The Rsocialdata.network extension: Handling and analysing egocentric network survey data in R

Emmanuel Rousseaux[*,1,2,3] and Gilbert Ritschard[4,5]

[1]PRN LIVES (http://www.lives-nccr.ch/) – Suisse
[2]Institute for Demographic and Life Course Studies (IDEMO) – University of Geneva, Suisse
[3]Institute of Information and Service Science (IISS) – University of Geneva, Suisse
[4]Institut d'études démographiques et du parcours de vie (IDEMO) – Université de Genève, Suisse
[5]PRN LIVES – Suisse

Résumé

This communication introduces the R package Rsocialdata.network. This package provides specific tools for dealing with network survey data in R. The package comes as an extension to the Rsocialdata package [1] by adding two new classes respectively designed to store (1) network survey variables (for instance emotional support, financial support, or conflict between family members) and (2) covariate variables describing members cited in the networks. These covariates generally include the family tie with the respondent as well as demographic variables: for instance gender, age, marital status, or education level. The framework allows to store network survey variables in one single variable instead of separated variables usually stored in a tabular form. All Rsocialdata features for survey data management are available, for instance the distinction between valid cases and missing values, handling several different missing value types, and a variable description. Most common network measures (centrality, density, in/out-degree, etc.) can directly be computed without any additional recoding or data preparation. Furthermore, the user has the possibility to link network variables with corresponding network covariate data. Specifically, the main features of the package are the net_extract and net_query functions. The net_extract function allows to stem new networks from existing networks using information from covariate variables. For instance, the user can generate a new network variable by removing for each respondent links coming from her/his father in one single step. This is particularly useful to assess the role played by a specific actor on the structure on the network (assessed for instance by the density or the in/out-degree). The net_query function offers facilities to easily retrieve networks based on criteria from both network and covariate information. For example the user can, in one single step, select networks based on the following properties: (1) the father is cited by the respondent (2) the father has a low education, and (3) the density of the network is lower than 0.2. This is particularly useful to model a hypothesis and generate dummies to be tested in regression models. We start this communication by introducing existing packages dealing with network data in R and motivating the development of this extension. Then we introduce the design of the package before to illustrate the package with an application on family configuration of the elderly in the Canton of Geneva, Switzerland. The package is currently available from the R-Forge. [1] Rousseaux E., Bolano D. and Ritschard G. (2013), "The Rsocialdata package: Handling survey data in R", In XXVII IUSSP International Population Conference, Busan, Republic of Korea, August 26th to 31th, 2013.

---

[*]Intervenant