



Marginal models with individual-specific effects for the analysis of longitudinal bipartite networks

Francesco Bartolucci¹ · Antonietta Mira^{2,3} · Stefano Peluso^{2,4}

Received: 10 June 2023 / Revised: 2 June 2024 / Accepted: 12 August 2024 /
Published online: 3 September 2024
© The Author(s) 2024

Abstract

A new modeling framework for bipartite social networks arising from a sequence of partially time-ordered relational events is proposed. We directly model the joint distribution of the binary variables indicating if each single actor is involved or not in an event. The adopted parametrization is based on first- and second-order effects, formulated as in marginal models for categorical data and free higher order effects. In particular, second-order effects are log-odds ratios with meaningful interpretation from the social perspective in terms of tendency to cooperate, in contrast to first-order effects interpreted in terms of tendency of each single actor to participate in an event. These effects are parametrized on the basis of the event times, so that suitable latent trajectories of individual behaviors may be represented. Inference is based on a composite likelihood function, maximized by an algorithm with numerical complexity proportional to the square of the number of units in the network. A classification composite likelihood is used to cluster the actors, simplifying the interpretation of the data structure. The proposed approach is illustrated on simulated data and on a dataset of scientific articles published in four top statistical journals from 2003 to 2012.

Keywords Association · Binary variables · Cluster analysis · Composite likelihood · Latent trajectories · Log-odds ratios

✉ Francesco Bartolucci
francesco.bartolucci@unipg.it

Antonietta Mira
antonietta.mira@usi.ch

Stefano Peluso
stefano.peluso@unimib.it

¹ Department of Economics, University of Perugia, Via A. Pascoli, 20, 06123 Perugia, Italy

² Euler Institute, Università della Svizzera Italiana, Lugano, Switzerland

³ Department of Sciences and High Technology, Università degli Studi dell'Insubria, Varese, Italy

⁴ Department of Statistics and Quantitative Methods, Milano-Bicocca University, Milan, Italy

Mathematics Subject Classification 91D30 · 62H12 · 62P25

1 Introduction

In many relational contexts, a set of events is observed, with each event involving an arbitrary number of actors and even a single actor. These events give rise to a so-called affiliation network in which there are two types of node: actors and events. Following the current literature, see Wang et al. (2009) among others, we refer to this structure as *bipartite network*, also known as *two-mode network*, contrarily to the *one-mode network* having a unique type of nodes. An example, which motivates the present paper, is that of academic articles in top statistical journals (Ji and Jin 2017), typically involving more than two authors but that might also be written by a single researcher. In these applications, the interest is in studying the relations between units, with the aim of modeling separately the tendency of each unit to be involved in an event, and the tendency of each pair of units to cooperate. We are also interested in studying the time evolution of social behaviors and thus the dynamics over time of both these tendencies. The dataset of statistical publications we aim to analyze has also a particular feature that is important for the following developments: the events are only partially ordered, since we know the year of publication of each article, but no order is available between articles published in the same year.

The literature on bipartite networks is mainly based on models having characteristics similar to those for one-mode networks in which direct connections are observed between certain pairs of actors, such as Exponential Random Graph Models (ERGMs; Frank and Strauss 1986; Wasserman and Pattison 1996); for a review see Snijders (2011) and Amati et al. (2018). One of the first model for the analysis of bipartite networks is proposed in Iacobucci and Wasserman (1990) for data with no temporal dimension and is based on an ERGM structure with specific effects for both types of node (i.e., actors and events) and strong assumptions of independence between the response variables. This approach was extended in several directions by Skvoretz and Faust (1999), whereas Wang et al. (2009) presented a flexible class of ERGMs for bipartite networks and related estimation methods. More recently, a review of models for this type of networks has been illustrated by Aitkin et al. (2014), including certain versions of the Rasch (1967) model and the latent class model (Goodman 1974). The approach proposed in the present paper is also related to models for the analysis of longitudinal one-mode networks, such as actor-oriented models (Snijders and van 1997; Snijders et al. 2010), dynamic ERGMs (Robins and Pattison 2001), hidden Markov models (Yang et al. 2011; Matias and Miele 2017; Bartolucci et al. 2018), and the models for relational events described in DuBois et al. (2013), Perry and Wolfe (2013), Butts and Marcum (2017), Stadtfeld et al. (2017), Fox et al. (2016), and Xia et al. (2016). All these mentioned works handle longitudinal data, but not of bipartite networks. The cited actor-oriented models, relational event models and, generally speaking, all models based on a point process approach deal with time-stamped data, whose dynamics is controlled by an intensity function. On the other hand, our approach, similarly to dynamic ERGMs, are based on time-discrete temporal data. Still, there are crucial differences in this regard between the proposed method and

ERGM: in the latter the temporal dynamics is parameterized through a Markov transition matrix, whilst we do not require it, and we can also handle data only partially ordered, as below for articles published within the same calendar year.

Furthermore, we stress that, relative to ERGM models, our methodology is not based on network summary statistics: since we directly model the vector of allocations of subjects to events, two equal adjacency matrices can be probabilistically treated as different. This allows for instance to consider the joint absence from an event as a point of closeness between two subjects, whilst in ERGMs it is not possible to disentangle from the adjacency matrix bilateral or unilateral absence from an event. Finally, with our marginal modelling approach we can model parsimoniously only subsets of the whole allocation vectors.

In the current paper we therefore model bipartite temporal networks through a marginal approach on the related event allocation vectors, instead of through the association matrix, as more common in the literature. More formally, for the analysis of bipartite networks, and in particular for the dataset of publications in top statistical journals (Ji and Jin 2017), we represent each event by a vector of response variables $\mathbf{Z}^{(e)} = (Z_1^{(e)}, \dots, Z_n^{(e)})'$, with $Z_i^{(e)}$ equal to 1 if unit i is involved in event e and 0 otherwise. First, our aim is to directly formulate a statistical model for the response vectors $\mathbf{Z}^{(e)}$ having a meaningful interpretation. In particular, we rely on a marginal model (Bergsma and Rudas 2002; Bergsma et al. 2009) based on first- and second-order effects. The first-order effects correspond to the logit of the marginal distribution of each $Z_i^{(e)}$ variable and represent the general tendency of actor i to be involved in event e . The second-order effects are the log-odds ratios (Agresti 2013, Ch. 2) for the marginal distribution of each pair of variables $(Z_i^{(e)}, Z_j^{(e)})'$, representing the tendency of actors i and j to be jointly involved in the same event e . However, as we show in detail in the sequel, this parameter may be directly interpreted as the tendency of i and j to cooperate. Moreover, even if we do not directly consider higher order effects, we do not pose any restrictions on these effects. At least to our knowledge, the use of marginal models for the analysis of social network data is new in the statistical literature.

Second, we pay particular attention to the parametrization of the above effects so as to account for the time evolution, and represent individual trajectories in terms of tendency to participate in an event and tendency to cooperate. This feature is common to latent growth models (Bollen and Curran 2006); however, in the proposed approach we use individual fixed parameters, rather than random parameters, applied to polynomials of time of suitable order. Then, the proposed approach is particularly appropriate when the interest is in the evaluation of the behavior of a single actor in terms of the tendencies mentioned above. The possibility to estimate fixed parameters is possible thanks to the amount of information that is typically huge in the applications of interest. For instance, in the motivating example, there are more than three thousand papers that play the role of events. For a related approach, based however on a simpler log-linear parametrization, see Bianchi et al. (2020); see also Bartolucci et al. (2023).

Third, in order to estimate the fixed individual parameters, we rely on a composite likelihood approach (Lindsay 1988; Varin et al. 2011), where, generally speaking, individual components of conditional or marginal densities are multiplied, whether or

not independent, and the resulting derivative constitutes an unbiased estimating equation. Composite likelihood methods have found applications in various areas, such as in spatial statistics (Besag 1974, 1975; Heagerty and Lele 1998), in longitudinal and panel studies (Henderson and Shimakura 2003), for the estimation of time-varying correlation matrices (Pakel et al. 2021), to cite a few. In network analysis, Chen et al. (2018) propose a sequential composite likelihood approach to efficiently estimate social intercorrelations in large-scale social networks, whilst Bartolucci et al. (2015) and Asuncion et al. (2010) estimate via composite likelihood, respectively, a Hidden Markov Model for dynamic networks and an Exponential Random Graph Model. In the current proposal, we use a likelihood function based on the marginal distribution of every ordered pair of actors. For each of these pairs, the likelihood component directly depends on the first- and second-order effects described above, and on individual parameters referred to the two actors. Then, to maximize the target function, we propose a simple iterative algorithm with $O(n^2)$ complexity, that is thus computationally tractable even if the number of actors is large. This is an important feature given the large scale of nowadays social network data; see also the discussion in Vu et al. (2013).

Forth, in presence of many statistical units, we show how to cluster them in groups that are homogenous in terms of tendency to be involved in an event or tendency to cooperate with other units. For this aim, we rely on a classification composite likelihood function that is related to that used for estimating the individual fixed parameters. This allows us to represent trajectories referred to homogeneous groups, rather than to individuals, so as to simplify the interpretation of the evolution of the data structure and of the social perspective of the phenomenon under study.

The paper is organized as follows. In the next section we describe assumptions and interpretation of the proposed approach. In Sect. 3 we outline the method of inference based on the use of fixed effects and clustering techniques. The properties of the proposed estimators are illustrated by a simulation study in Sect. 4, whilst the application is illustrated in Sect. 5. In the last section we draw main conclusions and outline some possible extensions, as the inclusion of third-order effects and of individual covariates. The estimation algorithm is implemented in a series of R functions that we make available to the reader upon request.

2 Proposed model

Let n denote the number of actors and r the number of observed relational events. Also let $Z_i^{(e)}$ be a binary outcome equal to 1 if the relational event e involves unit i and to 0 otherwise, with $i = 1, \dots, n$ and $e = 1, \dots, r$. As already mentioned, these variables are collected in the column vector $\mathbf{Z}^{(e)} = (Z_1^{(e)}, \dots, Z_n^{(e)})'$, a generic configuration of which is denoted by $\mathbf{z} = (z_1, \dots, z_n)'$. Moreover, let $Y_{ij}^{(e)}$ be a binary variable equal to 1 if units i and j are involved in event e , and to 0 otherwise. Note that

$$Y_{ij}^{(e)} = Z_i^{(e)} Z_j^{(e)}, \quad (1)$$

so that the set of variables $Y_{ij}^{(e)}$ is function of the set of variables $Z_i^{(e)}$; the vice-versa does not hold ($Y_{ij}^{(e)} = 0$ can be determined by $Z_i^{(e)} = 0$ or by $Z_j^{(e)} = 0$), confirming that the direct analysis of the $Y_{ij}^{(e)}$ leads, in general, to an information loss. About this point see also the discussion in Aitkin et al. (2014).

2.1 Marginal effects

The main issue is how to parametrize the distribution of the random vectors $\mathbf{Z}^{(e)}$. We adopt a marginal parametrization (Bergsma and Rudas 2002; Bergsma et al. 2009) based on hierarchical effects up to a certain order. This parametrization is less common than the log-linear parametrization, adopted even in ERGMs (Frank and Strauss 1986; Wasserman and Pattison 1996), in which

$$\log \frac{p(\mathbf{Z}^{(e)} = \mathbf{z})}{p(\mathbf{Z}^{(e)} = \mathbf{0})} = \mathbf{g}(\mathbf{z})' \boldsymbol{\gamma},$$

for all configurations \mathbf{z} different from the null configuration $\mathbf{0}$, where $\mathbf{g}(\mathbf{z})$ is a vector-valued function depending on \mathbf{z} .

Indeed, a marginal parametrization may be expressed on the basis of a sequence of log-linear parametrizations referred to the marginal distribution of selected subset of variables, that is,

$$\log \frac{p(\mathbf{Z}_{\mathcal{M}}^{(e)} = \mathbf{z}_{\mathcal{M}})}{p(\mathbf{Z}_{\mathcal{M}}^{(e)} = \mathbf{0}_{\mathcal{M}})} = \mathbf{g}_{\mathcal{M}}(\mathbf{z}_{\mathcal{M}})' \boldsymbol{\gamma}_{\mathcal{M}}, \tag{2}$$

where \mathcal{M} is the set of indices of such variables and $\mathbf{Z}_{\mathcal{M}}^{(e)}$ is the corresponding subvector of $\mathbf{Z}^{(e)}$.

Log-linear parameters are defined as certain sums and differences of logarithms of cell probabilities. Marginal parameters are log-linear parameters calculated from marginal probabilities. In our approach, in particular, we rely on first- and second-order effects, specified, for all e : starting from $\log p(Z_i^{(e)} = 1) = \eta_{i\cdot}^{(e)} + \eta_i^{(e)}$ and $\log p(Z_i^{(e)} = 0) = \eta_{i\cdot}^{(e)} - \eta_i^{(e)}$ for $i = 1, \dots, n$, we can obtain

$$\eta_i^{(e)} = \log \frac{p(Z_i^{(e)} = 1)}{p(Z_i^{(e)} = 0)}, \quad i = 1, \dots, n, \tag{3}$$

as a particular case of (2) with $\mathcal{M} = \{i\}$. Also, from the parameterization

$$\begin{cases} \log p(Z_i^{(e)} = 1, Z_j^{(e)} = 1) = \eta_{i\cdot}^{(e)} + \eta_{i\cdot}^{(e)} + \eta_{\cdot j}^{(e)} + \eta_{ij}^{(e)} \\ \log p(Z_i^{(e)} = 0, Z_j^{(e)} = 0) = \eta_{i\cdot}^{(e)} - \eta_{i\cdot}^{(e)} - \eta_{\cdot j}^{(e)} + \eta_{ij}^{(e)} \\ \log p(Z_i^{(e)} = 1, Z_j^{(e)} = 0) = \eta_{i\cdot}^{(e)} + \eta_{i\cdot}^{(e)} - \eta_{\cdot j}^{(e)} - \eta_{ij}^{(e)} \\ \log p(Z_i^{(e)} = 0, Z_j^{(e)} = 1) = \eta_{i\cdot}^{(e)} - \eta_{i\cdot}^{(e)} + \eta_{\cdot j}^{(e)} - \eta_{ij}^{(e)} \end{cases},$$

we obtain

$$\eta_{ij}^{(e)} = \log \frac{p(Z_i^{(e)} = 0, Z_j^{(e)} = 0)p(Z_i^{(e)} = 1, Z_j^{(e)} = 1)}{p(Z_i^{(e)} = 0, Z_j^{(e)} = 1)p(Z_i^{(e)} = 1, Z_j^{(e)} = 0)}, \quad i, j = 1, \dots, n, \quad j \neq i, \tag{4}$$

as a special case of (2) with $\mathcal{M} = \{i, j\}$. In terms of interpretation (see also Bartolucci et al. 2007), we can easily realize that the marginal logit $\eta_i^{(e)}$ is a *measure of tendency of unit i to be involved in the e -th relational event*. On the other hand, the log-odds ratio $\eta_{ij}^{(e)}$ is a *measure of the tendency of units i and j to cooperate with reference to the same e -th relational event*.

To better interpret the $\eta_{ij}^{(e)}$ effects, it is worth recalling that the log-odds ratio is a well-known measure of association between binary variables (Agresti 2013, Ch. 2), being 0 in the case of independence. In fact, an alternative expression for this effect is

$$\begin{aligned} \eta_{ij}^{(e)} &= \log \frac{p(Z_i^{(e)} = 1|Z_j^{(e)} = 1)}{p(Z_i^{(e)} = 0|Z_j^{(e)} = 1)} - \log \frac{p(Z_i^{(e)} = 1|Z_j^{(e)} = 0)}{p(Z_i^{(e)} = 0|Z_j^{(e)} = 0)} \\ &= \log \frac{p(Z_j^{(e)} = 1|Z_i^{(e)} = 1)}{p(Z_j^{(e)} = 0|Z_i^{(e)} = 1)} - \log \frac{p(Z_j^{(e)} = 1|Z_i^{(e)} = 0)}{p(Z_j^{(e)} = 0|Z_i^{(e)} = 0)}, \end{aligned} \tag{5}$$

corresponding to the increase in the logit of the probability that unit i (or j) is involved in the e -th event, given that unit j (or i) is present in the same event, with respect to the case the latter is not present. More details in this regard are provided in the following section.

Before illustrating how we parametrize in a parsimonious way the marginal effects defined above, it is worth recalling that, apart from the trivial case of $n = 2$ actors, the knowledge of these effects is not sufficient to obtain univocally the corresponding distribution of the vectors $\mathbf{Z}^{(e)}$. To formulate this argument more formally, let $\mathbf{p}^{(e)}$ denote the vector containing the 2^n joint probabilities $p(\mathbf{Z}^{(e)} = \mathbf{z})$ for all possible configurations \mathbf{z} in lexicographical order. Also let $\boldsymbol{\eta}_1^{(e)}$ be the vector containing the first-order effects $\eta_i^{(e)}$ for $i = 1, \dots, n$ and let $\boldsymbol{\eta}_2^{(e)}$ denote the corresponding vector of second-order effects $\eta_{ij}^{(e)}$ for $i = 1, \dots, n - 1$ and $j = i + 1, \dots, n$. It is possible to prove that

$$\boldsymbol{\eta}^{(e)} = \begin{pmatrix} \boldsymbol{\eta}_1^{(e)} \\ \boldsymbol{\eta}_2^{(e)} \end{pmatrix} = \mathbf{C} \log(\mathbf{M} \mathbf{p}^{(e)}) \tag{6}$$

for a suitably defined matrix of contrasts \mathbf{C} and a marginalization matrix \mathbf{M} with elements equal to 0 or 1. However this relation is not one-to-one, in the sense that it is not possible to obtain a unique probability vector $\mathbf{p}^{(e)}$ starting from $\boldsymbol{\eta}^{(e)}$.

In order to have an invertible parametrization, the structure of higher order effects must be specified. Just to give the idea, a third-order marginal effect between units i, j , and k may be defined as

$$\eta_{ijk}^{(e)} = \eta_{ijk|1}^{(e)} - \eta_{ijk|0}^{(e)},$$

where

$$\eta_{ijk|z}^{(e)} = \log \frac{p(Z_i^{(e)} = 0, Z_j^{(e)} = 0 | Z_k^{(e)} = z) p(Z_i^{(e)} = 1, Z_j^{(e)} = 1 | Z_k^{(e)} = z)}{p(Z_i^{(e)} = 0, Z_j^{(e)} = 1 | Z_k^{(e)} = z) p(Z_i^{(e)} = 1, Z_j^{(e)} = 0 | Z_k^{(e)} = z)}, \quad z = 0, 1.$$

This is the difference between the conditional log-odds ratio for units i and j given that unit k is present with respect to the case it is not present. This directly compares to the *triangularization effect* in an ERGM, as it measures how much the presence of unit k affects the chance that units i and j collaborate. In a similar way we may recursively define effects of order higher than three until order n (Bartolucci et al. 2007), so that including the specification of these effects, the parametrization in (6) becomes invertible.

In the present approach, however, we prefer to focus only on first- and second-order effects as formulated in (3) and (4), leaving the structure of higher-order interactions unspecified. In fact, as we show in detail in Sect. 3, to make inference on these effects it is not necessary to specify the structure of these higher-order effects as we base inference on a pairwise likelihood function. This is an advantage of the marginal parametrization with respect to the log-linear parametrization; the latter does not allow us to directly express the marginal distribution of a subset of variables without specifying the full set of interactions. The way of obtaining each bivariate probability vector

$$\mathbf{p}_{ij}^{(e)} = \begin{pmatrix} p(Z_i^{(e)} = 0, Z_j^{(e)} = 0) \\ p(Z_i^{(e)} = 0, Z_j^{(e)} = 1) \\ p(Z_i^{(e)} = 1, Z_j^{(e)} = 0) \\ p(Z_i^{(e)} = 1, Z_j^{(e)} = 1) \end{pmatrix} \tag{7}$$

on the basis of the parameters $\eta_i^{(e)}$, $\eta_j^{(e)}$, and $\eta_{ij}^{(e)}$, which are collected in the vector $\boldsymbol{\eta}_{ij}^{(e)} = (\eta_i^{(e)}, \eta_j^{(e)}, \eta_{ij}^{(e)})'$, is clarified in the Appendix.

2.2 Interpretation of the log-odds ratios

To clarify the interpretation of the log-odds ratio $\eta_{ij}^{(e)}$, it is useful to consider that it directly compares with the logit of the probability that there is a connection between units i and j , in the sense that both units are involved in the same event. In fact, from (1) we have

$$\tilde{\eta}_{ij}^{(e)} = \log \frac{p(Y_{ij}^{(e)} = 1)}{p(Y_{ij}^{(e)} = 0)} = \log \frac{p(Z_i^{(e)} = 1, Z_j^{(e)} = 1)}{1 - p(Z_i^{(e)} = 1, Z_j^{(e)} = 1)},$$

that is a commonly used effect in typical social network models; see, for instance, Hoff et al. (2002). There is an important difference between $\tilde{\eta}_{ij}^{(e)}$ and $\eta_{ij}^{(e)}$: the former corresponds to the tendency of units i and j to be involved in the same event, but it does not disentangle this joint tendency from the marginal tendency of each single unit to be involved in the same event. For instance, $\tilde{\eta}_{ij}^{(e)}$ could attain a large value only because both units have, separately, a high tendency to be involved in the event (both authors are very active in their publication strategy) even if there is not a particular “attraction” between them, namely with low values of $p(Z_i^{(e)} = 0, Z_j^{(e)} = 0)$. On the other hand, $\eta_{ij}^{(e)}$ is a proper *measure of attraction* because, as clearly shown by (5), it corresponds to the increase in the chance that one unit is present in the event given that also the other unit is present in the same event. This difference between parameters $\eta_{ij}^{(e)}$ and $\tilde{\eta}_{ij}^{(e)}$ is key in the proposed approach and is of particular relevance in the application of our interest, where each event may involve a variable number of actors and, possibly, also only one actor. Indeed, in our approach the general tendency of unit i to be involved in an event is meaningfully measured by effects $\eta_i^{(e)}$ defined in (3). Note that effects of this type cannot be directly included in an ERGM.

The above arguments may be further clarified considering that, for given marginal distributions $p(Z_i^{(e)})$ and $p(Z_j^{(e)})$, or, in other terms, for fixed $\eta_i^{(e)}$ and $\eta_j^{(e)}$, the log-odds ratio $\eta_{ij}^{(e)}$ is an increasing function of $p(Z_i^{(e)} = 1, Z_j^{(e)} = 1) = p(Y_{ij}^{(e)} = 1)$ and thus of $\tilde{\eta}_{ij}^{(e)}$. Moreover, for a given value of this joint probability, $\eta_{ij}^{(e)}$ is a decreasing function of $\eta_i^{(e)}$ and $\eta_j^{(e)}$. In particular, considering that

$$\begin{aligned} p(Z_i^{(e)} = 0, Z_j^{(e)} = 0) &= 1 - p(Z_i^{(e)} = 1) - p(Z_j^{(e)} = 1) + p(Z_i^{(e)} = 1, Z_j^{(e)} = 1), \\ p(Z_i^{(e)} = 0, Z_j^{(e)} = 1) &= p(Z_j^{(e)} = 1) - p(Z_i^{(e)} = 1, Z_j^{(e)} = 1), \\ p(Z_i^{(e)} = 1, Z_j^{(e)} = 0) &= p(Z_i^{(e)} = 1) - p(Z_i^{(e)} = 1, Z_j^{(e)} = 1), \end{aligned}$$

we can easily realize that

$$\begin{aligned} \frac{\partial \eta_{ij}^{(e)}}{\partial \tilde{\eta}_{ij}^{(e)}} &= p(Z_i^{(e)} = 1, Z_j^{(e)} = 1)[1 - p(Z_i^{(e)} = 1, Z_j^{(e)} \\ &= 1)] \sum_{z_1=0}^1 \sum_{z_2=0}^1 \frac{1}{p(Z_i^{(e)} = z_1, Z_j^{(e)} = z_2)} > 0. \end{aligned}$$

Similarly, we have

$$\frac{\partial \eta_{ij}^{(e)}}{\partial \eta_i^{(e)}} = -p(Z_i^{(e)} = 1)[1 - p(Z_i^{(e)} = 1)] \sum_{z_1=0}^1 \sum_{z_2=0}^1 \frac{1}{p(Z_i^{(e)} = z_1, Z_j^{(e)} = z_2)} < 0,$$

with a corresponding expression for $\partial \eta_{ij}^{(e)} / \partial \eta_j^{(e)}$. An illustration of this behavior is provided in Fig. 1, where for a pair of individuals we represent the value of $\eta_{ij}^{(e)}$ with

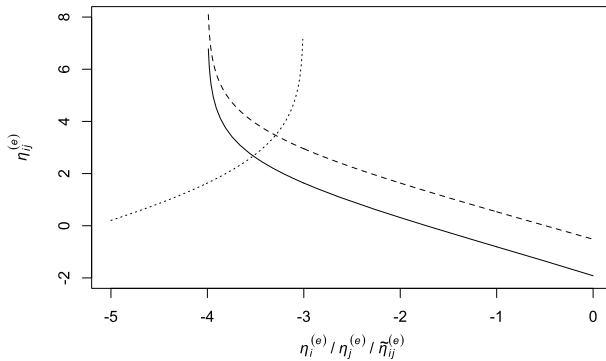


Fig. 1 Plot of the log-odds ratio $\eta_{ij}^{(e)}$ with respect to $\eta_i^{(e)}$ (solid line), $\eta_j^{(e)}$ (dashed line), and $\tilde{\eta}_{ij}^{(e)}$ (dotted line)

respect to $\eta_i^{(e)}$ (with $\eta_j^{(e)} = -2$ and $\tilde{\eta}_{ij}^{(e)} = -4$), to $\eta_j^{(e)}$ (with $\eta_i^{(e)} = -3$ and $\tilde{\eta}_{ij}^{(e)} = -4$), and to $\tilde{\eta}_{ij}^{(e)}$ (with $\eta_i^{(e)} = -3$ and $\eta_j^{(e)} = -2$).

Another advantage of $\eta_{ij}^{(e)}$ with respect to $\tilde{\eta}_{ij}^{(e)}$ is that the former induces a variational independent parametrization (Bergsma and Rudas 2002b). This means that the joint distribution of $(Z_i^{(e)}, Z_j^{(e)})'$ exists for any value in \mathbb{R} of the first-order effects $\eta_i^{(e)}$ and $\eta_j^{(e)}$ and of the second-order effect $\eta_{ij}^{(e)}$. More formally, the function relating $\eta_{ij}^{(e)} = (\eta_i^{(e)}, \eta_j^{(e)}, \eta_{ij}^{(e)})'$ with $\mathbf{p}_{ij}^{(e)}$ defined in (7) is one-to-one for $\eta_{ij}^{(e)}$ in \mathbb{R}^3 and $\mathbf{p}_{ij}^{(e)}$ in the four-dimensional simplex. This has advantages in terms of model interpretation and estimation. On the other hand, effect $\tilde{\eta}_{ij}^{(e)}$ has a limited range of possible values with bounds depending on $\eta_i^{(e)}$ and $\eta_j^{(e)}$, making the joint estimation and interpretation of $\tilde{\eta}_{ij}^{(e)}$, $\eta_i^{(e)}$ and $\eta_j^{(e)}$, more problematic.

2.3 Parametrization of marginal effects

Formulating a model for relational events requires to parametrize, in a parsimonious way, the effects $\eta_i^{(e)}$ and $\eta_{ij}^{(e)}$ of main interest. In absence of individual covariates, we propose the following parametrization of the first-order effects:

$$\eta_i^{(e)} = \mathbf{f}_1(t_e)' \boldsymbol{\alpha}_i, \tag{8}$$

where $\mathbf{f}_1(t_e)$ is a vector-valued function specific of time t_e of each event e . For instance, this function may contain the terms of a polynomial of suitable order of the day or year of event e starting from the beginning of the study. This parametrization is similar to that of a latent trajectory model (Dwyer 1983; Crowder and Hand 1996; Menard 2002), with the difference that, as we clarify in the sequel, each $\boldsymbol{\alpha}_i$ is considered as a vector of fixed individual parameters. In any case, it is possible to represent individual trajectories regarding the tendency over time to be present in an event.

Regarding the second-order effects, a natural extension of (8) would lead to a vector of specific parameters for each pair of units. However, to obtain a parsimonious model, we prefer to rely on an additive parametrization of type

$$\eta_{ij}^{(e)} = f_2(t_e)'(\beta_i + \beta_j), \tag{9}$$

where $f_2(t_e)$ is defined as $f_1(t_e)$ and, again, vectors β_i represent the evolution of the tendency, of unit i , to collaborate across time. We use two different functions, $f_1(t_e)$ and $f_2(t_e)$, to allow for a different order of the involved polynomials of time. Note, however, that the additive structure in (9) implies that $f_2(t_e)'\beta_i$ may be interpreted as the “general” tendency of individual i to collaborate with other individuals in an event at time t_e .

Overall, for each bivariate probability vector $p_{ij}^{(e)}$, the parametrization based on (8) and (9) is linear in the parameters. In particular, if we let $\delta_i = (\alpha_i', \beta_i)'$, we have that

$$\eta_{ij}^{(e)} = (D_{ij1} \ D_{ij2}) \begin{pmatrix} \delta_i \\ \delta_j \end{pmatrix}, \tag{10}$$

where D_{ij1} and D_{ij2} are suitable design matrices.

To clarify the proposed parametrization, consider a sample of $n = 9$ individuals for a single event, for different values of the intercepts α_i (from -3 to -1) and of β_i (from -1 to 1). These values are reported in Table 1 together with certain average probabilities that help to understand the meaning of these parameters. The single 2×2 tables for each pair of actors are reported in Table 2.

Table 1 Parameters α_i and β_i together with mean marginal probabilities and probabilities of two actors being involved in the same event

	α_i	β_i	mean($p(Z_i^{(e)})$)	mean($p(Z_i^{(e)}, Z_j^{(e)} = 1)$)
1	- 3.00	-1.00	0.05	0.00
2	- 3.00	0.00	0.05	0.01
3	- 3.00	1.00	0.05	0.01
4	- 2.00	-1.00	0.12	0.01
5	- 2.00	0.00	0.12	0.02
6	- 2.00	1.00	0.12	0.03
7	- 1.00	-1.00	0.27	0.02
8	- 1.00	0.00	0.27	0.04
9	- 1.00	1.00	0.27	0.05

Table 2 Single 2×2 tables for each pair of actors considered in Table 1

	<i>j</i> = 1		<i>j</i> = 2		<i>j</i> = 3		<i>j</i> = 4		<i>j</i> = 5		
	0	1	0	1	0	1	0	1	0	1	
<i>i</i> = 1	0	–	–	0.906	0.047	0.907	0.045	0.834	0.118	0.836	0.117
	1	–	–	0.047	0.001	0.045	0.002	0.047	0.001	0.045	0.002
<i>i</i> = 2	0	0.906	0.047	–	–	0.910	0.042	0.836	0.117	0.839	0.114
	1	0.047	0.001	–	–	0.042	0.005	0.045	0.002	0.042	0.006
<i>i</i> = 3	0	0.907	0.045	0.910	0.042	–	–	0.839	0.114	0.846	0.107
	1	0.045	0.002	0.042	0.005	–	–	0.042	0.006	0.035	0.012
<i>i</i> = 4	0	0.834	0.047	0.836	0.045	0.839	0.042	–	–	0.768	0.113
	1	0.118	0.001	0.117	0.002	0.114	0.006	–	–	0.113	0.006
<i>i</i> = 5	0	0.836	0.045	0.839	0.042	0.846	0.035	0.768	0.113	–	–
	1	0.117	0.002	0.114	0.006	0.107	0.012	0.113	0.006	–	–
<i>i</i> = 6	0	0.839	0.042	0.846	0.035	0.855	0.026	0.776	0.105	0.790	0.091
	1	0.114	0.006	0.107	0.012	0.098	0.022	0.105	0.014	0.091	0.028
<i>i</i> = 7	0	0.686	0.045	0.689	0.042	0.696	0.035	0.618	0.113	0.627	0.104
	1	0.267	0.002	0.263	0.006	0.256	0.013	0.262	0.006	0.254	0.015
<i>i</i> = 8	0	0.689	0.042	0.696	0.035	0.707	0.024	0.627	0.104	0.644	0.087
	1	0.263	0.006	0.256	0.013	0.246	0.023	0.254	0.015	0.237	0.032
<i>i</i> = 9	0	0.696	0.035	0.707	0.024	0.717	0.014	0.644	0.087	0.667	0.064
	1	0.256	0.013	0.246	0.023	0.235	0.034	0.237	0.032	0.213	0.055

	<i>j</i> = 6		<i>j</i> = 7		<i>j</i> = 8		<i>j</i> = 9		
	0	1	0	1	0	1	0	1	
<i>i</i> = 1	0	0.839	0.114	0.686	0.267	0.689	0.263	0.696	0.256
	1	0.042	0.006	0.045	0.002	0.042	0.006	0.035	0.013
<i>i</i> = 2	0	0.846	0.107	0.689	0.263	0.696	0.256	0.707	0.246
	1	0.035	0.012	0.042	0.006	0.035	0.013	0.024	0.023
<i>i</i> = 3	0	0.855	0.098	0.696	0.256	0.707	0.246	0.717	0.235
	1	0.026	0.022	0.035	0.013	0.024	0.023	0.014	0.034
<i>i</i> = 4	0	0.776	0.105	0.618	0.262	0.627	0.254	0.644	0.237
	1	0.105	0.014	0.113	0.006	0.104	0.015	0.087	0.032
<i>i</i> = 5	0	0.790	0.091	0.627	0.254	0.644	0.237	0.667	0.213
	1	0.091	0.028	0.104	0.015	0.087	0.032	0.064	0.055
<i>i</i> = 6	0	–	–	0.644	0.237	0.667	0.213	0.692	0.189
	1	–	–	0.087	0.032	0.064	0.055	0.039	0.080
<i>i</i> = 7	0	0.644	0.087	–	–	0.501	0.230	0.534	0.197
	1	0.237	0.032	–	–	0.230	0.039	0.197	0.072
<i>i</i> = 8	0	0.667	0.064	0.501	0.230	–	–	0.576	0.155
	1	0.213	0.055	0.230	0.039	–	–	0.155	0.114
<i>i</i> = 9	0	0.692	0.039	0.534	0.197	0.576	0.155	–	–
	1	0.189	0.080	0.197	0.072	0.155	0.114	–	–

3 Pairwise likelihood inference

Before introducing the proposed methods of inference for the model described above, we clarify the data structure used in applications. We start from the data

$$\mathbf{w}_{ij}^{(e)} = \begin{pmatrix} I(z_i^{(e)} = 0, z_j^{(e)} = 0) \\ I(z_i^{(e)} = 0, z_j^{(e)} = 1) \\ I(z_i^{(e)} = 1, z_j^{(e)} = 0) \\ I(z_i^{(e)} = 1, z_j^{(e)} = 1) \end{pmatrix}, \tag{11}$$

where $I(\cdot)$ is the indicator function, for $i = 1, \dots, n - 1, j = i + 1, \dots, n$, and $e = 1, \dots, r$. In our motivating application, $\mathbf{w}_{ij}^{(e)}$ denotes the joint participation of authors i and j in the scientific article e . For instance $\mathbf{w}_{ij}^{(e)} = (0, 1, 0, 0)'$ means that only author j is involved in article e , or $\mathbf{w}_{ij}^{(e)} = (0, 0, 0, 1)'$ that both authors are involved. Moreover, in the applications of interest, it is possible to group events that, by assumption, have the same distribution. For instance, in the application based on the academic articles published by statisticians, it is plausible to assume that for all articles published in the same year the distribution of the binary vector is the same, even because it is not possible to have the precise dates of the publication and thus their precise time order. In other words, it is sensible to group events in homogenous periods $t = 1, \dots, t(r)$, where $t(e)$ denotes the time of event e . Then, the relevant information is that contained in the frequency vectors

$$\tilde{\mathbf{w}}_{ij}^{(t)} = \sum_{e:t(e)=t} \mathbf{w}_{ij}^{(e)},$$

and consequently we denote by $\tilde{\mathbf{p}}_{ij}^{(t)}$ the corresponding probability vector having the same structure as in (7).

3.1 Fixed-effects estimation

It is possible to estimate the parameters of interest by maximizing the pairwise log-likelihood function (Lindsay 1988; Varin et al. 2011):

$$p\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{e=1}^r [\mathbf{w}_{ij}^{(e)}]' \log \mathbf{p}_{ij}^{(e)},$$

where $\boldsymbol{\theta}$ is the vector of all such parameters, that is the collection of individual parameter vectors $\boldsymbol{\delta}_i, i = 1, \dots, n$, used in (10). An alternative expression is

$$p\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \ell_{ij}(\boldsymbol{\delta}_i, \boldsymbol{\delta}_j),$$

$$\ell_{ij}(\delta_i, \delta_j) = \sum_{t=1}^{t(r)} [\tilde{\mathbf{w}}_{ij}^{(t)}]' \log \tilde{\mathbf{p}}_{ij}^{(t)}, \tag{12}$$

which is faster to compute as it relies on the frequency vectors defined in (11).

In order to maximize $p\ell(\theta)$, it is important to obtain the score vector of each components $\ell_{ij}(\delta_i, \delta_j)$. To this aim, it is convenient to introduce the log-linear effects $\tilde{\mathbf{p}}_{ij}^{(t)}$ which are collected in the vector $\tilde{\boldsymbol{\lambda}}_{ij}^{(t)} = (\lambda_i^{(e)}, \lambda_j^{(e)}, \lambda_{ij}^{(e)})'$, where

$$\begin{aligned} \lambda_i^{(e)} &= \log \frac{p(Z_i^{(e)} = 1 | Z_j^{(e)} = 0)}{p(Z_i^{(e)} = 0 | Z_j^{(e)} = 0)}, \\ \lambda_j^{(e)} &= \log \frac{p(Z_j^{(e)} = 1 | Z_i^{(e)} = 0)}{p(Z_j^{(e)} = 0 | Z_i^{(e)} = 0)}, \\ \lambda_{ij}^{(e)} &= \eta_{ij}^{(e)}, \end{aligned}$$

and e is any of the events at time occasion t . Also let $\tilde{\boldsymbol{\eta}}_{ij}^{(t)}$ denote the corresponding vector of marginal parameters. We have that

$$s_{ij}(\delta_i) := \frac{\partial \ell_{ij}(\delta_i, \delta_j)}{\partial \delta_i} = \sum_{t=1}^{t(r)} \mathbf{D}'_{ij1} \frac{\partial \boldsymbol{\eta}_{ij}^{(t)}}{\partial [\boldsymbol{\lambda}_{ij}^{(t)}]'} \mathbf{G}[\tilde{\mathbf{w}}_{ij}^{(t)} - m_{ij}^{(t)} \tilde{\mathbf{p}}_{ij}^{(t)}],$$

where $m_{ij}^{(t)}$ is the sum of the elements of $\tilde{\mathbf{w}}_{ij}^{(t)}$, that is, the number of events in time period t , whereas \mathbf{G} and the derivative of $\tilde{\boldsymbol{\eta}}_{ij}^{(t)}$ with respect to $\tilde{\boldsymbol{\lambda}}_{ij}^{(t)}$ are defined in Appendix.

The estimation algorithm is based on the following steps. First of all define an initial guess for the parameters δ_i , denoted by $\delta_i^{(0)}$, $i = 1, \dots, n$. Then, for every unit i , find the values of δ_i such that

$$\sum_{j=1, j \neq i}^n s_{ij}(\delta_i) = \mathbf{0},$$

so as to maximize

$$p\ell_i(\theta) = \sum_{j=1, j \neq i}^n \ell_{ij}(\delta_i, \delta_j), \tag{13}$$

with respect to δ_i , with all other parameters kept fixed. Iterate this process until convergence in $p\ell(\theta)$, and denote the final parameter estimates by $\hat{\delta}_i = (\hat{\alpha}'_i, \hat{\beta}'_i)'$, $i = 1, \dots, n$, which are collected in the vector $\hat{\theta}$. In practice, the algorithm steps may be implemented by using a readily available numerical solver.

3.2 Clustering

With large samples, it is typically of interest to find clusters of units presenting a similar behavior. In our approach this amounts to assume that there are h_1 groups of individuals having a similar behavior in terms of tendency to be involved in an event and h_2 groups of individuals having a similar tendency to collaborate in the network. For each group we have specific parameter vectors denoted by $\alpha^*_{g_1}$ and $\beta^*_{g_2}$, with $g_1 = 1, \dots, h_1$ and $g_2 = 1, \dots, h_2$, all collected in the parameter vector θ^* .

For unit i , let d_{i1} denote the cluster to which the unit is assigned with respect to the first type of tendency and d_{i2} the cluster assigned with respect to the second type of tendency. The corresponding classification pairwise log-likelihood has the same expression as $p\ell(\theta)$ defined in (12), with $\alpha_i = \alpha^*_{d_{i1}}$, $\beta_i = \beta^*_{d_{i2}}$, and then $\delta_i = ((\alpha^*_{d_{i1}})', (\beta^*_{d_{i2}})')$. This function is denoted by $cpl(\theta^*, \mathbf{d}_1, \mathbf{d}_2)$, where \mathbf{d}_1 is the vector with elements d_{i1} and \mathbf{d}_2 is that with elements d_{i2} , respectively, with $i = 1, \dots, n$.

To cluster units in homogeneous groups, we maximize $cpl(\theta, \mathbf{d}_1, \mathbf{d}_2)$ by an iterative algorithm that is initialized from the output of a k -means clustering of the individual estimates $\hat{\alpha}_i$ and $\hat{\beta}_i$. Then, it alternates the following three steps until convergence:

1. for $i = 1, \dots, n$ try to change the cluster d_{i1} of unit i by finding the cluster that maximizes the individual component of the classification pairwise log-likelihood, which is defined as in (13) accounting for the cluster structure, with all other parameters kept fixed;
2. for $i = 1, \dots, n$ try to change the cluster d_{i2} of unit i by the same procedure as above;
3. update the parameter estimates of $\alpha^*_{g_1}$, $g_1 = 1, \dots, h_1$, and $\beta^*_{g_2}$, $g_2 = 1, \dots, h_2$, by maximizing $cpl(\theta^*, \mathbf{d}_1, \mathbf{d}_2)$ with respect to θ^* with \mathbf{d}_1 and \mathbf{d}_2 kept fixed.

We select h_1 and h_2 as the smallest number of clusters such that the initial clustering of the estimates $\hat{\alpha}_i$ and $\hat{\beta}_i$, performed by the k -means algorithm, leads to a between sum of squares equal to at least 80% of the total sum of squares. Then, at convergence of the three steps illustrated above, we check that the number of clusters is adequate, comparing the maximum value of $cpl(\theta^*, \mathbf{d}_1, \mathbf{d}_2)$ with that of $p\ell(\theta)$, as we show in connection with the application in the next section.

4 Simulation

As clarified in Section 2.1, the knowledge of main and second-order marginal effects, without knowing higher-order effects, is not enough to generate data from an assumed model. Moreover, even with a small number of actors, generating data is computationally unfeasible, apart from the case in which all second- and higher-order effects are null. Given these constraints, we now implement a study on simulated data that mimic the scientific publications data analyzed in the next section, to provide relevant indications on the properties of the proposed estimation methods.

The simulation study is based on a benchmark design (DESIGN 0) and on alternative designs, variations of the benchmark. Under DESIGN 0, we assume that $r = 250$ events involve an overall number of $n = 500$ individuals and cover $T = 5$ consecutive periods. The individual parameters measuring the tendency to participate to the event, denoted as α_i , can equally likely assume $h_1 = 4$ possible values equally spaced from -5.0 to -2.5 . To simplify the generation of the simulated data, the tendency to reciprocate, β_i , are all equal to 0. The alternative simulation designs alter DESIGN 0, by using $n = 1,000$ rather than 500 (DESIGN 1), $T = 10$ rather than 5 (DESIGN 2), $k = 8$ rather than 4 clusters (DESIGN 3), and $r = 500$ events rather than 250 (DESIGN 4).

Under each design, 100 samples have been generated and the model based on a second-order time polynomial for both the tendency to connect and the tendency to reciprocate is assumed. This is the same model estimated in the following section to analyze the data on scientific publications. Estimation is based on both the fixed-effects method illustrated in Sect. 3.1 and the clustering estimation method illustrated in Sect. 3.2. The quality of these estimation methods is measured, for each sample, by the average across individuals of the absolute value of the difference between each parameter estimate and the corresponding true value. The average across samples of these sample-specific averages are reported in Table 3 for the fixed-effects estimation method and in Table 4 for the cluster estimation method, where the measures of errors are indicated by $\text{err}(\hat{\alpha}_l)$ and $\text{err}(\hat{\beta}_l)$ for $l = 1, 2, 3$. The clustering results in Table 4 also reports the value of the Adjusted Rand Index (ARI; Hubert and Arabie 1985) between the true and predicted clustering.

We note that as the number of actors n increases (DESIGN 1 vs 0), the average error of the fixed effects method does not change substantially because, as expected, the

Table 3 Average measure of error of the fixed-effects estimator of every parameter under each simulation scenario

n	T	h_1	r	$\text{err}(\hat{\alpha}_1)$	$\text{err}(\hat{\alpha}_2)$	$\text{err}(\hat{\alpha}_3)$	$\text{err}(\hat{\beta}_1)$	$\text{err}(\hat{\beta}_2)$	$\text{err}(\hat{\beta}_3)$
500	5	4	250	0.228	0.480	0.460	0.116	0.181	0.178
1000	5	4	250	0.228	0.480	0.463	0.113	0.159	0.158
500	10	4	250	0.132	0.407	0.401	0.073	0.120	0.122
500	5	8	250	0.211	0.450	0.434	0.103	0.162	0.159
500	5	4	500	0.133	0.289	0.285	0.046	0.078	0.078

Table 4 Average measure of error of the estimation method based on clustering for every parameter under each simulation scenario

n	T	h_1	r	$\text{err}(\hat{\alpha}_1)$	$\text{err}(\hat{\alpha}_2)$	$\text{err}(\hat{\alpha}_3)$	$\text{err}(\hat{\beta}_1)$	$\text{err}(\hat{\beta}_2)$	$\text{err}(\hat{\beta}_3)$	ARI
500	5	4	250	0.150	0.130	0.256	0.116	0.181	0.178	0.875
1000	5	4	250	0.016	0.024	0.027	0.113	0.159	0.158	0.895
500	10	4	250	0.011	0.033	0.033	0.073	0.120	0.122	0.976
500	5	8	250	0.568	0.773	0.622	0.103	0.162	0.159	0.424
500	5	4	500	0.059	0.099	0.080	0.046	0.078	0.078	0.965

greater amount of information is compensated by the larger number of parameters to be estimated. On the contrary and coherently, the error level of the cluster estimation method strongly improves with an increase in the number of subjects. Also, we observe that as the number of time periods T increases (DESIGN 2 vs 0), both methods improve, since there is a higher data availability but no additional parameters to estimate, with a more pronounced gain in the performance of the clustering method. When the number of clusters h_1 increases (DESIGN 3 vs 0) we observe that the fixed-effects method does not change its performance, whilst the cluster estimation method has a worse performance: the same amount of information is now split into a higher number of clusters, worsening the estimation performance of cluster-specific parameters. As a result, the ARI, while having satisfactory levels close to 1 in all other designs, in DESIGN 3 is significantly lower. Finally, as the number of events r increases (DESIGN 4 vs 0), both methods have a significant improvement in terms of performance: more occasions are available to learn marginal and joint tendencies of subjects to participate in the events.

5 Application

In order to illustrate the approach based on individual-specific effects, see assumptions (8) and (9), we propose an application based on the data recently made available by Ji and Jin (2017). The data refer to the publication history of all authors with at least one paper published in four top statistical journals (*Annals of Statistics*, *Biometrika*, *Journal of the American Statistical Association*, *Journal of the Royal Statistical Society - series B*) between 2003 and the first half of 2012. Overall, 3607 authors are involved who coauthored 3248 articles. In Table 5 we report some descriptive statistics, whereas in Fig. 2 we represent the distribution of the number of articles and the number of coauthors for each individual in the dataset.

From Table 5 we observe that the number of published papers per year does not vary considerably, even if there is an increase from 2003 to 2009 and a decrease after 2009 (year 2012 counts only partially). Moreover, the average number of authors moderately increases during the time span and it is important to note that the number of single-author articles is relevant in each year. Indeed, these articles represent the 16.7% of the total articles considered in the dataset and this justifies the use of the proposed

Table 5 Descriptive statistics on the number of authors per article

Year	# articles	# authors										Average
		1	2	3	4	5	6	7	8	9	10	
2003	296	78	132	63	19	2	2	0	0	0	0	2.125
2004	320	70	157	67	17	5	3	1	0	0	0	2.197
2005	328	64	166	77	16	4	0	1	0	0	0	2.189
2006	354	55	178	96	18	4	2	1	0	0	0	2.288
2007	350	56	158	105	24	3	2	1	0	0	1	2.363
2008	370	59	151	114	32	9	2	2	1	0	0	2.459
2009	409	53	177	128	42	4	2	0	1	1	1	2.489
2010	355	53	151	104	36	8	2	1	0	0	0	2.451
2011	325	39	135	107	34	5	4	1	0	0	0	2.529
2012	141	16	63	49	9	2	1	1	0	0	0	2.468
	3248	543	1468	910	247	46	20	9	2	1	2	2.357

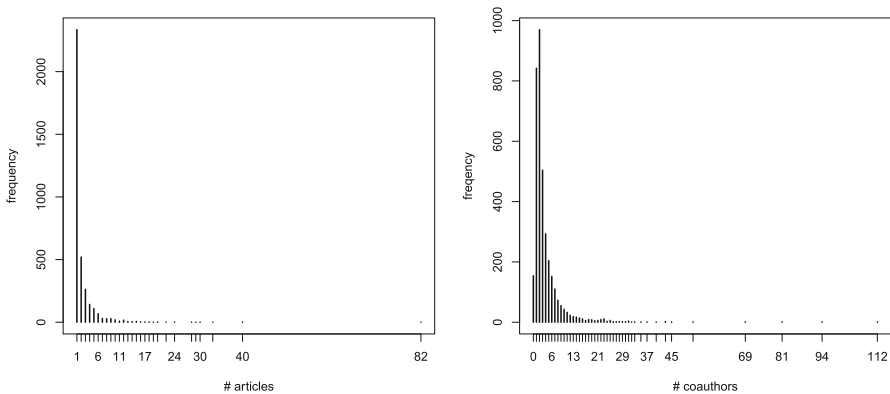


Fig. 2 Distribution of the number of articles per author (left panel) and of the number of coauthors (right panel)

approach for the analysis. Regarding Fig. 2, we note the concentration of the number of articles, with the majority of authors (2335) who published only one article in one of the four top journals considered in the reference period and 520 who published only two articles. On the other hand, the three most productive researchers published 33, 40, and 82 articles, with a total of 155 overall (ignoring possible overlapping). Even the distribution of the number of coauthors shows a very high concentration, although the situation is somehow different as the third modality (i.e., 2 coauthors) has the highest frequency, equal to 970. The authors with zero and one coauthors are 154 and 842, respectively, whereas the three highest modalities are 81, 94 and 112.

The application is based on two phases, that is, fixed-effects estimation and clustering, which are illustrated in Sects. 3.1 and 3.2, respectively. The fixed-effect estimation method is executed in about 27 h and the clustering method in about the double, on an Intel(R) Core(TM) i7-8565U CPU @ 1.80 GHz 1.99 GHz machine. Regarding the first phase, with reference to (8) and (9), we assume second order polynomials for the effect of time. In this way we estimate 3607 parameter vectors α_i and β_i of length 3. On the basis of these estimates it is possible to obtain trajectories both in terms of tendency to publish an article in a certain period and in terms of tendency to collaborate with other authors. We recall that, for the former, the effect that is represented is the logit defined in (3) and for the latter it is the log-odds ratio (4). The choice of a second order polynomial is driven by model parsimony, to have a reasonable low number of parameters with non-linear trajectories of tendencies to publish and collaborate. Still, we stress that the proposed method is thought more generally for f_1 and f_2 vector-valued functions of time, extendable, for instance and according to the application and the computational capacity, to higher-order polynomials or splines. In order to illustrate these results, we consider the five authors with the largest number of published articles in the period that we identify with the letters from A to E; the patterns of publication of these authors is reported in Table 6.

For these top five authors, we represent the estimated profiles in Fig. 3 where we can clearly identify author E as the most productive one with a profile that follows a reverse U-shape, having its pick around years 2006 and 2007, coherently with the data in Table 6. On the other hand, this author shows the lowest profile in terms of tendency to collaborate with other authors, in accordance to a ratio between number of coauthors and number of published papers in the table which, for author E, tends to be lower than for the other main authors. The conclusion is that the large number of coauthors of author E can be mostly ascribed to his tendency to publish; see Sect. 2.2 for general comments on the interpretation of these results. In a similar way we can interpret the profiles of the other authors. For instance, authors C and D, who published the same number of papers (namely 40), have very similar profiles in terms of tendency to publish, but according to the proposed model, D has a higher tendency to collaborate, with a difference that also increases in time, and in particular the curve for D always dominates that for C; this is again coherent with the data in Table 6. Finally, authors A and B have profiles which are in agreement with a smaller number of published papers that, at the same time, tends to increase from 2003 to 2012.

To improve the interpretability of these profiles, instead of using the logits defined in (3), we can also express the tendency to publish in terms of expected number of publications per year. These expected values are obtained by multiplying the probability to be involved in a publication by the yearly number of publications available in Table 5. The resulting profiles are reported in Fig. 4 and confirm the previous conclusions.

When we examine the overall sample of 3607 authors, the analysis may be effectively carried out by building clusters of authors. Following the approach described in Sect. 3.2, we find evidence of $h_1 = 6$ different profiles in terms of tendency to publish and $h_2 = 5$ profiles in terms of tendency to collaborate. The model with clustered profiles attains a maximum profile log-likelihood (normalized dividing by the number of ordered pairs of units) equal to -34.916 that is close that of the fixed-effects method, which is equal to -33.693 . On the other hand, the maximum pairwise log-likelihood

Table 6 Publication profiles of the five authors with the largest number of published papers in the period considered

Author	Year												Total
	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012			
A	# papers	2	1	6	3	1	4	4	4	3	2	30	
	# coauth	3	1	8	5	2	5	5	4	7	3	43	
	Ratio	1.50	1.00	1.33	1.67	2.00	1.25	1.25	1.00	2.33	1.50	1.43	
B	# papers	1	2	3	3	3	3	4	5	9	0	33	
	# coauth	1	2	3	3	4	6	4	8	12	0	43	
	Ratio	1.00	1.00	1.00	1.00	1.33	2.00	1.00	1.60	1.33	-	1.30	
C	# papers	2	5	3	1	6	6	6	5	3	3	40	
	# coauth	3	5	6	2	13	9	10	9	6	6	69	
	Ratio	1.50	1.00	2.00	2.00	2.17	1.50	1.67	1.80	2.00	2.00	1.73	
D	# papers	2	5	3	3	6	3	8	4	4	2	40	
	# coauth	6	13	7	3	14	5	17	15	10	4	94	
	Ratio	3.00	2.60	2.33	1.00	2.33	1.67	2.12	3.75	2.50	2.00	2.35	
E	# papers	7	7	10	12	10	8	11	9	6	2	82	
	# coauth	7	10	13	16	13	12	16	10	12	3	112	
	Ratio	1.00	1.43	1.30	1.33	1.30	1.50	1.45	1.11	2.00	1.50	1.37	

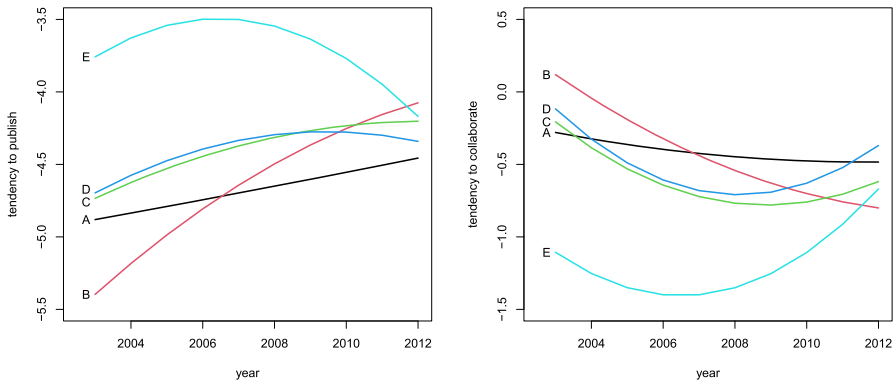
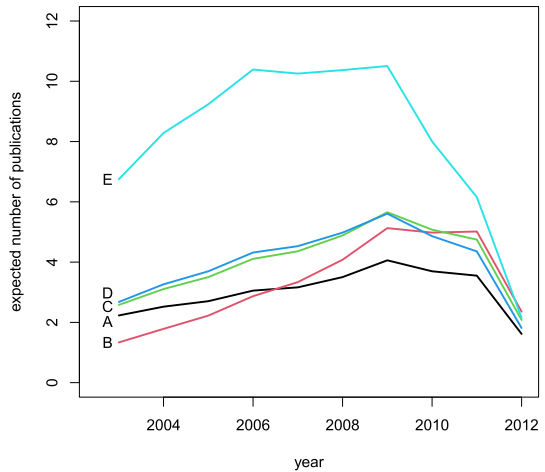


Fig. 3 Profiles in terms of tendency of authors to publish papers (left panel) and of collaborate (right panel), for the top five authors A, B, C, D, and E

Fig. 4 Profiles of tendency of authors to publish papers, measured by the expected number of yearly publications, for the top five authors A, B, C, D, and E



with only one cluster is equal to -75.401 . This means that using a structure of 6×5 clusters implies an improvement of the pairwise log-likelihood equal to 97.1% with respect of using only one cluster, despite the huge reduction in the number of parameters with respect to the fixed-effects model: rather than using n individual-specific parameter vectors α_i and β_i , we use a very limited number of parameter vectors denoted by α_g^* and β_g^* . The corresponding profiles are represented in Fig. 5.

It is interesting to note that the 6 cluster profiles in terms of tendency to publish cover different possibilities. In particular, profiles for the first 3 clusters have a reversed U-shape with picks located at different years. On the other hand, profiles for clusters 4 and 5 have a U-shape but are rather different. Finally, authors in cluster 6 tend to have a rather constant over time tendency to publish, which is higher than for the other clusters. In any case, the values of the logit even for this class corresponds to low probability levels, with an expected number of publications which is smaller than 1

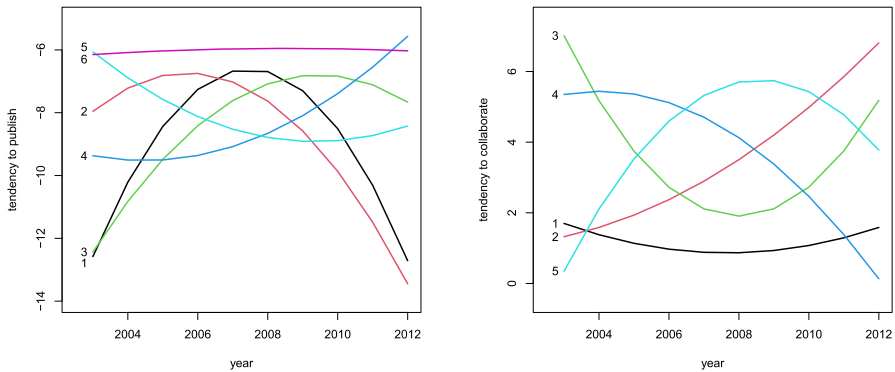


Fig. 5 Profiles in terms of tendency of authors to publish papers (left panel) and to collaborate (right panel), for clusters 1–5, obtained following the analysis in Section 3.2

for all years. In a similar way we can interpret the 5 clusters in terms of tendency to collaborate. For instance, individuals in the first cluster have a general tendency to collaborate lower than the other authors, which is rather constant in time.

It is important to stress that, in principle, any author may belong to any cluster of the first type (in terms of tendency to publish) and of the second type (in terms of tendency to collaborate). In order to better understand this aspect, we consider the cross classification of authors according to both criteria. This cross classification is reported in Table 7 that also shows the size of each cluster in terms of units assigned to it.

On the basis of the results in Table 7 we observe that all clusters have a comparable dimension without a neat prevalence in terms of size of a specific cluster, although in terms of tendency to publish cluster 3 is the largest and the same happens for the tendency to collaborate. Regarding the association between the two classification criteria, it is interesting to comment on certain regular patterns that appear evident. The most relevant one is that individuals in cluster 6 in terms of tendency to publish are

Table 7 Cross classification of authors in terms of tendency to publish and to collaborate

Tend. to Publish	Tend. to collaborate					Total
	1	2	3	4	5	
1	27	0	577	0	0	604
2	85	584	0	0	0	669
3	130	0	374	301	0	805
4	42	0	0	478	3	523
5	108	28	0	0	478	614
6	392	0	0	0	0	392
Total	784	612	951	779	481	3607

all in the first cluster in terms of tendency to collaborate. In summary, the authors with the highest tendency to publish have, at the same time, the smallest tendency to collaborate. As noted above, when commenting the results shown in Fig. 3 for author E, this means that, for the most productive authors, having a large number of coauthors is more plausibly ascribed to the tendency to publish than to a pure tendency to collaborate. Author E can be considered as a pivotal or representative author for this joint class.

Other patterns may be easily discovered by looking at Table 7 as, for instance, that authors in cluster 1 in terms of tendency to publish are mostly in cluster 3 in terms of tendency to collaborate. Comparing the two profiles we observe that they have an opposite shape, which is coherent with the previous reasoning according to which the tendency to have a large number of coauthors may be reasonably ascribed to the general tendency to publish than to a specific social behavior. This is likely due to the fact that scientific collaborations are viewed as long-term investments, and once a productive author establishes a team of researchers that effectively collaborate with each other, he/she tends to fully exploit these known scientific relations, with the aim of authoring even a large number of articles without changing the team. The whole analysis is repeated on a reduced dataset that excludes 2012, the partly available last year of the sample. The results are substantially equivalent to those reported above, and minor differences are commented in detail in the Supplementary Material.

6 Conclusions

We propose a new model for social networks arising from a sequence of events involving an arbitrary (one or more) number of actors. The main novelties, relative to available approaches, may be summarized as: (i) binary variables $Z_i^{(e)}$ for actor i being involved in event e are directly modeled instead of the tie variables $Y_{ij}^{(e)} = Z_i^{(e)}Z_j^{(e)}$; (ii) the model is based on marginal first- and second-order effects that have a meaningful interpretation in terms of tendency of an actor to participate in an event and tendency to cooperate; (iii) these effects are parametrized accounting for each event time and individual fixed-effects parameters, so that the evolution of individual behaviors may be represented by suitable trajectories; (iv) inference is based on a composite likelihood function, built on the distribution of each ordered pair of units, and maximized with numerical complexity of order $O(n^2)$, n being the network size; (v) units may be clustered in groups having the same behavior so as to simplify the interpretation of the data structure. The results of a simulation study confirms the adequacy of the proposed estimation methods.

In conclusion, it is worth noting that the proposed approach may be potentially used in contexts different from our motivating example. In particular, we can formulate third- (or higher) order effects to account for triangularizations. This amounts to rely on a different composite likelihood function: a composite likelihood based on triples or a pairwise conditional likelihood, depending on the type of third-order effects used. However, triples would necessarily lead to a higher computational complexity.

Furthermore, individual covariates can be easily incorporated. In fact, we rely on a parametrization based on a linear predictor with suitable polynomials of event times. This linear predictor can include, in a natural way, individual covariates with no increase in complexity. However, covariates specific to each pair of units (not to single units) and to events, contribute to increase the computational burden.

Finally, our fixed-effect model can be naturally extended to random-effects and lends itself to a Bayesian formulation. By simply adding suitable priors on model parameters, the same inferential algorithm for finding the composite maximum likelihood estimate can be adopted to find maximum a posteriori estimates, with minor adjustments.

Funding Open access funding provided by Università degli Studi di Perugia within the CRUI-CARE Agreement. Funding was provided by Swiss National Science Foundation (Grant Number 208249 and 200557).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix: Obtaining joint probabilities from marginal effects

A simple but crucial issue concerns how to obtain the joint distribution of two binary variables starting from the corresponding marginal effects. In particular, consider two binary variables A and B having joint probabilities

$$p_{AB}(a, b) := p(A = a, B = b), \quad a, b = 0, 1,$$

which are collected in lexicographical order in the vector \mathbf{p} . Also let $p_A(a) := p(A = a)$ and $p_B(b) := p(B = b)$ denote the corresponding marginal probabilities. The marginal parameters are defined as

$$\begin{aligned} \eta_A &:= \log \frac{p_A(1)}{p_A(0)}, \\ \eta_B &:= \log \frac{p_B(1)}{p_B(0)}, \\ \eta_{AB} &:= \log \frac{p_{AB}(0, 0)p_{AB}(1, 1)}{p_{AB}(0, 1)p_{AB}(1, 0)}, \end{aligned}$$

and are collected, following the order given above, in the 3-dimensional vector $\boldsymbol{\eta}$.

The inversion from $\boldsymbol{\eta}$ to \mathbf{p} is based on a formula studied in a more general context by Dale (1986). On the basis of some simple algebra, we find that the joint probability

$p_{AB}(1, 1)$ is equal to

$$p_{AB}(1, 1) = \frac{1 + [p_A(1) + p_B(1)](e^{\eta_{AB}} - 1) - \sqrt{\Delta}}{2(e^{\eta_{AB}} - 1)},$$

with

$$\Delta = [1 + (p_A(1) + p_B(1))(e^{\eta_{AB}} - 1)]^2 - 4(e^{\eta_{AB}} - 1)p_A(1)p_B(1)e^{\eta_{AB}}$$

and where $p_A(1) = e^{\eta_A} / (1 + e^{\eta_A})$, with $p_B(1)$ computed similarly. Obviously, if $\eta_{AB} = 0$, then the solution is simply $p_{AB}(1, 1) = p_A(1)p_B(1)$. In the end, we have

$$\begin{cases} p_{AB}(1, 0) = p_A(1) - p_{AB}(1, 1), \\ p_{AB}(0, 1) = p_B(1) - p_{AB}(1, 1), \\ p_{AB}(0, 0) = 1 - p_A(1) - p_B(1) + p_{AB}(1, 1). \end{cases}$$

Now consider the canonical (log-linear) parameters

$$\begin{aligned} \lambda_A &= \log \frac{p_{AB}(1, 0)}{p_{AB}(0, 0)}, \\ \lambda_B &= \log \frac{p_{AB}(0, 1)}{p_{AB}(0, 0)}, \\ \lambda_{AB} &= \log \frac{p_{AB}(1, 1)p_{AB}(0, 0)}{p_{AB}(0, 1)p_{AB}(1, 0)}, \end{aligned}$$

which are collected, following the order given above, in the 3-dimensional vector λ , and note that

$$p_{AB}(a, b) = \frac{e^{a\lambda_A + b\lambda_B + ab\lambda_{AB}}}{K(\lambda)},$$

where $K(\lambda)$ is the normalizing constant. For estimation purposes, the derivative of λ with respect to η is necessary. We obtain this as the inverse of the derivative of η with respect to λ , noting that

$$\begin{aligned} \eta_A &= \lambda_A + \log \frac{1 + e^{\lambda_B + \lambda_{AB}}}{1 + e^{\lambda_B}}, \\ \eta_B &= \lambda_B + \log \frac{1 + e^{\lambda_A + \lambda_{AB}}}{1 + e^{\lambda_A}}, \\ \eta_{AB} &= \lambda_{AB}. \end{aligned}$$

Then we have:

$$\frac{\partial \eta}{\partial \lambda'} = \begin{pmatrix} 1 & -\frac{e^{\lambda_B}(1 - e^{\lambda_{AB}})}{(1 + e^{\lambda_B})(1 + e^{\lambda_B + \lambda_{AB}})} & -\frac{e^{\lambda_B + \lambda_{AB}}}{1 + e^{\lambda_B + \lambda_{AB}}} \\ -\frac{e^{\lambda_A}(1 - e^{\lambda_{AB}})}{(1 + e^{\lambda_A})(1 + e^{\lambda_A + \lambda_{AB}})} & 1 & -\frac{e^{\lambda_A + \lambda_{AB}}}{1 + e^{\lambda_A + \lambda_{AB}}} \\ 0 & 0 & 1 \end{pmatrix}, \tag{14}$$

whereas regarding the inverse function we have

$$\frac{\partial \lambda}{\partial \eta'} = \left(\frac{\partial \eta}{\partial \lambda'} \right)^{-1} = \left(1 - \frac{\partial \eta_A}{\partial \lambda_B} \frac{\partial \eta_B}{\partial \lambda_A} \right)^{-1} \begin{pmatrix} 1 & -\frac{\partial \eta_A}{\partial \lambda_B} & \frac{\partial \eta_A}{\partial \lambda_B} \frac{\partial \eta_{AB}}{\partial \lambda_B} & -\frac{\partial \eta_{AB}}{\partial \lambda_A} \\ -\frac{\partial \eta_B}{\partial \lambda_A} & 1 & \frac{\partial \eta_{AB}}{\partial \lambda_A} & -\frac{\partial \lambda_A}{\partial \eta_{AB}} \\ 0 & 0 & 1 - \frac{\partial \eta_A}{\partial \lambda_B} \frac{\partial \eta_B}{\partial \lambda_A} & \end{pmatrix},$$

having elements directly taken from (14).

References

Agresti A (2013) *Categorical data analysis*, 3rd edn. Wiley-Interscience, Hoboken, New Jersey

Aitkin M, Vu D, Francis B (2014) Statistical modelling of the group structure of social networks. *Soc Netw* 38:74–87

Amati V, Lomi A, Mira A (2018) Social network modeling. *Ann Rev Stat Appl* 5:343–369

Asuncion A, Liu Q, Ihler A, Smyth, P (2010) Learning with blocks: composite likelihood and contrastive divergence. In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 33–40. *JMLR Workshop and Conference Proceedings*

Bartolucci F, Colombi R, Forcina A (2007) An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints. *Statistica Sinica* 17:691–711

Bartolucci F, Li Donni P, Mira A (2023) Temporal analysis of hospital network data by hierarchical Bayesian p2 models with covariates. *J Royal Stat Soc Ser A* 186:422–440

Bartolucci F, Marino MF, Pandolfi S (2015) Composite likelihood inference for hidden Markov models for dynamic networks. *MPRA Paper 67242*, University Library of Munich, Germany

Bartolucci F, Marino MF, Pandolfi S (2018) Dealing with reciprocity in dynamic stochastic block models. *Comput Stat Data Anal* 123:86–100

Bergsma W, Croon MA, Hagenaars JA (2009) *Marginal models: for dependent, clustered, and longitudinal categorical data*. Springer Science & Business Media, NY

Bergsma WP, Rudas T (2002) Marginal models for categorical data. *Ann Stat* 30:140–159

Bergsma WP, Rudas T (2002b) Variation independent parameterizations of multivariate categorical distributions. In: *Distributions with given marginals and statistical modelling*, pp. 21–27. Springer

Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. *J Roy Stat Soc B* 36:192–225

Besag J (1975) Statistical analysis of non-lattice data. *J Royal Stat Soc Ser D* 24:179–195

Bianchi F, Bartolucci F, Peluso S, Mira A (2020) Longitudinal networks of dyadic relationships using latent trajectories: evidence from the European interbank market. *J Royal Stat Soc Ser A* 69:711–739

Bollen KA, Curran PJ (2006) *Latent Curve Models*. Wiley, Hoboken, NJ

Butts CT, Marcum CS (2017) A relational event approach to modeling behavioral dynamics. In: *Group processes*, pp. 51–92. Springer

Chen Y, Qi Y, Liu Q, Chien P (2018) Sequential sampling enhanced composite likelihood approach to estimation of social intercorrelations in large-scale networks. *Quant Market Econom* 16:409–440

Crowder D, Hand M (1996) *Practical longitudinal data analysis*. Chapman and Hall, London, England

- Dale JR (1986) Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics* 42:909–917
- DuBois C, Butts CT, McFarland D, Smyth P (2013) Hierarchical models for relational event sequences. *J Math Psychol* 57:297–309
- Dwyer J (1983) *Statistical models for the social and behavioral sciences*. Oxford University Press, Oxford, England
- Fox EW, Short MB, Schoenberg FP, Coronges KD, Bertozzi AL (2016) Modeling e-mail networks and inferring leadership using self-exciting point processes. *J Am Stat Assoc* 111:564–584
- Frank O, Strauss D (1986) Markov graphs. *J Am Stat Assoc* 81:832–842
- Goodman LA (1974) Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* 61:215–231
- Heagerty PJ, Lele SR (1998) A composite likelihood approach to binary spatial data. *J Am Stat Assoc* 93:1099–1111
- Henderson R, Shimakura S (2003) A serially correlated gamma frailty model for longitudinal count data. *Biometrika* 90:355–366
- Hoff PD, Raftery AE, Handcock MS (2002) Latent space approaches to social network analysis. *J Am Stat Assoc* 97:1090–1098
- Hubert L, Arabie P (1985) Comparing partitions. *J Classif* 2:193–218
- Iacobucci D, Wasserman S (1990) Social networks with two sets of actors. *Psychometrika* 55:707–720
- Ji P, Jin J (2017) Coauthorship and citation networks for statisticians. *Ann Appl Stat* 10:1779–1812
- Lindsay BG (1988) Composite likelihood methods. *Contemp Math* 80:221–39
- Matias C, Miele V (2017) Statistical clustering of temporal networks through a dynamic stochastic block model. *J Roy Stat Soc B* 79:1119–1141
- Menard S (2002) *Longitudinal research*, 2nd edn. Sage, Newbury Park, Canada
- Pakel C, Shephard N, Shephard K, Engle RF (2021) Fitting vast dimensional time-varying covariance models. *J Business Econom Stat* 39:652–668
- Perry PO, Wolfe PJ (2013) Point process modelling for directed interaction networks. *J Roy Stat Soc B* 75:821–849
- Rasch G (1967) An informal report on a theory of objectivity in comparisons. In: Van der Kamp L, J. T, Vlek C. A. J (eds) *Psychological measurement theory. Proceedings of the NUFFIC international summer session in science at Het Oude Hof. University of Leyden, Leyden*, pp 1–19
- Robins G, Pattison P (2001) Random graph models for temporal processes in social networks*. *J Math Sociol* 25:5–41
- Skvoretz J, Faust K (1999) Logit models for affiliation networks. *Sociol Methodol* 29:253–280
- Snijders T, van Duijn M (1997) Simulation for statistical inference in dynamic network models. In: *Simulating social phenomena*, pp. 493–512. Springer
- Snijders TA (2011) Statistical models for social networks. *Ann Rev Sociol* 37:131–153
- Snijders TA, Van de Bunt GG, Steglich CE (2010) Introduction to stochastic actor-based models for network dynamics. *Soc Netw* 32:44–60
- Stadtfeld C, Hollway J, Block P (2017) Dynamic network actor models: Investigating coordination ties through time. *Sociol Methodol* 47:1–40
- Varin C, Reid N, Firth D (2011) An overview of composite likelihood methods. *Statistica Sinica* 21:5–42
- Vu DQ, Hunter DR, Schweinberger M (2013) Model-based clustering of large networks. *Ann Appl Stat* 7:1010–1039
- Wang P, Sharpe K, Robins GL, Pattison PE (2009) Exponential random graph (p^*) models for affiliation networks. *Soc Netw* 31:12–25
- Wasserman S, Pattison P (1996) Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^* . *Psychometrika* 61:401–425
- Xia D, Mankad S, Michailidis G (2016) Measuring influence of users in twitter ecosystems using a counting process modeling framework. *Technometrics* 58:360–370
- Yang T, Chi Y, Zhu S, Gong Y, Jin R (2011) Detecting communities and their evolutions in dynamic social networks: A Bayesian approach. *Mach Learn* 82:157–189