **Contingency tables (crosstabs)**

Click Sample|Contingency Tables from the "Console" window. A window named "Contingency Tables" will open (Figure 1.27). From the left-side panel, select the variable(s) for analysis (you can select more than one variable at a time) by highlighting the variable and putting it in the right-side panel using the blue arrows. Put at least one variable in the "Row" panel and one variable in the "Column" panel. In Figure 1.27, "lab\_HIV" is selected for the row and "lab\_HepC" is selected for the column. Click "Run".

FSW.Demo.RDS.Analyst	-		Row	
Filter:			lab_HIV	Cells
HV2_HepC				
HV14_1				Statistics
DA5_5				
 IP7				
W8_6_1				
W8_6_2_maturi				Results
W8_6_2_copii			Column	
N8_7	=		lab_HepC	
W8_7_1				
ab_HepB				
ab_Sifilis				
ab_candida				
ab_gonoree				
ab_trihomon				
lab_hlamidia			Stratify By	
ANYsit				
zSite				
W2_Rezult		-	Subset	
PS1livstat	-		÷	
-			Run Rese	t Cancel

## Figure 1.27 Contingency Tables: Selecting variables

On the right-hand site of the "Contingency Tables" window is a tab named "Cells". By selecting this tab, a window named "Table Cell Contents" will open (Figure 1.28). In addition to observation counts, this window allows you to select from the following:

- "Percentages"
  - o "Row" (within each row percentage)
  - o "Column" (within each column percentage)
  - o "Total" (percentage in cell).
- "Chi-squared"
  - "Expected" (the expected count of the cell if there were no relationship between the two variables)
  - o "Residuals" (the observed count minus the expected count)
  - "Standardized Residuals" (the residuals standardized such that, if the two variables were independent, they have mean 0 and standard deviation 1; these residuals are useful in determining which cells of a contingency table contribute most to a significant chi-squared test)
  - o "Adjusted Residuals" (these adjust the residuals by the row and column totals).

Click "OK".

Table Cell Contents	
Percentages          Image: Percentages	Chi-Squared Expected Residuals Standardized Residuals Adjusted Residuals
Don't print tables	OK Cancel

Figure 1.28 Contingency Tables: Selecting table cell contents

On the right-hand site of the Contingency Tables window is a tab named "Statistics". By selecting this tab, a window named "Table Statistics" will open (Figure 1.29). For this analysis, the "Chi-Squared" statistic is selected. Click "OK".

Table Statistics	X
Nominal By Nominal	Ordinal By Ordinal
Likelihood	🔲 Spearman's Rho N
Fisher's Exact	Nominal By Ordinal
Cross-Stratum	Kruskal-Wallis
Mantel Haenszel	N M
N III	
	OK Cancel

## Figure 1.29 Contingency Tables: Selecting table statistics

Figure 1.30 shows output from the contingency analysis. For instance, in the Console View in Figure 1.30, 58.9% of those infected with HIV, were also infected with hepatitis C.

File Edit Workspace	ce Data	Sample Po	pulation P	lots Packag	ges & Data Windo
2 🤳 🗖	-	4	· 🗐 💼		<i>8</i> % 🍳
rownames(FSW.Der	Console	View Elemen	t View		
[Frequencies] lab_H [Frequencies] lab_H [Frequencies] lab_H	Со	ntingen	cy Tab	oles	
[descriptive] lab_HIV [descriptive] F8	lab_H	IV by lab_H	epC across	levels of	
wtd.descriptive.table [descriptive] PS1			lab_	НерС	
[descriptive] PS1		lab_HIV	0	1	Row Total
[descriptive] PS1		Count	462	86	548
[contin.table tests] 1:	0	Row %	84.31%	15.69%	83.66%
		Column %	91.30%	57.72%	
		Count	44	63	107
	1	Row %	41.12%	58.88%	16.34%
		Column %	8.70%	42.28%	
	Co	lumn Total	506	149	655
	(	Column %	77.25%	22.75%	
		for lab_HIV Chi Squared	by lab_He	pC across	

Figure 1.30 Contingency Tables: Output

# **Recruitment homophily**

Click Sample|Recruitment Homophily from the "Console" window. A window named "Homophily" will open (Figure 1.31). From the left-side panel, select the variable(s) for analysis (you can select more than one variable at a time) by highlighting the variable and putting it in the right-side panel using the blue arrows. Click "Run".

left Homophily	-	-	-	X
FSW.Demo.RDS.Anal	yst 🔻		Variables	
Filter:			lab_HIV	
IP7		<b>I</b>		
W8_6_1				
W8_6_2_maturi		4		
W8_6_2_copii W8_7				
W8_7_1	=			
lab_HepB				
lab_HepC				
lab_Sifilis				
lab_candida				
lab_gonoree		Subse	et expression	
lab_trihomon				
lab_hlamidia ANYsit				-
ANTSIC	Ψ.			
	Ru	n	Reset	Cancel

#### Figure 1.31 Recruitment homophily: Selecting a variable

Figure 1.32 shows an example of recruitment homophily using HIV serostatus. In this case, it is the ratio of number of recruits that have the same HIV serostatus as their recruiter to the number we would expect if there was no homophily on HIV serostatus. This homophily is calculated from the recruitment chain (rather than the population of social ties). For example, if the recruitment homophily on HIV status is about 1 (specifically 1.069922), there is little effect of recruitment homophily on HIV serostatus (as the numbers of homophilous pairs are close to what we would expect by chance).

```
Recruitment Homophily for lab_HIV

Homophily = 1.069922

lab_HIV of recruit

lab_HIV of respondent 1 0

1 34 79

0 67 465

Number of cases in table: 645

Number of factors: 2

Test for independence of all factors:

Chisq = 21.599, df = 1, p-value = 3.36e-06
```

Figure 1.32 Recruitment homophily: Output

# Analysis of the population

Analysing the population (weighted analysis) is done using the "Population" tab on the Console window. RDS-A offers several estimators, all of which have variations on the methods used to derive estimates and confidence intervals.

## **RDS** estimators<sup>1</sup>

Table 1.2 provides information about the estimators available in RDS-A. For more information on these estimators, review the source provided in the first column and listed in the reference section at the end of this guide.

<sup>&</sup>lt;sup>1</sup> Not including Heckathorn, 1997.

RDS estimator and type	Data needed	Limitations	Variance estimation	Analysis feature	Software available
Heckathorn, 2002 Reciprocity model- based estimator (RDS I); Linear least squares used for >2 categories and data smoothing used to estimate >2 categories	Recruitment matrix; self- reported network sizes	Limited to categorical data and by RDS assumptions	Bootstrap	Controls for differences in network sizes, homophily across groups; uses data smoothing for narrower confidence intervals	Yes: STATA RDS estimator (Schonlau and Liebau, 2010)
Salganik and Heckathorn, 2004 Reciprocity model- based estimator (RDS I)	Recruitment matrix; self- reported network sizes	Limited to categorical variables and by RDS assumptions	Bootstrap	Proof that estimate is asymptotical-ly unbiased; estimate of average group network size	Yes: RDSAT, RDS Analyst
<b>Heckathorn,</b> 2007 Dual-component estimator (RDS I)	Recruitment matrix; self- reported network sizes	As other estimators, limited by RDS assumptions	Bootstrap	Allows analysis of continuous variables; controls for differential recruitment	Yes: RDSAT, RDS Analyst
Volz and Heckathorn, 2008 Probability-based estimator; data- smoothing (RDS II)	Recruitment matrix; self- reported network sizes	Limited to nominal variables	Analytic	Allows analysis of continuous variables; shows convergence between reciprocity- and probability- based RDS estimators; uses data smoothing to control for differential recruitment	Yes: RDS Analyst
Gile's successive sampling, 2011 Probability-based estimator	Working estimate of population size, recruitment matrix; self- reported network sizes	Currently limited to categorical variables (may be extended)	Bootstrap	Corrects for finite population effects	Yes: RDS Analyst
Gile and Handcock's model-assisted estimator, 2012 Probability-based estimator	Working estimate of population size, recruitment matrix; self- reported network sizes, self-reported composition of contacts helpful but not necessary	Currently limited to binary variables (may be extended)	Bootstrap	Corrects for finite population effects and for some forms of seed bias	To be made available in RDS Analyst

### Table 1.2 Estimators available in RDS-A<sup>1</sup>

<sup>&</sup>lt;sup>1</sup> Parts of this table were adapted from Wenjert C, Heckathorn DD. Respondent-driven sampling: Operational procedures, evolution of estimators, and topics for future research. In: Williams M, Vogt PW, eds. The SAGE handbook of innovation in social research methods. London, Sage Publications, 2011.

### **Frequency** estimates

To analyse frequency estimates click Population|Frequency Estimates. A window named "Population Frequency Estimates" will open. From the left-side panel, select the variable(s) for analysis (you can select more than one variable at a time) by highlighting the variable and putting it into the right-side panel using the blue arrows. In the example in Figure 1.33, "lab\_HIV", "lab\_HepB" and "lab\_HepC" have been selected. Also note the following items which are set as defaults in the program.

- "Weights": Gile's ss.
- "Confidence": 95%. The confidence interval is computed using Gile's bootstrap method. This is a computationally-intensive procedure and can take a minute or longer to complete.
- "Population Size"1: 1500 (this was added to the data viewer when loading the data set).
- Number of simulations/iterations: 1000.

Click "Run".

FSW.Demo.RDS.Analyst	Variables lab_HIV lab_HepB
HV14_1 DA5_5 IP7 W8_6_1 W8_6_2_maturi W8_6_2_copii W8 7	lab_HepC
W8_7_1 lab_sifilis lab_candida lab_gonoree lab_trihomon lab_hlamidia ANYsit zSite W2_Rezult PS1livstat UE14_1shrneedle PS2agefirstsex	Subset expression Weights Gile's SS RDS-I RDS-I/DS RDS-II Arithmetic Mean
Confidence # Bootstraps 0.95 50	Population Size         # Sim. / iter.           1500         1000

#### Figure 1.33 Population Frequency Estimates: Selecting variables

The output appears in the output panel of the "Console" window (Figure 1.34). In Figure 1.34, the frequencies for "lab\_HIV" show the values 0, which are HIV negatives and 1 which are HIV positives. The window displays the estimates (0 = 83.8%, 1 = 16.2%), confidence bounds (80.8 to 86.9 for group 0 and 13.1 to 19.3 for group 1), the estimated design effect (2.15), standard error (0.016) and number of cases for each category (0 = 551, 1 = 107). If your data have labels, the labels will show up in the left-hand column.

<sup>&</sup>lt;sup>1</sup> Needed for the Gile's successive sampling and Gile and Handcock's model-assisted estimator.

			Olic 3 00 Estimate			
	Point Estimate	95% Lower Bound	95% Upper Bound	Estimated Design Effect	Standard Error	Sample Size
0	0.8384	0.8075	0.8693	2.15	0.0158	551
1	0.1616	0.1307	0.1925	2.15	0.0158	107
					Total	658

Cile's SS Estimate for lab HIV

Figure 1.34	Frequency	estimate	of HIV:	Output
-------------	-----------	----------	---------	--------

#### **Parsing data**

In the "Population Frequency Estimates" window, it is also possible to parse data using the "Subset expression" panel (Figure 1.35). In the window in Figure 1.35, "lab\_HepC" is added to the variables panel and "lab\_HIV" (along with "==1", the code for HIV positive) is written into the "Subset expression" panel. Click "Run".

Population Frequency Estimate	s
FSW.Demo.RDS.Analyst         Filter:         W8_6_2_copii         W8_7         W8         Iab_How         Iab_hlamidia         ANYsit         ZSite         W2_Rezult         PS1ivstat         UE14_1shrneedle         PS2agefirstSELL         PS2agefirstSex_2	Variables lab_HepC Subset expression lab_HIV==1 Weights Gile's SS RDS-I RDS-I/DS
PS8recode	<ul> <li>RDS-II</li> <li>Arithmetic Mean</li> </ul>
Confidence # Bootstraps	Population Size # Sim. / Iter.
0.95 50	1500 1000
Run	Reset Cancel

## Figure 1.35 Parsing data

In the example in Figure 1.36, the sample sizes for each category are equal to 107 (the sample size for all those who were HIV positive in Figure 1.30). Figure 1.36 shows that among those who were HIV positive, 47.7% (95% CI: 37.1, 58.4) were also infected with hepatitis C and 52.3% (95% CI: 41.6, 62.9) were not infected with hepatitis C. The figure also displays design effects and standard errors.

			Gile's SS Estimate f	or lab_HepC		
	Point Estimate	95% Lower Bound	95% Upper Bound	Estimated Design Effect	Standard Error	Sample Size
0	0.5225	0.4157	0.6292	1.37	0.0545	44
1	0.4775	0.3708	0.5843	1.37	0.0545	63
					Total	107

Gile's SS Estimate for lab_HepC
---------------------------------

#### **Descriptive estimates**

To produce descriptive estimates, click Population|Descriptive Estimates. A window named "Population Estimates" will open. From the left-side panel, select the variable(s) for analysis (you can select more than one variable at a time) by highlighting the variable and putting it into the right-side panel using the blue arrows. In Figure 1.37, PS2 (age at first sexual intercourse) is selected. From the "Statistics" panel, "Minimum", "Maximum" and "Median" are selected. Click "Run".

Population Estimates	-traditional l	X
FSW.Demo.RDS.Analyst Filter: City F8 F9 PS1 PS7		PS2
PS9_3 PS9_5 PSS7 PSP7 PSP9 PSC5 PSC6 PSC19 SG1 SG3 SG5	Ţ	Stratify by:
Statistics 25th Percentile 75th Percentile	Minimum Maximum Median	Subset expression  Weights  Gile's SS  Arithmetic Mean  Population Size # Sim. / Iter.  I500 1000
		Run Reset Cancel

#### Figure 1.37 Descriptive estimates

In the output in Figure 1.38, the minimum age at first sexual intercourse was 6 years old, the maximum was 24 years old and the median was 16 years old.

```
> wtd.descriptive.table(vars=d(PS2), data=FSW.Demo.RDS.Analyst, func.names=c("Minimum", "Maximum", "Median"),
+ weights=compute_weights(FSW.Demo.RDS.Analyst, weight.type="Gile's SS", N=1500, number.ss.samples.per.iteratio
$`strata: all cases`
Minimum.PS2 Maximum.PS2 Median.PS2.50%
6 24 16
```

#### Figure 1.38 Descriptive estimates: Output

#### **Population crosstabs**

To produce population crosstabs, click Population|Population Crosstabs. A window named "Population Crosstabs" will open. From the left-side panel, select the variable(s) for analysis (you can select more than one variable at a time) by highlighting the variable and putting it into the right-side panel using the blue arrows. In the example in Figure 1.39, "lab\_HIV" is selected for the "Rows" and "EverHIVtest" (persons who reported ever having had an HIV test) is selected for the "Columns". Click "Run".

Population	Crosstabs	_		X
Population FSW.Demo.RD Filter: W8_6_2_copii W8_7 W8_7_1 lab_Hep8 lab_HepC lab_Sifilis lab_candida lab_gonoree lab_trihomon lab_hlamidia ANYsit zSite W2_Rezult PS1livstat UE14_1shrneed PS2agefirstsex PS2agefirstsex PS2agefirstsex PS9_totnumpat PS9_4_comnun ps9_3permpart ps9_5caspart educ W_8_71life agegrp otherincome PS9_3permy_n PS9_5casy_n	S.Analyst	Subset expre Weights © Gile's S: © RDS-II		
	Ru	n Re	eset Can	icel

## Figure 1.39 Population Crosstabs

The "Console View" (Figure 1.40) displays "EverHIVtest" as columns: 1 is yes, 2 is no. "Lab\_HIV" (HIV serostatus) is presented as the rows: 1 is yes, 0 is no. This figure shows that, among those who were HIV positive, 53.37% ever had an HIV test.

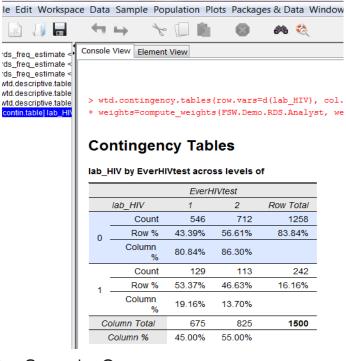


Figure 1.40 Population Crosstabs: Output

### **Population homophily**

To calculate population homophily, click Population|Population Homophily. A window named "Homophily" will open (Figure 1.41). From the left-side panel, select the variable(s) for analysis (you can select more than one variable at a time) by highlighting the variable and putting in the right-side panel using the blue arrows. In this example, "lab\_HIV" is selected. Click "Run".

left Homophily		X
FSW.Demo.RDS.Analyst	•	Variables
Filter:		lab_HIV
DA5_5	*	
IP7 W8_6_1		
W8_6_2_maturi		
W8_6_2_copii		
W8_7	Ξ	
W8_7_1 lab_HepB		Subset expression
lab_HepC		Subset expression
lab_Sifilis		
lab_candida lab_gonoree		· · · · · · · · · · · · · · · · · · ·
lab_trihomon		Population Size Estimate
lab_hlamidia	-	1500
_		
	Run	Reset Cancel

#### Figure 1.41 Population Homophily

The output is displayed in the "Console View" (Figure 1.42) as 1.39. This is the ratio of the number of recruits that have the same HIV serostatus as their recruiter to the number we would expect if there were no homophily on HIV serostatus and is calculated based on the population of social ties (rather than from the recruitment chain); also see recruitment homophily section above. For example, the homophily of 1.39 on HIV status demonstrates a moderate effect of homophily on HIV status (The numbers of homophilous pairs are more than we would expect by chance).

> print(homophily.estimates(rds.data=FSW.Demo.RDS.Analyst, outcome.va)
Population Homophily Estimate for lab\_HIV
1.389375

Figure 1.42 Population Homophily: Output

# Plots to visually explore RDS data

#### **Plot recruitment tree**

To create a recruitment tree, go to Plots|Plot Recruitment Tree. A window named "Plot Recruitment Tree" will open. Use the blue arrow keys to select a variable from the variables panel and add it to "Node Color" or "Node Size" of you want to highlight the categories in the variable. Otherwise, there is no need to select a variable and you can just click "Run". In Figure 1.43, "lab\_Sifilis" is selected and added to the "Node Color" panel. Output can be either in a "Graphics Window" or "PDF Report".

Plot Recruitment Tree			×	
FSW.Demo.RDS.Analyst	•			
Filter: W8_6_2_copii		(	Node Color (optional)	
W8_7 W8_7 1			lab_Sifilis	
lab_HIV			Node Size (optional)	
lab_HepB lab_HepC	_			
lab_candida	=			
lab_gonoree lab_trihomon lab_hlamidia			Node Label (by default the id)	
ANYsit zSite		Pause	after each plot?	
W2_Rezult		Output		
PS1livstat UE14_1shrneedle		Graphics windows		
PS2agefirstsex	-		Report	
	RI	ın	Reset Cancel	

## Figure 1.43 Plot Recruitment Tree window

Figure 1.44, displays a graphic of the recruitment trees. Red are those who had reactive tests for syphilis and black are for those who had non-reactive tests. Blue, which is not visible in this example, are for those who had indeterminate results. In this portion there are three recruitment trees. For more aesthetically pleasing recruitment trees there are other software programs such as NetDraw and Gephi.

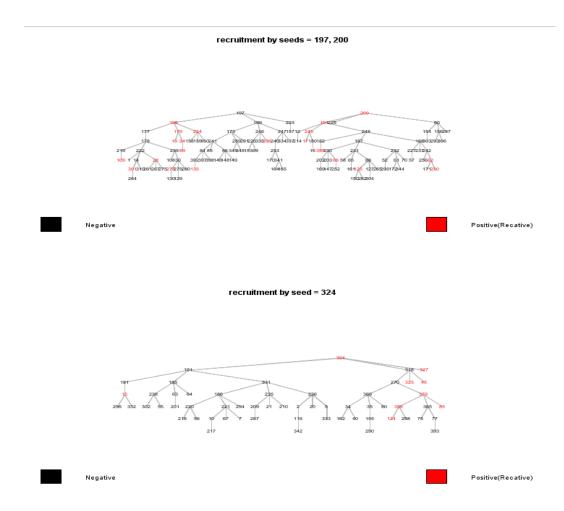


Figure 1.44 Plot recruitment tree with colour nodes: Output

#### **Recruitment diagnostics**

For other types of diagnostic graphics, go to Plots|Recruitment Diagnostics. This will open up a window called "Diagnostics Plots" (Figure 1.45). Under "Plots" (on the right-side of the window) you can click on the different types of plots you want to produce, including the "Recruitment tree". You can choose to stratify by a variable or not. No variable is selected in Figure 1.45. Each of the plot types are described in more detail below.

🛓 Diagnostic Plots	×
FSW.Demo.RDS.Analyst         Filter:         net         coupon         c1         c2         c3         chestionar         City         F8         F9         PS1         PS2         PS7         PS9_3         PS9_5         PSS7         PS9_5         PSS7         PS9_5         PS57         PS9_5         PS57         PS9_5         PS57	Stratify by (optional)  Plots  Plots  Recruitment tree  Network size by wave  Recruits by wave  Recruits per seed  Recruits per subject  Output  Graphics windows  PDF Report  Run Reset Cancel

Figure 1.45 Window for diagnostic plots

#### Network size by wave

Figure 1.46 shows a scatter plot of network size by wave. The network sizes for this sample range from 1 to just under 100. The number of waves is six. The legend to the right of the graph shows the number of subjects with identical network sizes. The largest black filled circle depicts 60 subjects. Most of the network sizes fall above zero (there should never be a network size of zero) to 25. The red lines through the network sizes are the mean of all network sizes for each wave. In the upper part of the graph are outliers having network sizes of between 75 and 100. Overall, the mean network sizes across the waves are similar.

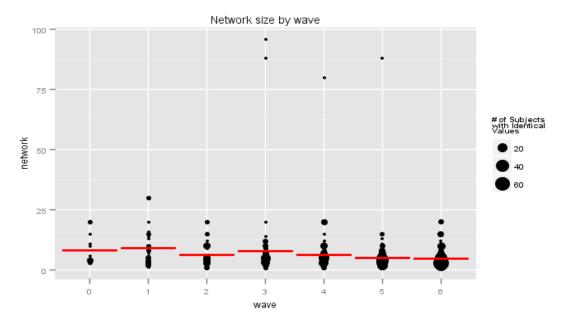


Figure 1.46 Network size by wave

#### **Recruits by wave**

Figure 1.47 shows a bar chart of recruitment by wave. The recruitment count ranges from zero to 200. The number of waves is six. There is a steady increase of recruitment by wave, starting with 0 (seeds) to wave 6, shown in the chart.

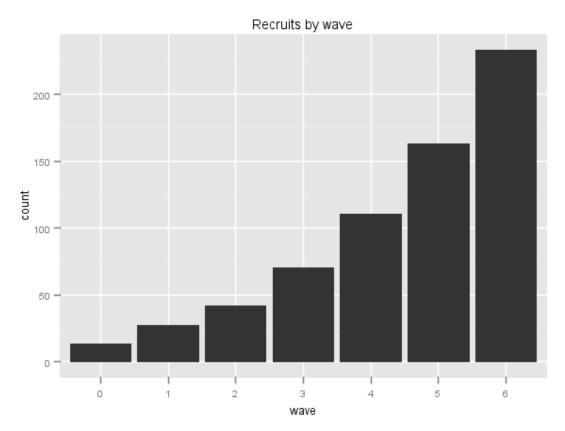


Figure 1.47 Recruits by wave

#### **Recruits per seed**

Figure 1.48 displays a bar chart of the number of recruits from each seed stratified by HIV serostatus. To the right of the graph is a legend showing that HIV positive persons are pink and HIV negative persons are blue. This example shows six seeds with different counts for recruits. The fourth bar from the left (seed ID number 322) appears to have the largest number of recruits in the sample and the largest number of HIV positive recruits.



Figure 1.48 Recruits per seed

#### **Recruits per subject**

Figure 1.49 shows a bar chart of the number of recruits per subject stratified by HIV serostatus. To the right of the graph is a legend showing HIV negative persons to be in pink, HIV positive persons to be in blue and data that are not available (i.e. recruitment coupons coded as missing data or no longer provided to participants because of the ending of the survey) to be in green. The bar to the far left shows that the majority of HIV positive recruits (in blue) did not recruit any others. In this particular survey, the recruitment process was fast so coupons were reduced from three to two and then from two to three early in the survey.

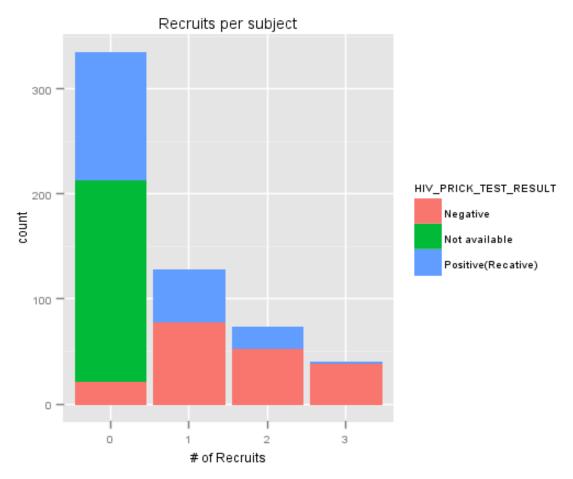


Figure 1.49 Recruits per subject

# Conclusion

This step-by-step guide was developed to provide an overview of how to use some of the most useful features of RDS-A. There are a number of other features in RDS-A which can be learnt through practice and by using the manual features available at the HPMRG website. Please be sure to sign up for the RDS-A users group on the HPMRG website in order to provide feedback, offer suggestions and ask for help.

### **References for the estimator table**

- Gile K. Improved inference for respondent-driven sampling data with application to HIV prevalence estimation. *Journal of the American Statistical Association*, 2011, 106:498:135–146.
- Gile K, Handcock MS. Network model-assisted inference from respondent-driven sampling data. Under revision. Available at: http://arxiv.org/pdf/1108.0298v1.pdf, accessed 27 January 2014.
- Heckathorn DD. Respondent driven sampling: a new approach to the study of hidden populations. *Social Problems*, 1997, 44(2):174–199.
- Heckathorn DD. Extensions of respondent-driven sampling: analyzing continuous variables and controlling for differential recruitment. *Sociological Methodology*, 2007, 37:151–207.
- Salganik MJ, Heckathorn DD. Sampling and estimation in hidden populations using respondent driven sampling. *Sociological Methodology*, 2004, 34:193–239.
- Schonlau M, Liebau E. Respondent driven sampling (August 2010). DIW Berlin Discussion Paper No.1048. Available at: http://papers.ssrn.com/sol3/papers. cfm?abstract\_id=1679748, accessed 27 January 2014.
- Volz E, Heckathorn DD. Probability-based estimation theory for respondent-driven sampling. *Journal of Official Statistics*, 2008, 24(Suppl. 1):79–97.

# 2

Using NetDraw for interpreting RDS data and for problem solving

# 2. Using NetDraw for interpreting RDS data and for problem solving

# **Getting started with NetDraw**

#### Introduction

NetDraw<sup>1</sup> is free software for graphing network data. It is commonly used with RDS data to graph recruitment chains. Because RDS data includes information about who recruited whom, we can produce graphs of recruitment chains.

To download NetDraw software go to the Analytic Technologies web site at:

https://sites.google.com/site/netdrawsoftware/home

- Click on "Download"
- Select either "Run/Open" or "Save"
- Double click on the installation file. The installation procedure will then begin. Just follow the prompts (you can just accept all the default choices).

# **Prepare data for NetDraw**

Preparation of data for NetDraw application use can be done in RDS Analyst (www. hpmrg.org) or RDSAT (www.respondentdrivensampling.org).

#### In RDS Analyst

- 1. Run data analysis in RDS Analyst (see RDS Analyst manual at www.hpmrg.org on how to format a dataset).
- 2. To save your data in a DL format, go to the RDS Analyst Console and click File|Save Data| (Figure 2.1).

<sup>&</sup>lt;sup>1</sup> Borgatti, SP. *NetDraw: Graph visualization software*. Harvard, Analytic Technologies, 2002.

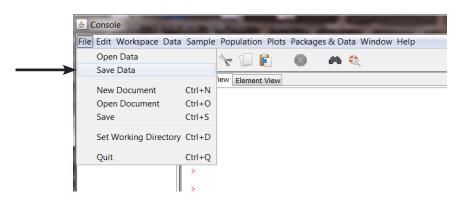


Figure 2.1 RDS Analyst: Console

3. A box will open asking you which dataset you want to save (if you are using multiple datasets at once, select the dataset you want) (Figure 2.2).

🛓 Console	of the local division in which the local division in the local div	and the second se
File Edit Workspace Data	Sample Popula	ation Plots Packages & Data Window Help
2 🥥 🖬 🖣	→ ~ [	🗆 🛍 🚳 🎮 🍳
• Balti\$recruiter.id <- get.recruite FSW.Demo.RDS.Analyst\$recr	Console View Ele	ement View
write.rdsobj(FSW.Demo.RDS.	>	
	>	
	>	🛓 Save RDS Data Set
	>	RDS Data
	>	FSW.Demo.RDS.An
	>	OK Cancel

#### Figure 2.2 RDS Analyst: Saving a file

4. Select the folder in which to place the file(s) and add the file name (Figure 2.3)

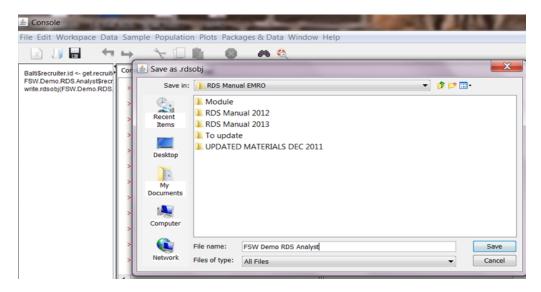


Figure 2.3 RDS Analyst: Naming the file and adding to folder

5. The program will automatically create three files: the DL file, the VNA file (see section on setting up the attribute file on page 52 below) and the RDSOBJ (RDS Analyst modified file) (Figure 2.4).

Organize 👻 📄 Op	en Share with 🕶 Burn New folder		
★ Favorites	Name	Date modified	Туре
🐌 Downloads 🚊	FSW DEMO RDS Analyst.DL	3/27/2013 9:32 AM	DL File
👢 Dropbox 👘	FSW DEMO RDS Analyst.rdsobj	3/27/2013 9:32 AM	RDSOBJ File
딇 Recent Place	📴 FSW Demo RDS Analyst	2/26/2013 2:59 PM	SPSS Statistics Dat.
💻 Desktop	FSW DEMO RDS Analyst	3/27/2013 9:32 AM	vna Document
Documents Libraries	🖺 FSW Demo RDS Analyst	3/27/2013 9:30 AM	Microsoft Excel W

Figure 2.4 RDS Analyst: DL, VNA and RDSOBJ files saved

#### **In RDSAT**

- I. Run data analysis in RDSAT (see RDSAT step-by-step manual on how to format a dataset at www.respondentdrivensampling.org).
- 2. To save your data in a DL format, go to File|Export DL Network File| (see Figure 2.5).
- 3. Change the name of your file by adding DL to the end of the file name.
- 4. Save in the folder of your choice.

File Analyze Help								
New RDS View/Edit RDS Save RDS Analysis	ACT	SILINODCIE	BASES'R	EPORT V	ARIABLES FOR NETDRAW.txt	\$	Dipen New RDS	Analyze Parl
Print					Reload			
Export DL Network File							Add Data	Analyze Brea
Export Population Weights						1	Edit Data	Change Opt
Export Individualized Weight Export Estimation Table	nts							
Export Estimation Table Export Table of Recruitmer						۲		
Options	Size	s and Hom	ophily Gr	aphics ar	nd Histograms			
Exit	nt h	y HIV						
(Rec	ruitment							
	ition Prob							
Person who Recruite					Data-Smoothed Recruitments:			
	Group O	Group 1	Group 2	2 Total	Group 0 Group 1 Group 2			
Group O	136.0	85.0	5.0	226.0	Group 0 131.514 80.233 6.798			
	0.602	0.376	0.022	1.0	Group 1 80.233 83.705 9.995			
Group 1	72.0	77.0	11.0	160.0	Group 2 6.798 9.995 0.73			
	0.45	0.481	0.069	1.0				
Group 2	12.0	11.0	1.0	24.0				
	0.5	0.458	0.042	1.0				
					Data-Smoothed Transition			
Transition Proba					Probabilities:			
Group O Gro	_	_			Group 0 Group 1 Group 2			
Group 0 0.602 0.3					Group 0 0.602 0.367 0.031			
	81 0.06				Group 1 0.461 0.481 0.057			
Group 2 0.5 0.4	58 0.04	42			Group 2 0.388 0.57 0.042			
Demographically	Adjusted				Sample Pop. Initial			
Recruitment I					Sizes: Recruits:			
Group 0 Gro		<b>р</b> 2			Group 0 222 Group 0 2			
Group 0 131.514 82.		-						

Figure 2.5 RDSAT analysis page

#### **DL** files

The DL protocol is a flexible language for describing data. Your dataset in a DL file will look something like Figure 2.6.

REPO	ORT	VARIA	ABLE	s fo	r ne	TD	RAV	V DI	L - N	lote	epao	ł																														-		2	<b>K</b>
Eile         E           DL         n           596063         11411           115515         19619           23723         27827           319324         360366           401400         0         0           0         0         0         0           0         0         0         0           0         0         0         0           0         0         0         0           0         0         0         0	= 4: 162 511 615 719 823 928 032: 136 240 0 0 0 0 0 0 0 0 0	141a 6364 6117 7158 8199 9240 0281 1322 2363 3404 0 0 0 0 0 0 0 0	bel 656 118 200 241 282 323 364 405 0 0 0 0 0 0	s:1 667 119 160 201 242 283 324 365 406 0 0 0 0	234 686 120 202 243 284 325 366 407 0 0 0 0 0 0	56 597 012 16 220 324 28 532 536 740 0 0 0 0 0 0 0 0	789 071 112 216 320 424 528 632 736 840 0 0 0 0 0	172 221 531 542 452 273 583 094 0 0 0 0	737 231 641 052 462 872 283 693 104 0 ( 0 ( 0 ( 0 ( 0 ( 0 ( 0 (	747 124 165 206 247 288 329 370 370 370 0 0 0 0 0 0 0 0 0 0 0 0 0 0	57 12 520 520 520 520 520 520 520 52	677 512 616 720 824 929 033 137 241 0 0 0 0 0 0	78 61 92 92 902 13 23 13 23 134 0 0 0 0 0	798 271 681 092 502 912 323 733 14d 0 0 0 0 0 0 0 0 0 0	08: 28: 69: 10: 51: 92: 33: 74: 0 0 0 0 0 0	L82 L29 L70 211 252 293 334 375 a:0 0 0 0 0	838 130 171 212 253 294 335 376 0 0 0 0 0 0 0 0	48: 13: 254 29: 330 37: 0 ( 0 0 0 0 0 0	586 113 217 321 425 529 533 737 0 0 0 0 0 0	878 213 317 421 525 629 733 837 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	88 31 52 52 62 72 83 79 83 79 83 79 83 79 83 70 93 70 70 70 70 70 70 70 70 70 70 70 70 70	990 341 751 162 572 982 393 803 0 0 0 0 0 0 0 0	919 351 761 172 582 993 403 813 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	293 361 771 2182 592 300 341 382 0 ( ) 0 ) 0 ) 0 ) 0	394 L37 L78 219 260 301 342 383 0 1 0 0 0 0	959 138 179 220 261 302 343 384 1 0 0 0 0 0 0 0 0	697 139 180 221 262 303 344 385 0 0 0 0 0 0 0 0 0	(98) (14( )18] (222) (26) (304) (34) (34) (38( ) 0 (0 (0 (0 (0) (0) (0) (0) (0) (0) (0) (	991 014 118 222 326 430 534 534 538 0 0 0 0 0 0 0 0 0 0 0 0 0	001 114 218 322 426 530 634 738 0 0 0 0 0 0 0 0	011 214 318 422 526 630 734 838 0 0 0 0 0 0 0 0 0 0 0 0 0	021 314 418 522 626 730 834 939 0 ( 0 0 0 0 0 0	L031 4414 3518 2622 5726 0830 4935 0039 0039 0039 0039 0039 0039 0039 0	04: 514 8618 2722 5826 933 503 9139 0 ( 0 0 0 0 0 0 0 0 0 0 0 0	105 461 871 282 592 103 513 923 00 0 0 0 0 0	106 471 881 292 702 11 52 933 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	5107 481 302 712 53 94 0 ( ) 0 0 ( ) 0	710 L49 L90 231 272 313 354 395 0 0 0 0 0 0	810 150 232 273 314 355 396 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	91: 15: 19: 23: 27: 31: 35: 39: 0 0 0 0 0 0 0	101 115 219 323 427 531 635 739 0 0 0 0 0 0 0 0	111 215 319 423 527 631 735	121 31 419 52 627 731 83	L13 54 95 36 77 L8 59	•
$\begin{smallmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 &$				000000000000000000000000000000000000000		) 0 ) 0 ) 0 ) 0 ) 0 ) 0 ) 0 ) 0 ) 0 ) 0	000000000000000000000000000000000000000	000000000000000000000000000000000000000	0 ( 1 1 0 ( 0 ( 0 ( 0 ( 0 ( 0 ( 0 ( 0 (		) 0 ) 0 ) 0 ) 0 ) 0 ) 0 ) 0 ) 0 ) 0 ) 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0			000000000000000000000000000000000000000			000000000000000000000000000000000000000			000000000000000000000000000000000000000			000000000000000000000000000000000000000				0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	000000000000000000000000000000000000000				) 0 ) 0 ) 0 ) 0 ) 0 ) 0 ) 0 ) 0 ) 0 ) 0	000000000000000000000000000000000000000			000000000000000000000000000000000000000					0 0 0 0 0 0 0	0 0 0	
$\begin{smallmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 &$		0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0					000000000000000000000000000000000000000	000000000000000000000000000000000000000	000000000000000000000000000000000000000			000000000000000000000000000000000000000	000000000000000000000000000000000000000	0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0	000000000000000000000000000000000000000	00000000100			000000000000000000000000000000000000000			000000000000000000000000000000000000000			000000000000000000000000000000000000000		000000000000000000000000000000000000000								000000000000000000000000000000000000000									000000000000000000000000000000000000000	



Where:

DL is the format of the file

N =sample size (414)

Labels: if not modified the NetDraw software will assign a number as label to every node

Data: actual data

Your data are now ready for use in NetDraw.

# **Generating recruitment chains**

Double click on the NetDraw icon to launch application. The NetDraw screen is shown in Figure 2.7.



Figure 2.7 NetDraw screen

#### Loading the file into Netdraw

To load your file, select the newly created DL file using the browser. Click on File|Open |Ucinet DL text file| (see Figure 2.8). Select appropriate network type ("Network 1-mode"). The screen in Figure 2.9 will appear. Click "OK".

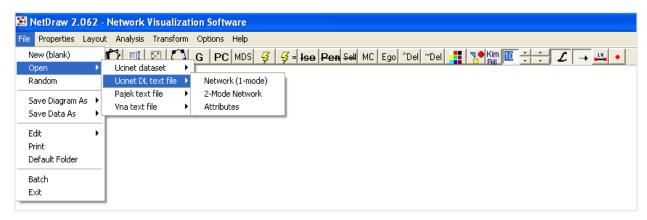
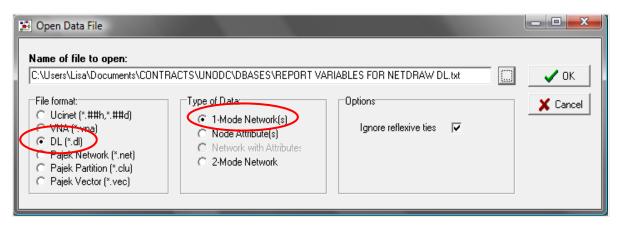


Figure 2.8 NetDraw screen file drop-down box



#### Figure 2.9 Open data file in NetDraw

Upon clicking "OK", a graphic looking something like that displayed in Figure 2.10 or Figure 2.11 will appear. This is a chain made up of 414 nodes. Sometimes the graphic will appear as several isolated nodes. The numbers next to nodes indicate 1D numbers generated by NetDraw. In the right lower corner of the screen are check boxes for nodes attributes.

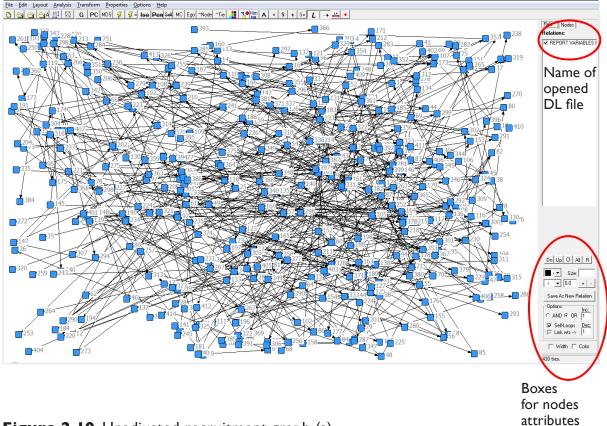
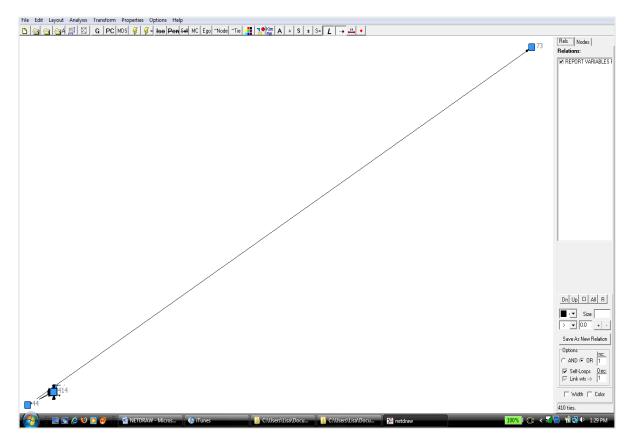


Figure 2.10 Unadjusted recruitment graph (a)



**Figure 2.11** Unadjusted recruitment graph (b)

#### Graphing the recruitment chain

To graph the recruitment chains, Click on Layout|Graph-Theoretic layout|Spring embedding| as shown in Figure 2.12.

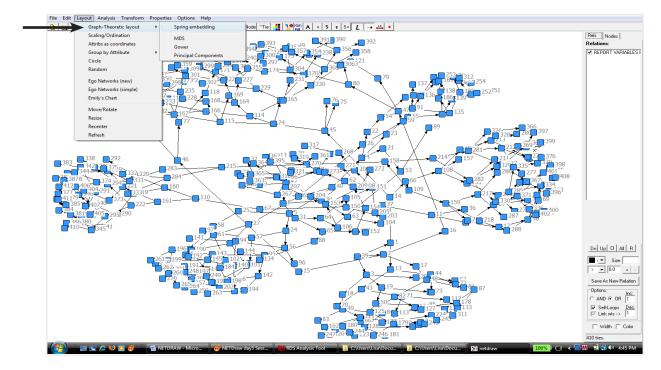
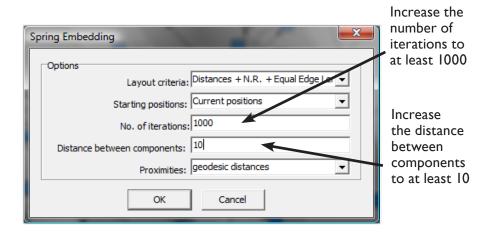


Figure 2.12 Layout

The "Spring Embedding" pop-up box appears (see Figure 2.13). In the menu the number of iterations needs to be increased to at least 1000 and the distance between components needs to be increased to at least 10, but you can play around with it and increase it to 13, 14 or higher. These values will affect how the graph looks. You may have to experiment with the options to see which produce the best display for your data.

You can also use the 😼 icon along the top of the display page to make more iterations of the recruitment chains.



#### Figure 2.13 Spring Embedding pop-up box

#### What to look for in the graph

Once the graph has been organized, it should portray the following information (see Figure 2.14):

- Node numbers should match up to the respondent's coupon number in the original dataset.
- There should be the same number of nodes as the sample size. This graph has 414 nodes which is the same as the sample size.
- On the bottom right, there is a display showing the number of ties. This graph has 410 ties (sample size minus the four seeds).
- The number of recruitment clusters should be the same as the number of seeds. This graph has four recruitment clusters, which is equivalent to four seeds.
- Isolated nodes in the upper right hand corner display seeds that have not recruited anyone or are nodes that have been incorrectly numbered. All of the seeds for this sample recruited others and there were no incorrectly numbered participants. If an independent node appears that is not a seed, it will be important to verify why that node is not connected. This may happen if the recording of the identification of who recruited whom is incorrect. To correct this, use the NetDraw recruitment graph to determine which number node is not connected and compare that to the sequential identification numbers (column 2) in the original RDSAT Excel file to find and correct the recruiter's coupon number.
- Correct file name appears on the relations box.

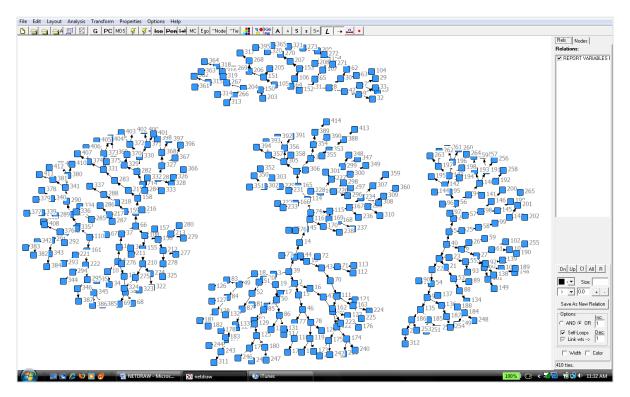
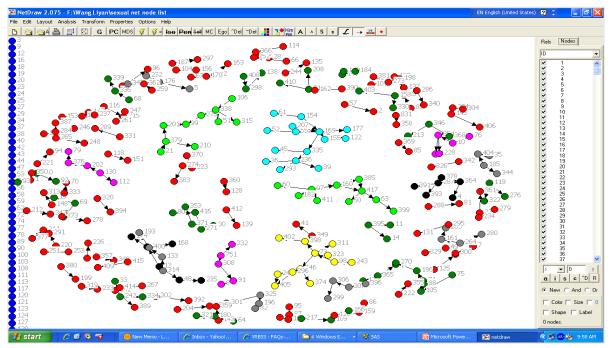
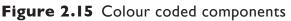


Figure 2.14 A recruitment graph after spring embedding<sup>1</sup>

#### **Colour code components**

It is also helpful to be able to see the different recruitment chains. This can be done by using different colours for each network (called components in NetDraw). Click on Analysis menu|Components|Select color|OK on the components dialog box. Once you do this each cluster will have a different colour (see Figure 2.15).





 $<sup>\</sup>overline{}^{1}$  For this file the number of iterations was set to 1000 and the distance between components was set to 16.

The attribute file contains information on the attributes of each case in the data and will allow you to highlight differences (by colour, shape, size, etc.) in the recruitment chains based on the attributes or variables (e.g. HIV positive can be a green colour and shaped as a triangle and HIV negative can be a blue circle).

#### Setting up the attribute file in RDS Analyst

As described already, prepare data for NetDraw. RDS Analyst automatically creates three files when you save your file from the program. You will need the VNA file for making graphics using attributes (Figure 2.16).

Organize 👻 📄 Op	oen Share with ▼ Burn New folder		
🚖 Favorites 🗂	Name	Date modified	Туре
🐌 Downloads 😑	FSW DEMO RDS Analyst.DL	3/27/2013 9:32 AM	DL File
👢 Dropbox	FSW DEMO RDS Analyst.rdsobj	3/27/2013 9:32 AM	RDSOBJ File
😹 Recent Place	🕞 FSW Demo RDS Analyst	2/26/2013 2:59 PM	SPSS Statistics Dat.
🕅 Desktop	FSW DEMO RDS Analyst	3/27/2013 9:32 AM	vna Document
Documents	🕙 FSW Demo RDS Analyst	3/27/2013 9:30 AM	Microsoft Excel W
Libraries			

Figure 2.16 RDS Analyst: DL, VNA and RDSOBJ files saved

#### Setting up the attribute file in RDSAT

Setting up the attribute file in RDSAT requires it to be modified manually from the original RDS data file (see Figure 2.17a and 2.17b). To create the attribute file in RDSAT:

- I. Open the original RDS data file.
- 2. Replace the cell for RDS (cell A1 in Excel) with "\*node data".
- 3. Replace the sample size (cell A2 in Excel) with "ID".
- 4. Replace the number of coupons cell (B2 in Excel) with "Degree". Degree should now be above the column of participants' social network size.
- 5. Delete the columns of coupon numbers (this will include the participants' coupons and the coupons given to participants to use in recruiting peers; for instance in this survey a maximum of three coupons were used so that columns C, D, E and F are removed).
- 6. Save the file as a "Tab delimited text file". Be careful not to overwrite your original RDS files.
- 7. Close the text file.

$\sim$	Home	Insert F	Page Layout	Formulas	Data R	eview Vi	ew Nitro	D PDF	0.	- =
Norm	nal Page Layout			Rew Wi	All		Save orkspace	Switc	vs -	Macros Macros
	А	В	С	D	E	F	G	н	1	J
1 F	RDS									
2	414	3	999				LAB	HIV	HCV	HBV
3	1	5	21	211	212		R1-008	1	1	0
4	2	20	22	221	222		R1-009	0		~
5	3	10	23	231	232		R1-007	0		-
6	4	100	31	311	312		R1-006	1		-
7	5	50	32	321	322		R1-010	1		-
8	6	30	33	331	332		R1-065	0		-
9	7	12	51	511	512		R2-004	0		-
10	8	4	52	521	522		R2-007	0		-
11	9	150	53	531	532		R2-006	0		-
12	10	50	61	611	612		R2-016	1		-
13	11	19	62	621	622		R2-002	0	-	
14	12	10	63	631	632		R2-015	0	-	-
15	13	100	211	2111	2112		R1-011	0		-
16	14	50	213	2131	2132		R1-012	0		-
17	15	2	221	2211	2212 2222		R1-028	1		-
18	16 17	45	222	2221	2222		R1-027 R1-024	1		
14 4				26 09 DBA	S She		N LIT M			•
Read							0% 🗩		<b>U</b>	(+

	Home	V CH 🔹 🔣 Insert Pa		Formulas		RAW [Con view Vie				
	mal Page Layout			Rew Win	All		Save orkspace W	Switch	s *	lacros lacros
	А	В	С	D	E	F	G	н	1	J
	*node data									
	ID	3	999	011	040	042	LAB	_	HCV	
3	2	5 20	21 22	211	212 222		R1-008	1	1	0
4	2		22	221 231	222		R1-009 R1-007	0	1	0
6	3		31	311	312		R1-007	1	1	0
7	5		32	321	312		R1-000	1	1	0
8	6		33	331	332		R1-065	o o	1	Ö
9	7		51	511	512		R2-004	ŏ	1	Č
10	8		52	521	522	523	R2-007	0	1	0
11	9	150	53	531	532	533	R2-006	0	1	0
12	10	50	61	611	612	613	R2-016	1	1	0
13	11	19	62	621	622	623	R2-002	0	1	0
14	12	10	63	631	632	633	R2-015	0	1	0
15	13		211	2111	2112		R1-011	0	1	0
16	14		213	2131	2132		R1-012	0	1	0
17	15		221	2211	2212		R1-028	1	1	0
18	16		222	2221	2222		R1-027	1	1	0
19	17 ▶ ▶ REF		223 ABLES DEC	2231 26 09 DBAS	2232 Sheet	2233 4 IIII	R1-024	1	1	
Rea			DELO DEO	20 05 0000			0% 🖃 —	U	1	-+

Figure 2.17a Attribute file set-up I

9		~~	REPOR Layout	RT VARIABLES		RAW [Com			- I		
	mal Page Layout	Show/Hide	Zoom	Rew Win	All 🔤	山 山 記 田 Wo Window	Save S rkspace Wir	witch	N	lacros	
		<b>D</b>		2			6				
	A	В	С	D	E	F	G	н		J	
1	*node data		000					1.10.7	11014		
2	ID 1	3	999 21	211	212	242	LAB R1-008	HIV 1	HCV 1	HBV	
4	2	20	21	211	212		R1-008	0	1	0	
5	3	10	22	221	232		R1-005	0	1	0	
6	4	100	31	311	312		R1-006	1	1	0	
7	5	50	32	321	322		R1-010	1	1	0	
8	6	30	33	331	332		R1-065	0	1	Ő	
9	7	12	51	511	512		R2-004	0	1	0	
10	8	4	52	521	522	523	R2-007	0	1	0	
11	9	150	53	531	532	533	R2-006	0	1	0	
12	10	50	61	611	612	613	R2-016	1	1	0	
13	11	19	62	621	622	623	R2-002	0	1	0	
14	12	10	63	631	632	633	R2-015	0	1	0	
15	13	100	211	2111	2112	2113	R1-011	0	1	0	
16	14	50	213	2131	2132		R1-012	0	1	0	
17	15	2	221	2211	2212		R1-028	1	1	0	
18	16	45	222	2221	2222		R1-027	1	1	0	
19	17 REPOR		223	2031	2232		R1_024	1	1		-
Rea		I VANLADI	ES DEC	20 09 DBAS	Share 1		» (=)		]	-+	.:
	-,		T	T							
						/					

Delete these columns

	8	Home		⇒ REPO	_	RIABL	ES FOR Data	_	AW [Comp	atibility Nitro PDF		×
						New V Arrang Freeze	ge All	-	다. Sa 관련 Work: Vindow	ve Swi space Wind	tch ows + Mac	ros
		А	в	С	D	E	F	G	Н	1	J	
	1	*node data	<u> </u>									E
	2		Degree	LAB	HIV	HCV			REAPART	FEMNET	Q103AGE	SE
	3	1		R1-008	1	1	0	0	2	0	32	
	4	2		R1-009	0	1	0	0	2	10	41	
	5	3		R1-007	0	1	0	0	2	0	39	
	6	4		R1-006	1	1	0	0	2	12	42	
	7	5		R1-010	1	1	0	0	6	3	50	
	8	6		R1-065	0	1	0	0	2	1	27	
	9	7		R2-004	0	1	0	0	8	4	19	
	LO	8		R2-007	0	1	0	0	9	0	32	
	11	9 10		R2-006 R2-016	0	1	0	0	2	2	27	
	L2 L3	10		R2-016 R2-002	0	1	0	0	2	9	26	
	L3 L4	11		R2-002 R2-015	0	1	0	0	6	2	20	
	L4 L5	12		R1-011	0	1	0	0	6	2	36	
	16	13		R1-012	0	1	0	0	6	4	36	
	LO L7	14		R1-012	1	1	0	1	6	2	53	
	18	16		R1-020	1	1	0	0	6	5	37	
	19	10		R1-027	1	1	0	0		1		
	•			ABLES DEC		09 DB		Shee				
1	Rea	dy Scroll Lo	ock						100%			÷.,

Figure 2.17b Attribute file set-up 2

#### Importing the attribute file

Once your attribute file is created, import it into NetDraw by doing the following: click File|Open |Vna text file|Attributes| (See Figures 2.18 and 2.19). Use the pop-up browser to find and open the attribute file. Before clicking "OK", make sure the file format is set to "VNA(\*.vna)" and type of data to "Node Attributes".

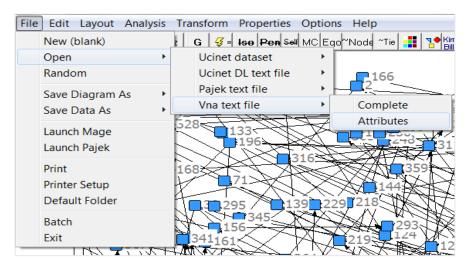


Figure 2.18 Opening an attribute file in NetDraw (a)

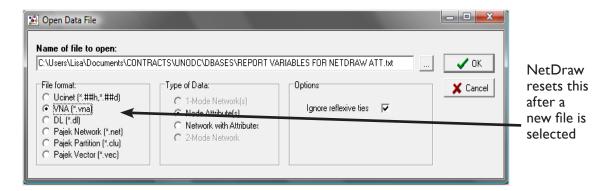
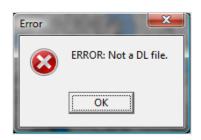


Figure 2.19 Opening an attribute file in NetDraw (b)

If you get the error message (see Figure 2.20), you will need to import the attribute file again. Make sure the VNA (\*.vna) file format is selected. This should be the last thing you do before clicking "OK". Another reason you might get this error box is if RDSAT is still open or if your original attribute file database (in Excel) is still open. Make sure you close these before opening the attribute file in NetDraw.



#### Figure 2.20 Error message

If the attribute file is successfully downloaded into NetDraw, the box in Figure 2.21 ("Node Attribute Editor") may appear, depending on your version of NetDraw. The "InternalID" column (column 1) should match the "ID" column (column 2). If the "InternalID" and the "ID" columns match, then close this box and the recruitment chains will appear. If you do not get this box, do not worry, just continue to follow the instructions below. You may have to repeat the spring embedding settings or click the icon in Figure 2.22 to get distinct clusters of recruitment chains.

<u>F</u> ile <u>E</u> dit						
InternalID	ID	Degree	LAB	HIV	HCV	
1	1	5	R1-008	1	1	
2	2	20	R1-009	0	1	
3	3	10	R1-007	0	1	
4	4	100	R1-006	1	1	
5	5	50	R1-010	1	1	
6	6	30	R1-065	0	1	
7	7	12	R2-004	0	1	
8	8	4	R2-007	0	1	
9	9	150	R2-006	0	1	
10	10	50	R2-016	1	1	
11	11	19	R2-002	0	1	
12	12	10	R2-015	0	1	
13	13	100	R1-011	0	1	
14	14	50	R1-012	0	1	

Figure 2.21 Node Attribute Editor box

#### Figure 2.22 Recruitment train icon

#### **Moving nodes**

Selected nodes can be dragged to other parts of the graph by clicking on a particular node and then dragging it (see Figure 2.23). You will know if the node has been properly selected when it turns from a solid colour to a hashed colour. If multiple nodes are selected and one is moved, the rest will follow, maintaining their structure or you can create a highlight box (Note: highlight box only works starting at the upper left corner) and dragging the nodes into the box.

**₹** 

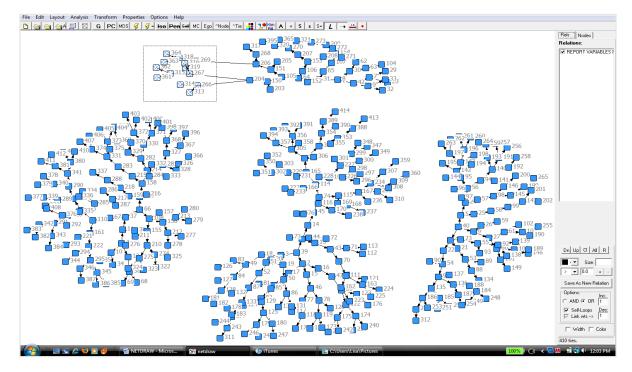


Figure 2.23 Selecting nodes and dragging them

#### Creating graphs with the attribute file

To create graphs with the attribute file, you will have to use the "Properties" menu. The Properties menu will allow you to change nodes, lines and background sizes, shapes and colours. There are two types of options:

- 1. "General" option: this assigns the same sizes, shapes and colours to all visible object in a category.
- 2. "Attribute-based" option: this assigns properties by attribute.

To select the General or Attribute-based options for colour click Properties Nodes Symbols [Color]General-all active nodes] or [Attribute-based] (see Figure 2.24).

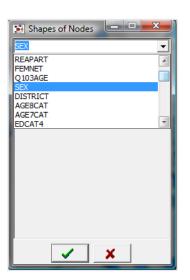


Figure 2.24 Selecting general or attribute-based options

#### Attribute-based options

This section will provide some examples of using the attribute-based node properties. To select a shape to distinguish nodes by attributes, click Properties|Nodes|Symbols|Shape| Attribute-based. A box entitled "Shapes of Nodes" will appear. Use the drop down menu to select a variable (see Figure 2.25). The variable selected for this example is "Sex" (box 1) and the shapes selected are "Circle" for 1 (males) and a "Square" for 2 (females). There are several shapes to choose from. After selecting the shapes, click the box with the green check symbol.

The same procedure can be used to set node colour and size.



Box I

Figure 2.25 Shapes of nodes boxes

Box 2

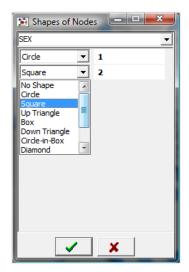


Figure 2.26 shows one of the four recruitment chains with squares and circles. Nodes are shape coded for sex.

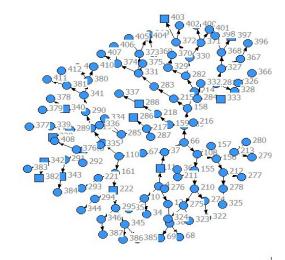


Figure 2.26 Shape coded recruitment chain

#### Nodes tab

The right side of the NetDraw display window has a series of useful tabs.

- The "Rels" and "Ties" tabs are not used for RDS data.
- The "Nodes" tab allows the selection of visible nodes by variable value (upper part of box). To locate a variable in the dataset, use the down arrow (see Figure 2.27). In this example the variable "SEX" is selected.
- By unchecking the box all nodes matching that value are no longer visible (Figure 2.28). In this example, value 1 for the variable SEX is being selected. Only value 2 remains in the recruitment chains.

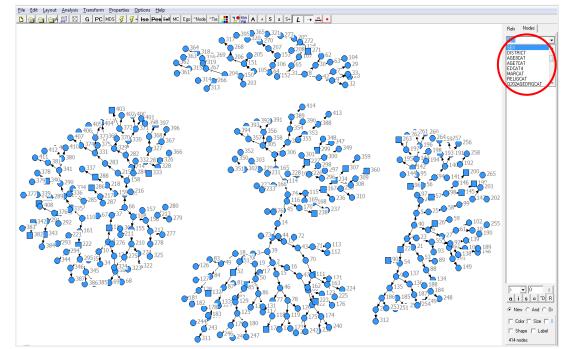


Figure 2.27 Displaying nodes by variable values (a)

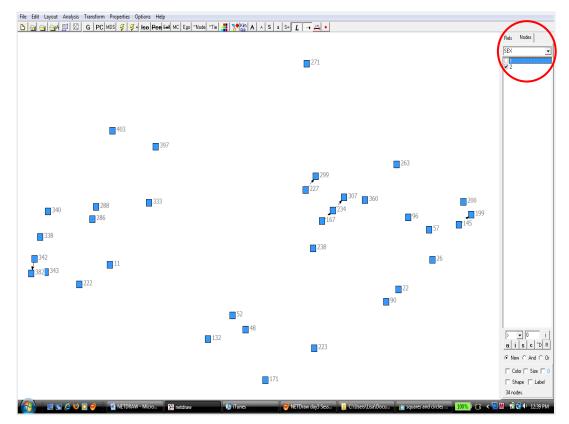


Figure 2.28 Displaying nodes by variable values (b)

- On the lower part of the right hand side of the NetDraw display window are six "short-cut" buttons that are useful for manipulating visible and invisible nodes (see Figure 2.29). These buttons are:
  - **a** All categories—nodes visible
  - i Switch—checked categories are visible and unchecked categories are hidden
  - **s** Next category—selects the next category and unselects the current category
  - **c** Add next—adds the next category to the selection
  - ^D Permanently delete—permanently deletes all hidden nodes
  - R Return—returns to the original respondent ID variable.
- The bottom of the right hand side of the NewDraw display window has boxes that can be checked for defaulted colour, size, shape and label.

> • 0 ! a i s c ^D R
⊙ New ○ And ○ Or
🗖 Color 🗖 Size 🔲 0
🗖 Shape 🔲 Label
34 nodes.

Figure 2.29 Short-cut buttons

#### **Displaying labels**

To get rid of node labels click Properties|Nodes|Labels|Visible and the "Node Labels" box in Figure 2.30 should appear. To get rid of labels (e.g. ID numbers), click "Off" under "Turn labels..." (left side of box). You can apply this to all nodes or only active nodes (right side of box). An easier way to do this is to use the icons at the top of the NetDraw display window. The icon "L" allows labels to be turned on and off.

🛃 Node Labels							
Turn labels	Apply to:						
C Off	<ul> <li>Only active nodes</li> </ul>						

#### Figure 2.30 Node Labels box

To label nodes by variable value, use the box for "Label" at the bottom of right-hand side of the NetDraw display window. By selecting the variable in the "Nodes" box (upper right hand side of the NetDraw display window) and then checking the "Label" box at the bottom, the nodes will be labelled according to the values given to each category.

To assign text labels, it is possible to click Properties|Nodes|Labels|Text| which will display the "Label Text" box (see Figure 2.31). Under the "Label" column it is possible to put in a text label such as "Male" or "Female" for the variable "Sex", corresponding to the appropriate value (males for 1 and females for 2).

🖼 Label Text 📃 💻 💌							
Node Attribute to Use as Labels:							
SEX 💌							
For numeric variables, number of decimal places: 0							
	Label						
1	1						
2	1						
3	1						
4	1						
5	1						
6	1						
7	1						
8	1						
9	1						
10	1						
VIC Cancel Apply							

An easier way to do this may be to go back to your original database and copy and paste the sex column into a new column (see "SEXTXT", in Figure 2.32) and "find" the values (e.g. 1) and "replace" them with the corresponding text (e.g. "Male ") (Figure 2.32).

		- (° - 📉							M 💶 🗖	
	Home	Insert	Page Layo	ut Form		Data Revi	ew View	/ Nitro P	DF 🕜 –	■ X
1	<u> </u>	Arial	- 10 -	= = ;		Number	- A	🗧 🗝 Inser	t• Σ•	<u>A</u> γ-
		BIU	A A	= = :	≣ 🔤 -	\$ - %	, _	🔁 🄁 Delet	te - 🗔 -	<b>#A</b> -
Pa	ste 🧹	- 😓 -	<u>A</u> -		≫~-	€.0 .00 .00 →.0	Styles	Form		
Clint	board 🖻	Font		Alignm	-	Number	G I	Cells		ing
Cip										
	С	D	E	F	G	Н		J	K	L 🔺
1								-		=
2	999				LAB	Q103AGE		SEXTXT	DISTRICT	
3	21		212		R1-008	32		Male	3	
4	22		222		R1-009	41		Male	1	
5	23 31		232		R1-007	39 42		Male	1	
6	31		312 322		R1-006 R1-010	42		Male Male	4	
8	33		332		R1-010	27		Male	4	
9	51		512		R1-065 R2-004	19		Male	5	
10	51		512		R2-004	32		Male	5	
11	53		532		R2-007	27		Male	5	
12	61		612		R2-016	32		Male	7	
13	62		622		R2-002	26		Female	7	
14	63		632		R2-015	33	_	Male	7	
15	211		2112		R1-011	36		Male	1	
16	213		2132		R1-012	36	1	Male	2	
17	221		2212		R1-028	53		Male	1	
18	222	2221	2222	2223	R1-027	37	1	Male	1	
19	223	2231	2232	2233	R1-024	31	1	Male	1	
20	231	2311	2312	2313	R1-030	40	1	Male	2	
21	232		2322		R1-017	42	1	Male	2	
22	233		2332		R1-046	42		Male	1	
23	311		3112		R1-015	43		Male	1	
24	312		3122		R1-016	35		Female	1	
25	313		3132		R1-014	43		Male	1	
26	321		3212		R1-035	42		Male	1	× 1
H 4 > H REPORT VARIABLES DEC 26 09 DBAS Sheet2 1 4										
Rea	dy							0% 🖃 —		-+ .::

#### Figure 2.32 Creating text labels in Excel

If text is created after already downloading your DL and Attribute files into NetDraw, it will be necessary to create and reload the new files with the text coding into NetDraw. Once the DL and then the attribute files (NetDraw must read the DL file before it can read the attribute file) with the text is loaded into NetDraw:

- go to the upper right-hand side of the display window and select the variable (in this example the variable is "SEXTXT")
- go to the lower right-hand side of the display window and select the label box.

The labels should appear next to the appropriate nodes as seen in Figure 2.33.

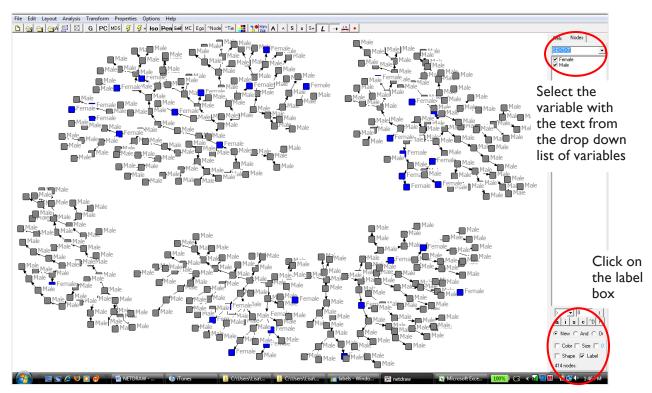


Figure 2.33 Recruitment chains with text labels

#### **Finding seeds**

To find seeds in the recruitment graphs it is necessary to look in the original database to find the IDs connected with the seeds. In the database below (see Figure 2.34) the four seeds have ID numbers 40, 41, 42 and 43 (use the ID numbers assigned by Excel in the far left column [not column A] as these will be the same ID numbers in the attribute file).

9		~ (° • 🐼	⇒ REPOI	RT VARIABL	ES FOR NE	TDRAW2 [		x
	Home I	insert Page	Layoi Formu	ulas Data	Review V	iew Nitro P	DF 🕜 🗕 🗖	<sup>3</sup> X
No	rmal Page	II Show/I		Window	Macros			
W	orkbook Viev			Ť	Macros			
	Α	в	C	D	E	F	G	
1	414	3	999	U	E .	F	DATE	
23	22	100	312	3121	312	2 3123	10/27/200	9
24	23	1000	313	3131	313		10/27/200	
25	24	2	321	3211	321		10/28/200	
26	25	50	322	3221	322		10/27/200	
27	26	35	331	3311	331	2 3313	10/30/200	19 F
28	27	500	332	3321	332	2 3323	10/30/200	19 F
29	28	40	511	5111	511	2 5113	11/4/200	19 F
30	29	50	512	5121	512	2 5123	11/4/200	9 F
31	30	20	521	5211	5212	2 5213	10/31/200	9 F
32	31	60	522	5221	522	2 5223	10/31/200	9 F
33	32	30	531	5311	531	2 5313	11/3/200	9 F
34	33	4	532	5321	532		11/3/200	
35	34	15	611	6111	611		11/23/200	
36	35	40	612	6121	612		11/23/200	
37	36	16	621	6211	621		11/3/200	
38	37	15	622	6221	622		10/30/200	
39	38	30	631	6311			11/24/200	
40	39	5	999	21			10/26/200	
41	40	285	999	31	3		10/26/200	-
42	41	25	999	61	6		10/28/200	
43	42	17	999	51	5		10/28/200	
44	43 44	6 10	2111 2112	21111 21121	2111		10/27/200	
45	44	10	2112	21121 21311	2112		10/27/200	
40	45	45	2131	21311 22111	2131		10/27/200	
47			ABLES DEC			22113	10/20/200	
	Average: 2					% 😑	-0(	÷:

Figure 2.34 Seeds in the Excel file

Find the seeds by looking at ID numbers under the "Nodes" tab. Upon selecting the corresponding ID number, the seed will disappear. In Figure 2.35, the red circle indicates where the seed for ID 42 was located. The identified seed can be distinguished by right clicking the mouse and selecting one of the options. In Figure 2.35, seeds 40, 41 and 43 are distinguished by large size and different colour.

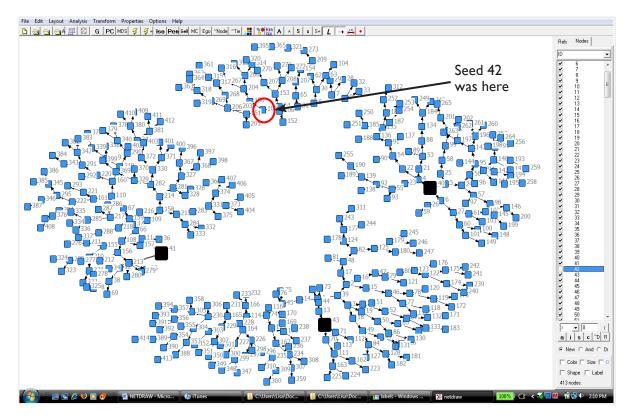


Figure 2.35 Finding and distinguishing seeds

#### **Making seeds with rims**

To change the node rim size and colour of the seeds, click "i" (one of the six buttons at the bottom of the right hand side of display screen) to uncheck all nodes (all nodes will disappear from the screen). Then check just the seed nodes (i.e. 40, 41, 42 and 43) and they will appear (see Figure 2.36). Then click Properties|Nodes|Rims|Color|General-all active nodes| and the "Rim Color" box will appear. Click a rim colour. In the example in Figure 2.37, the rim colour selected is "dLime".

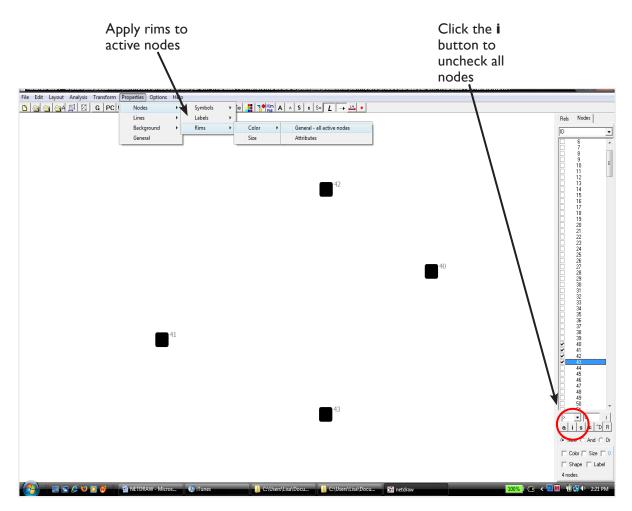


Figure 2.36 Isolating the seeds

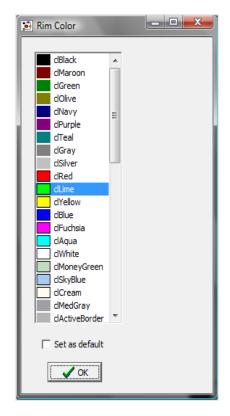


Figure 2.37 Selecting a rim colour to apply to seeds

#### **Exporting a graph**

To export a graph, click File|Save diagram as| then select the preferred picture format (Metafile, Bitmap, JPEG), all of which will produce the same image.

#### Saving your workspace

To save your workspace (the network and attribute data) so you can return to work with it at a later time without again importing the network and attribute data, select "VNA complete \*". Note that if you have deleted nodes they will not be saved.

Surveillance is the systematic, regular collection of information on the occurrence, distribution and trends of a specific infection, disease or other health-related event. HIV surveillance is designed to collect and integrate data reported from a variety of sources, including behavioural surveillance, case reporting, seroprevalence surveillance, and sexually transmitted infections surveillance. The goals of second generation HIV surveillance are to help countries better understand the HIV epidemic trends over time, to better understand the behaviours driving the epidemic, to focus on subpopulations at highest risk for infection and to better use surveillance data for planning the response to the epidemic.

HIV surveillance in the Eastern Mediterranean Region needs to be strengthened in order to fill the gaps in our understanding of the dynamics of the epidemic and to be in a better situation to plan appropriately for an effective response. This training course is part of a series of 4 training modules and has been adapted to the regional context from a module originally developed by the Centers for Disease Control and Prevention, USA. It describes how to plan and implement a respondent-driven sampling (RDS) survey. The supplement provides guidance on the use of the software RDS Analyst and NetDraw for analysis of RDS surveys. The course is intended for university and ministry of health staff, public and private public health researchers, and surveillance officers, who will be involved in the planning, organizing, monitoring or implementing of biological and/or behavioural surveillance surveys on HIV/AIDS and associated risk factors in key populations at higher risk of HIV exposure. Countries are free to further adapt these modules or to translate them into local use.