SHAMP: Seattle HIV/AIDS Modeling Project

Jeanette Birnbaum

2018-12-20

Contents

Chapter 1

Overview

Note that the sections below are denoted with bold, not section headings, because collapse: section in the yaml will work for regular chapters but not the index.Rmd chapter.

To add custom chapters, simply create new .Rmd files in this folder and add their names in the desired point in the list of .Rmd files in _bookdown.yml.

The chapters included are:

Chapter	Content
Setup	Reads in the "Parameters.R" file that establishes the baseline variables by which the descriptive statistics will be stratified. Load packages and shows sessionInfo.
Data	Notes on (and/or links to other documents containing notes on) how the raw data were prepared for this analysis.
Ego Characteristics	Demographic and other characteristics of the egos.
Degree	Momentary degree (e.g., extant/active ties only) and concurrency of egos.
Mixing	Mixing matrices of egos and alters on nodal attributes
One-Times	Rates of one-time relationships
Edge Age	Age of active/extant ties
Decisions	Analytical decisions made after considering the descriptives, including model specification choices.
ERGMs	ERGM fits and diagnostics
Parameters	List of parameter values for modeling. See https://norcalbiostat.netlify.com/post/ fancifying-excel-tables-in-r-markdown-with-kable-extra/ for importing from an Excel sheet and maintaining hyperlinks.
Appendix	Any appendices
Tables & Figures	Hyperlinked index of tables and figures, by caption
References	Reference list

Chapter 2

Data

Prepped rda data objects are available in `SHAMP/egonet/data`. They can be loaded after running `library(egonet)`, e.g. `data(nsfg.obj_c)` loads the cohab egodata object.

2.1   KC population weights

- See `egonet/inst/extdata/kc_ipums` for information on how ACS-5yr estimates for King County were obtained from IPUMS. This was necessary to get cross-tabs by foreign-born status, which was the proxy we used for immigration status.
- Confusingly, see `egonet/data-prep/make_rda/kc_demographics/kcmort.Rmd`, section 2, for how the raw IPUMS data were imported and processed
- Ignore `egonet/data-prep/make_rda/kc_demographics/kcdemog.Rmd` and the corresponding "kcdemog"' object - this was a placeholder hack used before we got the ACS data.
- The formatted pop data is available within `data(kcmort)`.

2.2   KC mortality

- See `SHAMP/egonet/data-prep/make_rda/kc_demographics/kcmort.Rmd`. Imports one year of mortality data from KC. Later, we got 10 years of data, but this analysis has not been updated to include all 10 years.
- The formatted pop data is available within `data(kcmort)`.

2.3   Network data

The README in SHAMP/egonet/data-prep/make_rda has a fairly extensive guide to the data prep process and the formatted objects available with a `data()` call. Some major points to note:

- Raw data are saved in `egonet/inst/extdata/`. This obscure folder structure was used as per the (at the time current) advice given in the R Packages book
- Th `inst/extdata` folder is not synced to the GitHub repo. To completely replicate the initial data prep, this folder must be transferred locally. The initial reasoning was data privacy. However SHAMP is a private repository, and given that subsequent prepped forms of these same data are synced to the repo, perhaps this inconvenience is not necessary.
- Martina did extensive data prep in SPSS to create the "Final2006-15.sav" NSFG data saved in `egonet/inst/extdata`

Chapter 3

Decisions

3.1 Data prep

See also {#data} chapter.

3.1.1 Definition of partner/network type

See this post of issue 4.

3.1.2 Construction of sex.ident

See 31. We use sex.ident aka sexgrp.yr.adj, which is based on activity in the last year except for those who have no activity in the last year. For them, we use the lifetime value.

3.1.3 Race grouping for Asians and Other race

See 59

3.1.4 Active one-times = Persistent

Issue 79.

3.1.5 Capping degree

See 78 - we cap cohabs at 1, pers at 2.

3.1.6 Imputing partner race-immigration

See issue 96 for a discussion of the problem. We decided to

- Impute from Cohab males to Cohab females - see 97. We maintain the race data and impute immigration.
- Impute from Cohab males to pers - see 98
- Ignore for OT network - see [99]https://github.com/statnet/SHAMP/issues/99)

3.2   Model specification

3.2.1   Modeling one-time rates

See issue 145. We adjust the edges coefficient to reflect the time period over which the one-time data were collected. For SHAMP we decided to look at the one-times in the last month, given the potential recall bias going further back. Issue 6 briefly compares this to the 12-mo recall period.

3.2.2   Nodematch(race), not nodemix

See 64

3.2.3   Modeling MSMF as higher activity

We decided to do this for all networks. Issue 140 shows convergence of Sept 2018/"final" models adding MSMF nodefactor using the sex.ident variable. 73 contains some prior graphs/tables when we considered it previously.

3.2.4   Modeling xnet concurrency

- Issue 141 notes that xnet conc rates don't vary noticeably by sex.ident, so we chose just M vs F
- We left the cohab model with the simpler deg.pers.c indicator of any pers partner. In 142 we justify this saying "For cohab, we checked the data via Martina's NME presentation and decided that there wasn't evidence of meaningful variation by xnet race. We did try replacing nodefactor("deg.pers.c", base=1) with nodefactor("ds.pers", base=1) (see cohab-bbi-conc-C445) and decided even the sex subgroups were not necessary." # Issue 139: we added the simple deg.pers.c and deg.cohab.c to the one-time model
- We settled on using the "sb" and "dsb.cohab" variables in the pers network because variables with more sex-race categories did not converge. See 143 on the creation of these variables.
- Issue #102 explores the possibility of crossing with age as well, which we did not attempt. Issue 146 has the diagnostics for the pers model including these terms.

3.3   ergm.ego pseudopop

3.3.1   Pseudopop size in ergm.ego fits

See issue 136. 50K is sufficient because the MCMC diagnostics are similar between 50K and 100K. That being said, this post does Pavel's recommended (N=sample size)/(w=smallest weight) calculation and shows that we would need a pop size of ~830K to meet that criteria. However, only 2.2% of our weights violate the N/w rule at a pop size of 50K.

3.3.2   Pseudopop nodal attribution

We decided to use ppop.wt='sample'. Otherwise nodal attribution is poor for very small groups, given their very small weights. Issue 123.

3.4 Mean age targets

3.4.1 Target stats for nodefactor(race) dissolution

Issue 21 has a lot of the backstory of worrying about censoring and left truncation. Ultimately we decided to go with a exponentially-distributed ages based on a mean age parameter, assuming left truncation and censoring cancel as per Pavel's proof.

See issue 129 for calculation of the mean age parameter given weights. We decided to:

1. Using each partnership twice, once with race=ego race and once with race=alter race
2. Giving each partnership the ego's weight/2
3. Taking the median duration by race
4. Converting to mean using exponential transform of med/ln(2)

Prior discussions/investigations of this calculation are in 115.

3.4.2 Hitting cohab duration targets

See 147 for our process and 125 for the cohab duration density graphs.

3.5 Notes on how to cross-reference

Analytical decisions made after considering the descriptives, including model specification choices. To cross-reference, check the Rmd file for the section ID. The general format for section IDs of the automated content is #chapterID-sectionname-variablename. For example, 0.73% of our egos are MSMF is linked using the following code:

```
[0.73% of our egos are MSMF](#egos-1way-sex.ident)
```

If the Rmd code does not show a section ID (e.g. you see ## Section name but not ## Section name {#section-id}), either add a section ID or use the default, which is the section name with special characters removed. Encase in square brackets not () and exclude pound sign.

For example, link to the [ERGM fits][ERGM fits] section with the following code:

```
[ERGM fits][ERGM fits]
```

This is obviously problematic for duplicate section names.

See https://bookdown.org/yihui/bookdown/cross-references.html for more detailed rules.