Missing data treatments for the network data

Anja Žnidaršič
University of Maribor, Faculty of Organizational Sciences
anja.znidarsic@fov.uni-mb.si

Patrick Doreian
University of Pittsburgh, Department of Sociology and
University of Ljubljana, Faculty of Social Sciences
pitpat@pitt.edu

Anuška Ferligoj
University of Ljubljana, Faculty of Social Sciences
NRU HSE Moscow
anuska.ferligoj@fdv.uni-lj.si

Abstract

Networks are used to represents a huge variety of relations among units. In social network analysis majority of data are still gathered by survey, while other options include diaries, archival data, transactions data, etc. Regardless the source of gathered network data, all networks are prone to measurement errors. The way the data are collected has an impact to the amount and type of error.

Relations among units under investigation have an instant influence on the type of gathered data. Stated differently, based on the measurement level of ties among units has the classifications of networks could be made. The network could be binary, where only a presence of a tie is recorded, in the valued network the magnitude of tie value is taken into account, while in signed networks the distinction between positive and negative relationships is important.

Missing network data could appear in several forms, here we will focus on missing ties due to non-response in two forms: actor and item non-response. Item non-response appears when an actor participate in a survey, but forget or deny to respond on particular items in the research and a consequence is a row with few missing entries in a matrix representation of a network. Actor non-response appears when an actor refuse to participate in a survey, and therefore a complete row is missing in a matrix representation of a network.

One of the best treatments in case of actor (and item) non-response is the k-nearest neighbours approach which searches for k closest actors according to their incoming ties and then calculates the outgoing ties of the non-respondent as a median of outgoing ties of selected nearest neighbours. Here we will try to determine the appropriate number of the nearest neighbours used to impute missing ties. Simulation results based on several real networks and different blockmodeling approaches will be presented.

Keywords: missing network data, actor non-response, item non-response, k-nearest neighbours, imputations