

# Node-specific effects in latent space modelling of multidimensional networks

## Effetti nodo-specifici nell'analisi a spazi latenti di reti multidimensionali

Silvia D'Angelo and Marco Alfò and Thomas Brendan Murphy

Abstract Social network analysis is a growing and popular field in statistics since the second half of the last century. Although single networks have been largely studied and a variety of models has been developed for their analysis, multidimensional networks are still a young and quite unexplored subject. In most cases, the attention has been focused on the specific case of dynamic networks. The aim of the present work is to provide an extension of latent space models for network analysis to the case of multidimensional networks. We also introduce node-specific effects (sender and receiver effects), typical to the single network literature, to allow for a more flexible representation of the multidimensional network. Finally, a real data application will be presented.

Abstract A partire dalla metà dello scorso secolo, l'analisi delle reti sociali è un campo popolare e in pieno sviluppo in statistica. Sebbene le reti singole siano state ampiamente studiate e siano disponibili una larga varietà di modelli per la loro analisi, l'estensione al caso di reti multidimensionali è ad oggi poco esplorata. Nella maggior parte dei casi, l'attenzione è stata rivolta al caso particolare delle reti dinamiche. In questo lavoro presenteremo un'estensione dei modelli a spazi latenti alle reti multidimensionali. Inoltre, introdurremo effetti nodo-specifici (sender e receiver), tipici in letteratura nel caso di reti singole, per descrivere un modello flessibile di reti multidimensionali. Infine, verrà presentata un'applicazione a dati reali.

Key words: Multidimensional networks, Latent space models

---

Silvia D'Angelo

Sapienza, Università di Roma, Piazzale Aldo Moro 5, Roma. e-mail: [silvia.dangelo@uniroma1.it](mailto:silvia.dangelo@uniroma1.it)

Marco Alfò

Sapienza, Università di Roma, Piazzale Aldo Moro 5, Roma. e-mail: [marco.alfò@uniroma1.it](mailto:marco.alfò@uniroma1.it)

Thomas Brendan Murphy

University College Dublin, Belfield, Dublin 4, Ireland. e-mail: [brendan.murphy@ucd.ie](mailto:brendan.murphy@ucd.ie)

## 1 Introduction

Social network analysis is a well known and vibrant branch of statistics. As network structures arise in many different contexts, network analysis has seen application in a broad variety of fields.

In general, a network is defined by a set of units (the nodes) among which a relation can be established. In the most simple case, the relation to be recorded between a pair of nodes (a dyad), is either present or not present. If it is present, the dyad is said to be linked by an edge. A large variety of models for the analysis of single networks (where a single relation is recorded) has been proposed in the literature (see [7] for a review) and has paved the way for the analysis of more complex network structures, such as multidimensional networks. Recent works on multidimensional networks are those of Gollini and Murphy [1], D'Angelo et al. [6], Salter-Townshend and McCormick [8], Durante et al. [9] and Sewell and Chen [10]. In the present work, we illustrate a model to describe the ties observed in a multidimensional network by means of underlying similarities among the nodes and node-specific characteristics. Section 2 describes the proposed model, while section 3 presents an application of the model to the well known Vickers data [3].

## 2 The model

Let us define a multidimensional network (or multiplex) as a collection of  $K$  networks (or views), defined on the same node set  $N$ . Each network defines a different relation among the nodes in  $N$ . That is, the edge set  $E^{(k)}$ ,  $k = 1, \dots, K$  may be different in each view. A multidimensional network can be represented by means of graph theory as a collection of graphs ( $G$ ), defined on a constant node set  $N$ :

$$G = (N, \{E^{(k)}; k = 1, \dots, K\}),$$

where a single network is  $G^{(k)} = (N, E^{(k)})$ ,  $k = 1, \dots, K$ . If the multiplex collects binary relations, that is whether something is verified or it is not, the realization of  $G$  will be denoted by a collection of adjacency matrices

$$\mathbf{Y} = \{y^{(1)}, \dots, y^{(k)}, \dots, y^{(K)}\}, \quad k = 1, \dots, K,$$

with  $n$  the number of observed nodes. The general entry of the  $k^{th}$  matrix will be  $y_{ij}^{(k)} = 1$  if the relation  $k$  is present between nodes  $i$  and  $j$  and  $y_{ij}^{(k)} = 0$  otherwise. Our aim is at describing the association structure underlying the multidimensional network and at estimating edge probabilities.

A first approach in the context of latent space models for networks, has been introduced by Hoff and Raftery [2]. This class of model assumes that the probability of an edge between two nodes depends on how similar they are;

this similarity refers to their distance in a so called latent (or social) space. The coordinates of the nodes in this space (and therefore the distances) are unknown and the aim is at recovering them to depict the similarities among the nodes and reconstruct the edge probabilities. This class of model was partially extended to the context of multidimensional networks by Gollini and Murphy [1] and D'Angelo et al. [6].

As we are modelling binary adjacency matrices, it is reasonable to model edge probabilities with a logistic regression (as in [6]):

$$Pr\left[y_{ij}^{(k)} = 1 \mid d(z_i, z_j), \alpha^{(k)}, \beta^{(k)}\right] = \frac{\exp\left\{f\left(\alpha^{(k)}, \beta^{(k)}, d(z_i, z_j)\right)\right\}}{1 + \exp\left\{f\left(\alpha^{(k)}, \beta^{(k)}, d(z_i, z_j)\right)\right\}} \quad (1)$$

where  $z_i$  and  $z_j$  are the latent coordinates of nodes  $i$  and  $j$  and  $d(\cdot)$  is the squared Euclidean distance. As the aim is at recovering the similarities among the nodes, the latent space is supposed to be common to all the networks. The sets of parameters  $\alpha = (\alpha^{(1)}, \dots, \alpha^{(K)})$  and  $\beta = (\beta^{(1)}, \dots, \beta^{(K)})$  help to distinguish network connectivity and, in general, link probabilities in the different networks.

We may define a more flexible specification of the edge probabilities by introducing node-specific parameters, to account for the direction of the edge in direct multiplex:

$$Pr\left[y_{ij}^{(k)} = 1 \mid d(z_i, z_j), \alpha^{(k)}, \beta^{(k)}, \theta_i^{(k)}, \gamma_j^{(k)}\right] = \frac{\exp\left\{f\left(\alpha^{(k)}, \beta^{(k)}, d(z_i, z_j), \theta_i^{(k)}, \gamma_j^{(k)}\right)\right\}}{1 + \exp\left\{f\left(\alpha^{(k)}, \beta^{(k)}, d(z_i, z_j), \theta_i^{(k)}, \gamma_j^{(k)}\right)\right\}} \quad (2)$$

The set of parameters  $\Gamma = (\Gamma^{(1)}, \dots, \Gamma^{(K)})$ , where  $\Gamma^{(k)} = (\gamma_1^{(k)}, \dots, \gamma_j^{(k)}, \dots, \gamma_n^{(k)})$  and  $\Theta = (\Theta^{(1)}, \dots, \Theta^{(K)})$ , where  $\Theta^{(k)} = (\theta_1^{(k)}, \dots, \theta_i^{(k)}, \dots, \theta_n^{(k)})$  describe, respectively, receiver and sender effects (see for example Krivitsky et al. [5]). Within each view, sender and receiver parameters are assumed to have a multiplicative effect on the intercept parameter  $\alpha^{(k)}$  and to be bounded in  $[-1, 1]$ .

Estimation of the proposed model parameters is carried out by employing a hierarchical Bayesian framework, using a Metropolis within Gibbs algorithm.

### 3 Application

The model proposed in section 2 has been applied to a benchmark dataset, the Vickers data [3], see for example [4]. In this case three different kind of relations among 29 students have been collected. In particular, the three corresponding networks describe:

1. if student  $i$  gets on with student  $j$ ,
2. if student  $i$  is best friend with student  $j$ ,
3. if student  $i$  works with student  $j$ .

None of the networks is sparse and the observed densities are, respectively, 0.445, 0.223 and 0.244. The observed out-degree (the number of ties a node sends) and in-degree (the amount of ties a node attracts) distributions in the three networks show that only a few nodes interact with most of the others, while the majority of them is less active.

Figure 1 shows the estimated latent positions of the students. The nodes have been coloured according to students gender, with blue and black numbers representing, respectively, males and females. A group structure seems to emerge in the latent space, with reference to the gender. This is confirmed by the heatmap in figure 2, which represents the estimated distances among the nodes and exhibits a clear block structure. Indeed, it seems that male students interact more with other male students, while females prefer the company of other females, exception made for a small group of nodes. These are three boys (5, 8, 11) and two girls (16, 21), all lying in the centre of the latent space. These five serve as a bridge between the two groups in the class. Figure 3 shows the estimated sender (and receiver) effects, together with the observed out-degrees (and in-degrees), for each network. The estimated coefficients are in line with what observed in the data, showing that the proposed model is a good candidate in representing the Vickers multiplex.

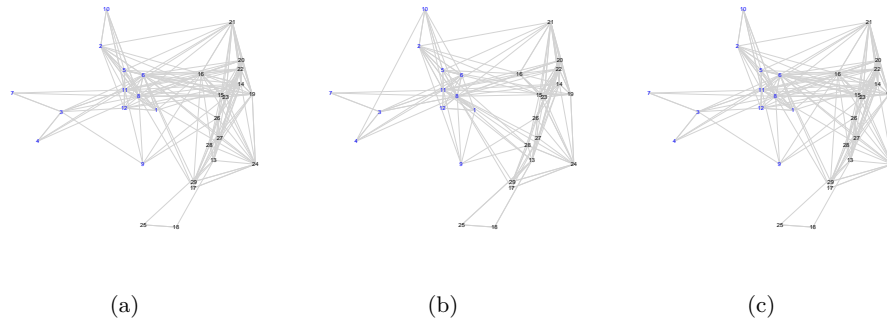


Fig. 1: Estimated latent coordinates for the nodes and observed edges. The relations represented are: get on well (figure *a*), best friend (figure *b*) and work with (figure *c*). Black numbers correspond to female students and blue numbers to males.

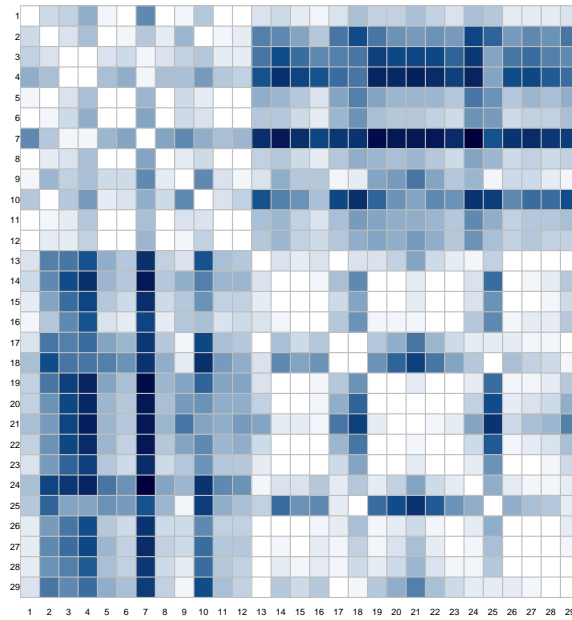


Fig. 2: Estimated distances between the nodes in the latent space. Nodes 1 – 12 are male students, while nodes 13 – 29 are females.

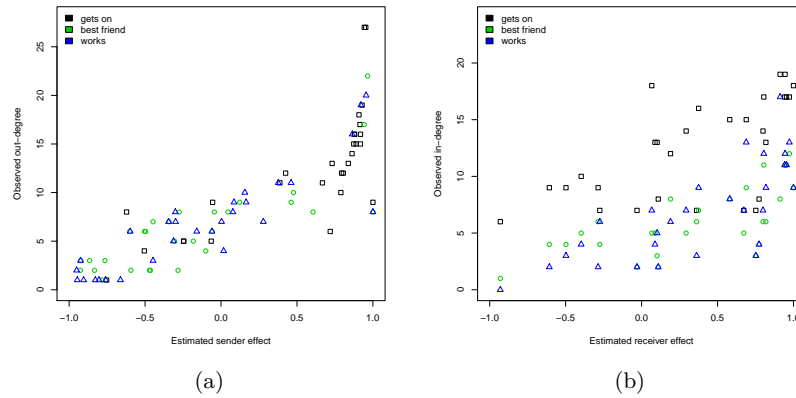


Fig. 3: Estimated sender effects vs. observed out-degrees (figure *a*) and estimated receiver effects vs. observed in-degrees (figure *b*) in the Vickers seventh grade networks.

## 4 Discussion

In this work we have proposed a novel approach to model multidimensional networks that builds both on latent space model for networks and on two typical instruments of social networks analysis: sender and receiver effects. A single latent space is employed to model the similarities between the nodes and node-specific parameters are introduced to model the possible presence of heterogeneity across the multiple networks. This approach allows a flexible reconstruction of the edge probabilities and could serve as an alternative to modelling each network of the multiplex via its own latent space [1]. Indeed, the use of sender and receiver effects prevents from choosing the dimensions of the latent spaces, which could be a relevant issue in some contexts. Unidimensional node-specific parameters would capture network-specific behaviours while the common latent space would represent the overall relationships between the actors in the multiplex. Also, a single latent space allows a straightforward visualization of the multiplex, a useful feature, especially when the number of views is high.

## References

1. Gollini, I., Murphy, T.B.: Joint Modeling of Multiple Network Views. *Journal of Computational and Graphical Statistics*. 25, 246–265 (2016)
2. Hoff, P.D., Raftery, A.E.: Latent space approaches to social network analysis. *Journal of the American Statistical Association*. 97, 1090–1098 (2002)
3. Vickers, M., Chan, S.: *Representing Classroom Social Structure*. Melbourne: Victoria Institute of Secondary Education. (1981)
4. Wasserman, S., Pattinson, P.: Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and  $p^*$ . *Psychometrika*. 61, 401–425 (1996)
5. Krivitsky, P.N., Handcock, M.S., Raftery, A.E., Hoff, P.D.: Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks*. 31, 204–213 (2009)
6. D'Angelo, S., Murphy, T.B., Alfö, M.: Latent space modeling of multidimensional networks with application to the exchange of votes in Eurovision Song Contest. [arXiv:1803.07166](https://arxiv.org/abs/1803.07166). (2018)
7. Salter-Townshend, M., White, A., Gollini, I., Murphy, T. B.: Review of statistical network analysis: models, algorithms, and software. *Statistical Analysis and Data Mining*. 5, 243–264 (2012).
8. Salter-Townshend, M., McCormick, T. H.: Latent space models for multiview network data. *Annals of Applied Statistics*. 11, 1217–1244 (2017).
9. Durante, D., Dunson, D.B., Vogelstein, J.T.: Nonparametric Bayes modeling of populations of networks. *Journal of the American Statistical Association*. 112, 1516–1530 (2017).
10. Sewell, D.K., Chen, Y.: Latent space models for dynamic networks. *Journal of the American Statistical Association*. 110, 1646–1657 (2015).