Install Instructions for NE4NS R Package

Kyle Vincent

[kyle.shane.vincent@gmail.com](mailto:kyle.shane.vincent@gmail.com)

The paper for which the estimation procedure this package implements is titled “New Estimates for Network Sampling”, authored by Steve Thompson, and can be found here: https://arxiv.org/pdf/2002.01350.pdf.

**Requires:**

1. A recent version of R. See here <https://www.r-project.org/>
2. A recent version of RStudio. See here https://www.rstudio.com/products/rstudio/download/
3. Microsoft Visual Studio 2017: See here <https://visualstudio.microsoft.com/vs/older-downloads/>
   1. Download Visual Studio Community 2017 (version 15.9)
4. The user should review the C file to ensure they have the proper C libraries installed.

**Instructions:**

1. Visit <https://github.com/KyleShaneVincent/NE4NS> and download the files via clicking on the green “Code” tab and Download Zip.
2. Extract the files to a local drive/path, such as the Desktop.
3. Open the …/NE4NS/R/NE4NS/NE4NS/NE4NS.Rproj R project file in RStudio.
4. In RStudio, click on the Build tab, and select Clean and Rebuild.
5. At the command line in R, enter library(NE4NS).

**Function:**

There is only function, namely Network\_Sample\_Estimates(). The parameters of the function and its corresponding default values, along with a description, are given as follows:

* t\_step = 1000; the length of the MCMC chain
* paad\_0 = 0.0167; The approximate proportion of the initial sample that should act as the resampled initial sample
* naad\_0 = 1; Leave as 1 (no description provided at this time)
* paad\_trace = 0.5; The tracing rate
* aadres\_eed\_design = 1; Choose 1 for random reseeds and 2 for reseeding with probability proportional to degree
* paad\_random = 0.01; Probability for reseeding
* naad\_target = 400; The targeted final sample size of the resample
* coupon\_s = 0; Leave as 0 (no description provided at this time)
* output\_file\_path = as.character("c:\\data\\output.txt"); the path and file name of the estimation output from implementing the resampling procedure
* node\_input\_file\_path = as.character("c:\\data\\nodes.txt"); the path and file name of the nodes-based information from the sample
* node\_fi\_output\_file\_path = as.character("c:\\data\\nodesfi.txt"); the path and file name of the resampling frequency of the sample respondents, can be used to get network weights
* edge\_input\_file\_path = as.character("c:\\data\\edges.txt"); the path and file name of the edges-based information from the sample

**An Example:**

After installation, load the NE4NS package from the R command line with

>lilbrary(NE4NS)

The NE4NS folder should have “nodes.txt” and “edges.txt” files. I found it was easiest to create a “data” folder within my C drive and then to place these files in this folder. Do be sure to familiarize yourself with the nodes and edges files as the data must be in this format to implement the estimation procedure; the nodes file contains artificial data for a hypothetical network sample and the edges file gives the corresponding directional links. At the current time, the nodes file can only contain the ids and three other variable observations. However, all that is needed are the sample weights and then estimates can be obtained with the `survey’ package.

Run the following function at the R command line

>Network\_Sample\_Estimates(t\_step = 1000, paad\_0 = 0.0167, naad\_0 = 1, paad\_trace = 0.5, aadres\_eed\_design = 1, paad\_random = 0.01, naad\_target = 400, coupon\_s = 0, output\_file\_path = as.character("c:\\data\\output.txt"), node\_input\_file\_path = as.character("c:\\data\\nodes.txt"), node\_fi\_output\_file\_path = as.character("c:\\data\\nodesfi.txt"),   
edge\_input\_file\_path = as.character("c:\\data\\edges.txt"))

You should be able to see some output that first summarizes the nodes.txt and edges.txt files, and then provides point estimates and confidence intervals for observations made on the “bug”, “degree”, and “concurrent” variables. You should also be able to view this output in the output.txt file. You can obtain information on the frequency at which the resampling procedure chose each sample respondent for the resample in the nodesfi.txt file.

Please note that the standard errors, as calculated with this program, may be incorrect at this time. My recommendation is to use the simple variance estimator as given in equation (7) of the paper. Here is one approach on how to do this: The NE4NS package can be used in combination with the ‘survey’ package in R to provide point estimates and standard errors. The following provides some R code that gives an example of how this can be achieved. The standard error calculations are based on the suggested simple variance estimator.

#Extract the information to construct the sample weights

>setwd("C:/data")

>sample\_info = read.table("nodesfi.txt", header=TRUE)

>head(sample\_info)

>sample\_info[,"fi"] = as.numeric(sample\_info[,"fi"])

>C = ncol(sample\_info)

>sample\_info[,C+1] = NA

>colnames(sample\_info)[C+1] = "network\_weight"

>sample\_info[,"network\_weight"] = (1/sample\_info[,"fi"]) / (sum(1/sample\_info[,"fi"]))

>colnames(sample\_info) = c("id", "bug", "degree", "concurrent", "fi", "network\_weight")

>head(sample\_info)

#Declare an unequal probability sampling design with weights based on the resampling procedure

>library(survey)

>netdesign = svydesign(id=~1, weights=~network\_weight, data=sample\_info)

>svymean(~bug, netdesign)

>svymean(~degree, netdesign)

>svymean(~concurrent, netdesign)

**Notes:**

1. The coupon ids must be in a numeric format.