Introduction to ego-network analysis with R

Raffaele Vacca[*1]

[1]University of Florida – United States

Abstract

Prerequisites:
None.

Topics:

- Short introduction to ego-network research and data.
- Data structures for ego-networks in R: vectors, data frames and lists.
- Network objects in R.
- The split-apply-combine strategy: creating ego-level summary variables.
- Looping over multiple ego-networks (for, while, repeat loops).
- Writing your own R functions.
- Applying your functions to multiple ego-networks: The "apply" family of functions (apply, lapply, sapply, etc.).
- The plyr package for easier split-apply-combining.

More information is available at www.raffaelevacca.com/teaching/workshops/ego-network-r/.

Technology requirements:

A laptop with RStudio installed. More details will be emailed to participants before the workshop.

Background:

This workshop offers an introduction to ego-network analysis with R, presenting essential facilities available in R to store and manipulate ego-network data, to visualize ego-networks, and to perform compositional and structural analysis on large collections of ego-networks.

The central idea behind ego-network analysis is that the people (alters) that an individual (ego) knows, and the way that these people interact with each other, affect outcomes in that individual's life such as mental wellbeing, smoking behavior, or assimilation to a foreign culture. A typical ego-network study involves identifying a sample of focal individuals (the egos), and collecting a network of personal contacts (the alters) from each. Ego is asked about characteristics of each alter, characteristics of each ego-alter relation, and characteristics of alter-alter relations. This information is then frequently aggregated into ego-level

---

[*]Speaker

variables that summarize ego-network characteristics, which can subsequently be linked to other ego attributes and outcomes.

Typical ego-network analysis requires handling dozens or hundreds of datasets, each representing one ego-network with ego attributes, alter characteristics, and alter-alter ties. The analysis involves running the same set of operations on each dataset, e.g. to extract compositional and structural summary variables on each ego-network; and joining the resulting metrics into a single dataset, together with other ego-level or alter-level variables. This has been called the split-apply-combine process in data analysis, in which raw data are split into pieces (in this case, each piece representing one ego), the same analysis is applied on each piece, and results are then combined together into a single dataset.

Handling the split-apply-combine process in traditional point-and-click software for statistical analysis is inefficient. Pointing and clicking is repetitive, boring and prone to errors. It typically does not allow users to run the same set of operations on many objects in batch, without the user's intervention. Perhaps more importantly, pointing and clicking makes research not reproducible. R overcomes these limitations and opens up a whole different way of doing ego-network analysis. It eliminates pointing-and-clicking entirely, and allows users to write reproducible scripts that batch analyze hundreds or thousands of ego-networks simultaneously in few seconds.

This workshop will use real-world ego-network data, in combination with the main R packages for network analysis (igraph and statnet). The workshop can be taken as an introduction to the workshop "Simplifying ego-centered network analysis in R with egonetR" by Till Krenz and Andreas Herz. Students interested in a general introduction to social network analysis with R should also consider taking the workshop "Using R and igraph for Social Network Analysis" by Michal Bojanowski.

Keywords: egocentric networks, ego networks, ego nets, personal networks, R, quantitative methods