Social Networks 48 (2017) 46–56

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Actor non-response in valued social networks: The impact of different non-response treatments on the stability of blockmodels

Anja Žnidaršič[a,*], Anuška Ferligoj[b], Patrick Doreian[b,c]

[a] University of Maribor, Faculty of Organizational Sciences, Kidričeva 55a, 4000 Kranj, Slovenia
[b] University of Ljubljana, Faculty of Social Sciences, Kardeljeva ploščad 5, 1000 Ljubljana, Slovenia
[c] University of Pittsburgh, Department of Sociology, USA

ARTICLE INFO

Article history:

Keywords:
Valued network
Missing data
Actor non-response
Actor non-response treatment
Blockmodeling

ABSTRACT

Social network data usually contain different types of errors. One of them is missing data due to actor non-response. This can seriously jeopardize the results of analyses if not appropriately treated. The impact of missing data may be more severe in valued networks where not only the presence of a tie is recorded, but also its magnitude or strength. Blockmodeling is a technique for delineating network structure. We focus on an indirect approach suitable for valued networks. Little is known about the sensitivity of valued networks to different types of measurement errors. As it is reasonable to expect that blockmodeling, with its positional outcomes, could be vulnerable to the presence of non-respondents, such errors require treatment. We examine the impacts of seven actor non-response treatments on the positions obtained when indirect blockmodeling is used. The start point for our simulation are networks whose structure is known. Three structures were considered: cohesive subgroups, core-periphery, and hierarchy. The results show that the number of non-respondents, the type of underlying blockmodel structure, and the employed treatment all have an impact on the determined partitions of actors in complex ways. Recommendations for best practices are provided.

1. Introduction

A key advantage of valued relations in network data, where the strength, intensity, weight, or frequency is recorded instead of only the simplified presence of ties, is a better description of the real world relational data they are trying to capture. However, the recorded tie values are prone to having measurement errors. Not only the misspecified presence or absence of a tie possible (Holland and Leinhardt, 1973), also incorrect tie values can be recorded. Here, we will focus on one specific type of error where one or more actors provide no information regarding all other network members, i.e. actor non-response. Patterns of ties are important in revealing both macro and micro network structure. Misspecification of tie values could severely affect the obtained clusters of actors. To examine this, we investigated the stability of partitions of actors obtained from indirect blockmodeling of valued networks after seven actor non-response treatments are applied.

The paper is organized as follows: Section 2 discusses valued networks. Section 3 focuses on actor non-response. Section 4 presents suitable treatments for actor non-response in valued networks. The basic concept of indirect blockmodeling is presented in Section 5. Section 6 presents the simulation study with regard to the overall design of the simulations, three types of blockmodels used in simulation of network data, and provide the numerical summary of the simulation study. Results are presented in Section 7 by graphical and model representations. Section 8 presents conclusions with an emphasis on recommendations for researchers.

2. Valued networks

Valued network data have their ties measured in terms of magnitudes rather than only the presence or absence of a tie (Wasserman and Faust, 1998; Scott, 2013).[1]

For social networks, vertices represent social actors over which many social relations can be defined. In most settings, relations

* Corresponding author.
E-mail address: anja.znidarsic@fov.uni-mb.si (A. Žnidaršič).

[1] Some authors refer to those networks as *weighted networks*, but we regard value as a broader, more general, concept than a weight. For example, Horvath (2011) defined a weight as a real number between 0 and 1. Here the term '*valued*' is used instead of '*weighted*' when referring to networks, unless we cite from an original source.

http://dx.doi.org/10.1016/j.socnet.2016.06.001
0378-8733/

can be operationalized to include values. Doing this is not always straightforward. Consider, for example, friendship'. Its meaning can range from 'acquaintance' to 'best friend' (Holland and Leinhardt, 1973). Operationalizing magnitude can be done in many ways. Girard et al. (2015) investigated factors influencing formation of a new social network in a new environment by gathering data from university freshmen on 4-point Likert type scale from 'university-acquaintance' to 'a close friend', while Van De Bunt et al. (1999) used a 6-point scale of friendship from 'best friend' to 'troubled relationship'. Grund (2012) studied interactions among soccer players in a longitudinal study. He emphasized the intensity of interactions, and not only the interaction itself, as being especially relevant for small teams where all team members are linked to each other.

A range of values for some relations can be defined unequivocally, for example, trade between countries and rail passengers traveling between cities. Other examples include the co-occurrence of keywords or collaboration of authors in bibliographic analysis of papers, and other networks which are calculated from 2-mode networks. More often, the range of values is established during the preparatory phase of data collection and not only by the question wording. A scale's level of measurement has to be precisely defined. For example, the frequency of collaboration between individuals or departments could be measured by the *number* of e-mails or face-face meetings. Alternatively, it could be estimated on a scale of frequencies ranging from 'never' to 'daily' or 'more often'. Potter et al. (2015) measured contacts between coworkers on 4-point ordinal scale of duration of daily contact (from up to 5 min to at least an hour to eight hours). Hlebec and Ferligoj (2002) used 5-point scales (from 0 (not at all) to 4 (certainly)) for three different sociometric questions: how likely would you borrow study materials from a classmate, likelihood of asking classmates for information about important study assignment, and how likely would you invite a classmate to a birthday party.

Apart from noting that the measurement of valued ties can vary greatly with many operationalizations being available, our intent is not to enter these debates about which measurement approach is the most appropriate for specific empirical situations. Instead, we assume that researchers can and do collect valued network data. These values can vary greatly in magnitude and range. We focus primarily on social relations among individuals. Operationally, our choice was to limit the tie values in our simulations of valued network data in Section 6 to the range 0 to 5, where 0s indicate absence of ties. In principle, this can be extended to other values and larger networks than those considered here.

Investigating the sensitivity of network properties in valued networks with introduced errors has been limited. Some attempts to evaluate regular blockmodeling structures on valued networks with *random* errors were carried out by Žiberna (2009). Páez et al. (2008) investigated impacts of erroneously omitting relevant ties and erroneously including irrelevant *subsets of* ties in the weight matrices for social influence analysis. They emphasized both situations as resulting in biased parameter estimates in network autocorrelation models. Given our concern with *actor non-response*, this study tackles a related problem for valued networks regarding blockmodeling (in Section 5).

3. Non-response in social networks

Actor non-response in social networks is one source of errors in network data (Žnidaršič et al., 2012). Each non-respondent in a network with n actors implies $(n-1)$ missing ties. While all outgoing ties are missing for each non-respondent, the incoming ties are still observed. Fig. 1(a) presents a demonstration network with 15 actors. Suppose it was a real network to be measured. Suppose, further, three actors (A_2, A_{10}, and A_{14} denoted with horizontal gray rectangles) had refused to respond. The actor response rate is 80% – the same as the overall relational response rate reported by Stork and Richards (1992). Fig. 1(b) represents the same network reorganized to have respondents in the upper rows and non-respondents at the bottom. The columns have been reorganized in the corresponding fashion. Missing ties consist of: (i) absent ties between non-respondents and respondents (bottom left part placed in a larger gray rectangle in Fig. 1(b)) and (ii) absent missing ties between non-respondents (right bottom part contained in the larger white square). This distinction for missing ties is important under the different treatments presented in Section 4.

While many studies report response rates the subsequent analyses most often deal only with the data from respondents about other respondents. In effect, this is a data collection imposition raising the well known boundary specification problem (Laumann et al., 1983) because the effective network boundary excludes non-respondents. Studies dealing seriously with boundary problems for networks remain quite rare. Doreian and Woodard (1994) discussed another variant where an 'official' list of the relevant organizations for an inter-organizational study left out many relevant organizations, organizations that were included subsequently by an expanding selection strategy. In general, omitting units is consequential. See also Kossinets (2006), Wang et al. (2012). The obvious question is: does this matter as far as the results of the subsequent analyses?

When the actor response rates are reported, literature reviews reveal a broad range in the number of reported non-respondents. Based on sample of 59 networks, Costenbader and Valente (2003) reported response rates between 51% and 100%.[2] Stork and Richards (1992) reported response rates varying from 65% to 90% of actors. Johnson et al. (2012) reported a 57% overall response rate in a sociometric survey (on friendship, advice, and information flow networks) among employees in Central European bank before investigating only three departments with the highest response rates varying from 63% to 71%. Ellwardt et al. (2012) reported on three waves of a longitudinal study of gossip and friendship relations among employees in organizations with response rates between 85% and 87%. Scherer and Cho (2003) studied risk perception among individuals involved in a community environmental conflict over a hazardous waste site cleanup and they reported 49.5% response rate.[3]

Clearly, actor non-response is a prevalent problem in studying social networks. Having non-respondents in a network be around half may appear to be an extreme case. But it is not so rare in empirical sociometric research that it can be ignored. As a result, we took this notion into account when we included in the simulations seemingly extreme rates of non-response (see Section 6).

Effects of actor non-response on different network properties in binary networks such as network density, average vertex degree, outdegree, indegree, clustering coefficients, transitivity, assortativity, mean inverse geodesic distance and blockmodel structures have been examined previously (Stork and Richards, 1992; Costenbader and Valente, 2003; Borgatti et al., 2006; Kossinets, 2006; Huisman, 2009; Wang et al., 2012; Žnidaršič et al., 2012; Niu et al., 2015). Some of these studies delete the non-respondents and compare the results of analyses but other studies impose different actor non-response treatments (Stork and Richards, 1992; Huisman, 2009; Žnidaršič et al., 2012). It was necessary to establish a set of

[2] Four networks were excluded from their analysis as more than 50% of the actors were non-respondents. This exclusion *might not* have been necessary given some of the results reported below.

[3] The results they reported were based on omitting respondents with any missing data.