

# Network Data Collection: comparison of self-assessment and observational methodologies

Michelle González Amador - UNU Merit/Maastricht University

Sebastián Martínez - University of Glasgow

## Abstract

Network data collection methods, in their most general form, can be divided into two categories: self-assessed connections and researcher-observed links. The former is based on self-reported peer networks, and the latter alludes to observable structures such as clusters of countries, but could also include human connections when observed through social media platforms, for example. These two methods make different assumptions about the way individuals connect with each other:

Self-assessed connections require surveyed nodes to think of the reach of their personal network, and can be thought of as subjective closeness or repeated social interactions. Measures of closeness rely heavily on node subjectivity, e.g. family, partners, admired local leaders; ultimately, they allow researchers to analyse network structures as perceived by the nodes. On the other hand, time-type networks follow a logic in which behaviour (re: social learning and choices) is influenced by relevant - subjectively close - node behaviour and the most popular behaviour of the larger network, not unlike a DeGroot learning process. They may partially overlap with subjective closeness networks, but tend to be more diverse and consequently exhibit a lesser degree of homophily. Finally, researcher-observed links (of individuals) provide the opportunity to combine closeness and time-type networks through the construction of artificial weights that include some measure of subjective relevance and a count of social interactions in the observed network; the challenge in this case is whether or not the constructed network is representative of the real one.

Each one of these three structures responds to very particular social contexts. But when the context is not known a priori, e.g. exploratory survey, which network structure could better help us analyse social phenomena? We propose a simulation that allows the same group of individuals to connect in different ways. The simulation is an attempt to understand the extent to which data gathered in one context fits into another. Each simulated individual has a ?closeness? and a ?time? variable, and the respective connections are based on distance to others on that same scale, plus a stochastic component to adjust for randomness. We sample these variables from different distributions,

allowing for different types of networks to emerge. An external, independent observer will determine connections as a mix between subjective closeness and repeated social interactions (re: time). We use a Cobb-Douglass type of function to determine these connections and create networks based on proximity of the 'closeness', 'time', and 'observed' scales. Cobb-Douglass functions allow for a certain equivalence between connections: we assume a strong personal influence has a similar effect to many weaker connections. The resulting statistics answer a set of [sampled] questions that include social learning outcomes and group-based choices and outcomes. Results suggest that researcher-observed and time-type networks? outcomes approximate each other in predictive capacity. Given the natural differences between these three structures, we conclude that these two structures may be more efficient, and allow researchers to use network data for a larger sample of phenomena.