

# CORRESPONDENCE ANALYSIS WITH DOUBLING FOR TWO-MODE VALUED NETWORKS

G. Ragozini<sup>1</sup>, D. De Stefano<sup>2</sup> and D. D'Ambrosio<sup>3</sup>

<sup>1</sup> Dip. di Scienze Politiche, Università di Napoli Federico II  
(e-mail: giragoz@unina.it)

<sup>2</sup> Dip. di Scienze Politiche e Sociali, Università di Trieste  
(e-mail: ddestefano@units.it)

<sup>3</sup> Dip. di Scienze Economiche e Statistiche, Università di Napoli Federico II  
(e-mail: daniela.dambrosio2@unina.it)

**ABSTRACT:** In this paper we discuss the use of Correspondence Analysis, in the Greenacre's doubling perspective, to analyze and graphically represent ordinal valued two-mode networks. We discuss how this approach: *i*) takes into account the nature of relational data and the asymmetry of the entities in two-mode networks; *ii*) permits to directly analyze valued relational data, avoiding transformations; *iii*) deals with the nature of the ratings and their bipolar character; *iv*) improves results interpretation.

**KEYWORDS:** Weighted two-Mode Networks, Doubling, Correspondence Analysis

## 1 Introduction

A two-mode network is a data relational structure in which relations are collected or on two different sets of actors (dyadic), or one set of actors and one set of events (affiliation). In some cases, the relational information can be discrete or continuous valued, e.g. the level or strength of ties are coded by a set of ordered categories. Many theoretical and analytical tools used to deal with one-mode networks must have been adapted in order to deal with two-mode networks. Usually, when relationships are valued, the data are dichotomized (often by adopting an arbitrary level of dichotomization) resulting in some loss of information (Borgatti and Everett, 1997). When the interest is visualize and graphically analyze the relational structures embedded in the network, it is possible to use weighted bipartite graphs, spring embedding, and correspondence analysis (CA). In particular, CA allows to derive low dimensional spaces in which it is possible to represent points corresponding to the actors and the events in order to evaluate similarities of participation/attendance patterns.

In this work we will discuss how CA, applied after a proper doubling coding, can be useful to analyze ordinal valued two-mode networks. This ap-

proach has been specially designed to handle bipolar variables (Greenacre, 1984), namely, ordinal variables or ratings, like those resulting from the detection of degree or level of the relations with rating scale (i.e. strength or intensity of the ties). In a nutshell, the proposed method allows to suitably represent the underlying weighted relational distance among actors and events. Moreover, the positions of actors and events in their respective factorial spaces have a nice “relational” interpretation, depending on the level/strength of the observed ties. We present the proposed approach by analyzing a subset, drawn from Wasserman *et al.* (1989).

## 2 Two-Mode Ordinal Valued Networks and Correspondence Analysis

Two-Mode Ordinal Valued Network is a two-mode network, with directed or undirected ties encoding a value that indicates its strength, intensity or frequency measured on a rating scale. The two-mode network that we are considering is a valued digraph  $G(N, E, R)$  composed of two disjoint sets of nodes,  $N$  and  $E$ , with  $N \cup E = \emptyset$ .  $N = n_1, n_2, \dots, n_I$  represents the first set of  $I$  actors (initiators/senders of relations) whereas  $E = e_1, e_2, \dots, e_J$  represents the second set of  $J$  actors (recipients/receivers of relations) or events to which actors participate.  $R$  is the set of ties  $R \subseteq N \times E$ . The edge  $r_{ij} = \{n_i, \rightarrow e_j\}$ ,  $r_{ij} \in R$ , is an ordered pair, defining an ordinal valued social relation and recording the level of intensity or strength of the tie from actor  $n_i$  to actor/event  $e_j$ . Let  $C = \{0, 1, 2, \dots, C-1\}$  be the set containing the discrete values assumed by  $R = \{r_{ij} = 0, 1, 2, \dots, C-1\}$  where the ties were measured on an ordinal scale.

The relational data can be represented by an affiliation (or incidence) matrix  $\mathbf{Y} = (y_{ij})$  with  $y_{ij} = 1, \dots, C-1$  if  $(n_i \rightarrow e_j) \in R$  and 0 otherwise.

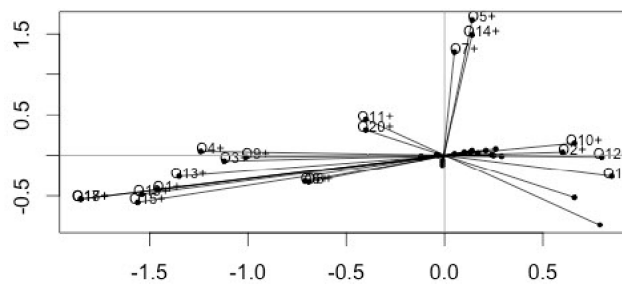
Our approach to analyze two-mode network can be seen as a direct approach – without dichotomization – to visualize and graphically analyze the relational similarity among actors (the set  $N$ ), among actors/events (the set  $E$ ) and between them. Given  $\mathbf{Y}$  we propose to perform CA by applying the usual CA algorithm – SVD of the normalized and centered profile matrix – on a doubled matrix  $\mathbf{Z}$  derived from  $\mathbf{Y}$ , “in order to take into account the absolute nature of the ratings and the fact that they are bipolar, CA may be applied to the doubled data matrix comprising both the original and reflected form of the data” (Greenacre, 1984).

In a doubled  $I \times 2J$  data matrix  $\mathbf{Z}$  each column of the original matrix is recoded into two columns. The  $j_+$  column indicating the measure of positive association where values are  $y_{ij}$ ; the  $j_-$  column indicating the complementary

measure of negative association where values are  $t_j - y_{ij}$ , and  $t_j$  is the upper bound of the  $j$ -th bipolar variable,  $j = 1, \dots, J$ . Doubling establishes a symmetry between poles of each bipolar variable and the CA is invariant with respect to the choice of scale direction. In such a way we consider not only the affiliation pattern, but also the non-affiliation pattern. For details on geometric properties of Correspondence Analysis on doubled matrix see Greenacre (1984). By this approach, *i*) the nature of relational data and the intrinsic asymmetry of the two different sets of entities in two-mode networks are taken into account; *ii*) the valued relational data are considered avoiding transformations (dichotomization) and loss of potentially useful information; *iii*) the nature of the ratings and their bipolar character are properly treated; *iv*) the distinctive features in both rows and columns spaces are taken into account; *v*) the distances in the factorial space reflect relational similarities (i.e. actors and events are closed to each other to the extent they have similar patterns in terms of connections and intensity attached to those); *vi*) the variability of relational patterns can be analyzed by looking at the length of the line (or the plane) passing through the origin of the factorial space. If available, actor and/or event additional covariates may be used to enhance results interpretation. All the usual quantity involved in the analysis have interpretations and are algebraically related to network characteristics (D'Esposito *et al.*, 2014; Ragozini *et al.*, 2015).

As illustrative example, we use data on the 1980 monetary donations from corporations to non-profit organizations in the Minneapolis-st. Paul metropolitan area. Data were analyzed by Wasserman *et al.* (1989). The amount of corporate contributions received from each nonprofit organizations was recorded like categorical data (ordinal variable) with values from 1 (No donation) to 9 (127,000 and over). We analyze the differences between senders actors with respect to the donation propensity patterns and the differences between receivers actors with respect to the capacity of financing patterns. Finally, we identify the joint relational structure between senders and receivers (detecting the donations patterns from corporations to nonprofit agencies). As example of results, in Figure 1, we show the two-dimensional map of the receivers actors in the senders actors space. In the case of donation data we note the high polarization on the negative pole indicating a prevalence of low levels of donations and at least three groups of receivers indicating three principal donations patterns. The most polarized actors are O18, O17, O15, O19, on the left-hand side of the map, and O5, O14 and O7, on the up right-hand side. These non profit organizations are characterized by a few donations focused on specific level/intensity (the higher the polarization, the lower the donations levels vari-

ance). From a relational point of view, the organizations located in each group have a similar donation pattern, indeed they received similar donations levels from the same corporations. Moreover, we can see that there are two opposite groups, O3, O4, O9, O13, on the left side of the map, and O2, O10, on the right. This means they have a similar donation pattern, in terms of the intensity of donation, but from different corporations senders.



**Figure 1.** CA map of receivers actors in the senders actors space

## References

- Borgatti, S. P., Everett, M. G., 1997. Network analysis of 2-mode data. *Social networks*, **19(3)**, 243-269.
- D'Esposito, M.R., De Stefano, D., Ragozini, G., 2014. On the use of Multiple Correspondence Analysis to visually explore affiliation networks. *Social Networks*, **(38)**, 28-40.
- Greenacre, M. J., 1984. *Theory and applications of correspondence analysis*. New York, Academic Press.
- Ragozini, G., De Stefano, D., D'Esposito, M. R., 2015. Multiple factor analysis for time-varying two-mode networks. *Network Science*; **3(01)**, 18-36.
- Wasserman, S., Faust, K., Galaskiewicz, J., 1989. Correspondence and canonical analysis of relational data. *Journal of Mathematical Sociology*; **1(1)**, 11-64.