Regression of directed graphs on independent effects for density and reciprocity

Zijlstra, B.J.H.

Published in:
Journal of Mathematical Sociology

DOI:
10.1080/0022250X.2017.1387858

Creative Commons License (see https://creativecommons.org/use-remix/cc-licenses):
CC BY-NC-ND

Citation for published version (APA):
Regression of directed graphs on independent effects for density and reciprocity

Bonne J.H. Zijlstra

To cite this article: Bonne J.H. Zijlstra (2017) Regression of directed graphs on independent effects for density and reciprocity, The Journal of Mathematical Sociology, 41:4, 185-192, DOI: 10.1080/0022250X.2017.1387858

To link to this article: https://doi.org/10.1080/0022250X.2017.1387858

© 2017 The Author(s). published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 25 Oct 2017.

Submit your article to this journal

Article views: 209

View related articles

View Crossmark data
Regression of directed graphs on independent effects for density and reciprocity

Bonne J.H. Zijlstra

Child Development and Education, Universiteit van Amsterdam Faculteit der Maatschappij- en Gedragswetenschappen, Amsterdam, Netherlands

ABSTRACT

In common models for dyadic network regression, the density and reciprocity parameters are dependent on each other. Here, the $j_1$ and $j_2$ models are introduced with a density parameter that represents the log odds of a single tie. Consequently, the density and reciprocity parameters are independent and the interpretation of both parameters more straightforward. Estimation procedures and simulation results for these new models are discussed as well as an illustrative example.

KEYWORDS
MCMC; network density; network reciprocity; network regression; social networks; social support

1. Introduction

In social network analysis, a network is commonly represented by a directed graph, or digraph, reflecting presence and absence of social relations in a given set of actors. These can be partitioned into dyads, representing pairs of actors and their relations. In a null dyad, actors are mutually unconnected, an asymmetric dyad consists of a directed tie from an actor to another, which is not returned by the other, in a symmetric or reciprocal dyad both actors send a tie to each other.

One of the first statistical models for directed graphs was the $p_1$ model by Holland and Leinhardt (1981). It estimates a density and a reciprocity parameter from the data. The density parameter models the odds of an asymmetric dyad versus a null dyad and is subdivided into an overall estimate and sender and receiver effects. The reciprocity parameter models the ratio between the odds of a symmetric dyad versus an asymmetric dyad and the odds of an asymmetric dyad versus a null dyad. The $p_2$ model (Van Duijn 1995; Van Duijn, Snijders and Zijlstra, 2004) is a successor to the $p_1$ model. It enables the inclusion of predictor variables through the introduction of random sender and receiver parameters. Predictors can be included on the actor level for sender and receiver effects and at the dyad level for density and reciprocity.

The $p_1$ and the $p_2$ models are based on a multinomial representation of the four possible dyadic outcomes. Consequently, the modeled probability of one outcome affects the probability of other outcomes and the density and reciprocity parameters are dependent on each other. This complicates the interpretation of both.

In this paper, the $j_1$ and $j_2$ models will be presented, in which the density parameter reflects the odds of a single tie, resulting in independent density and reciprocity parameters. Since these parameters are unconfounded, their interpretation is more straightforward. In addition, the density parameter is now comparable to parameters in more familiar models like logistic regression.

The $j_2$ model has random parameters for sending and receiving ties (activity and popularity). This sets it apart from the well-known class of exponential random graph models (ERGMs) (Robins et al. 2007). Even though both can be expressed as exponential models, ERGMs are more focused on, and...
very flexible in, modeling structure and predicting subgraphs. The $j_2$ model specifically aims at predicting ties.

In the remainder of this paper, the $p_1$ model will be further explained and the $j_1$ and $j_2$ models will be introduced in Section 2. Since estimates of the reciprocity parameter in the $j_2$ model are relatively likely to become unstable, specific adaptations to the estimation procedure are discussed in Section 3. This is followed by simulation results in Section 4. An illustrative example using empirical data is presented in Section 5, contrasting the $j_2$ and $p_2$ models. Finally, concluding remarks are given in Section 6.

2. Models

2.1 The $p_1$ model

The $p_1$ model (Holland and Leinhardt, 1981) describes the probabilities of two directed dichotomous tie observations in dyads of actors $i$ and $j$ as

$$
P(Y_{ij} = 0, Y_{ji} = 0) = 1/k_{ij}
$$

$$
P(Y_{ij} = 1, Y_{ji} = 0) = \{\exp(\theta_{ij})\}/k_{ij}
$$

$$
P(Y_{ij} = 0, Y_{ji} = 1) = \{\exp(\theta_{ji})\}/k_{ij}
$$

$$
P(Y_{ij} = 1, Y_{ji} = 1) = \{\exp(\theta_{ij} + \theta_{ji} + \rho_{ij})\}/k_{ij},
$$

where $Y_{ij}$ is a directed tie from actor $i$ to actor $j$ taking value 1 in the presence of a tie and 0 in the absence, and

$$
k_{ij} = 1 + \exp(\theta_{ij}) + \exp(\theta_{ji}) + \exp(\theta_{ij} + \theta_{ji} + \rho_{ij}) \text{ for all } i < j,
$$

$$
\theta_{ij} = \theta + \alpha_i + \beta_j \text{ for all } i < j,
$$

$$
\theta_{ji} = \theta + \alpha_j + \beta_i \text{ for all } i < j,
$$

$$
\rho_{ij} = \rho \text{ for all } i \neq j.
$$

In Equation (1), $\theta_{ij}$ is the density parameter for a specific dyad, solving for which yields

$$
\exp(\theta_{ij}) = \frac{P(Y_{ij} = 1, Y_{ji} = 0)}{P(Y_{ij} = 0, Y_{ji} = 0)},
$$

the odds of an asymmetric dyad versus a null dyad. In Equation (2), $\theta_{ij}$ is substituted by $\theta$ for the overall log odds and $\alpha_i$ and $\beta_j$ reflecting individual deviations in sending and receiving ties in an asymmetric dyad for actors $i$ and $j$, respectively. The sender and receiver effects $\alpha$ and $\beta$ sum to zero. Since an asymmetric tie never involves the same actor as sender and receiver, the sender and receiver effect tend to have a (slightly) negative relation.

Solving Equation (1) for $\rho_{ij}$, the reciprocity parameter for a specific dyad, results in

$$
\exp(\rho_{ij}) = \frac{P(Y_{ij} = 1, Y_{ji} = 1)P(Y_{ij} = 0, Y_{ji} = 0)}{P(Y_{ij} = 1, Y_{ji} = 0)P(Y_{ij} = 1, Y_{ji} = 0)P(Y_{ij} = 0, Y_{ji} = 0)P(Y_{ij} = 0, Y_{ji} = 0)} = \frac{P(Y_{ij} = 0, Y_{ji} = 0)P(Y_{ij} = 1, Y_{ji} = 1)}{P(Y_{ij} = 1, Y_{ji} = 0)P(Y_{ij} = 0, Y_{ji} = 1)},
$$

the ratio of the odds of a reciprocal dyad versus an asymmetric dyad and the odds of an asymmetric dyad versus a null dyad. In Equation (2), this ratio is substituted by the overall estimate $\rho$.

An implication of the multinomial representation of the four possible outcomes for the dyads in Equation (1) is that a change in the modeled probability of one outcome affects the probability of other outcomes. As a result, the parameters $\theta$ and $\rho$ in the $p_1$ model (Holland and Leinhardt, 1981) are dependent on each other. This can also be observed by recognizing that the denominator in the expression for $\rho_{ij}$ (Equation (2)) equals the expression for $\theta_{ij}$ (Equation (3)). A visual representation of the dependence between $\theta_{ij}$ and $\rho_{ij}$ for different values of the probability of a single tie, $P(Y_{ij} = 1) = p$, is given in Figure 1.
As Figure 1 illustrates, in the $p_1$ model the parameters $\theta_{ij}$ and $\rho_{ij}$ have differently shaped negative relations for different values of the probability of a tie $P(Y_{ij} = 1) = p$. Both $\theta_{ij}$ and $\rho_{ij}$ are also related to $p$ since for larger values of $p$ both increase. When $\rho_{ij} = 0$, $\theta_{ij}$ equals the log odds of $p$. For values $\rho_{ij} > 0$, $\theta_{ij}$ is larger than the log odds of $p$ while the opposite holds for $\rho_{ij} < 0$.

2.2 The $j_1$ model

The new $j_1$ model applies probabilities to the different dyadic outcomes through the density parameter $m_{ij}$ and a parameter $c_{ij}$ which is indirectly used to model the reciprocity:

\[
\begin{align*}
P(Y_{ij} = 0, Y_{ji} = 0) &= (1 + c_{ij})/k_{ij} \\
P(Y_{ij} = 1, Y_{ji} = 0) &= \exp(m_{ij}) - c_{ij}/k_{ij} \\
P(Y_{ij} = 0, Y_{ji} = 1) &= \exp(m_{ji}) - c_{ij}/k_{ij} \\
P(Y_{ij} = 1, Y_{ji} = 1) &= \exp(m_{ij} + m_{ji}) + c_{ij}/k_{ij},
\end{align*}
\]

(5)

with

\[
k_{ij} = 1 + \exp(m_{ij}) + \exp(m_{ji}) + \exp(m_{ij} + m_{ji}).
\]

The density parameter $m_{ij}$ models the log odds of a single tie, which can be seen by solving Equation (5) for $m_{ij}$,

\[
\exp(m_{ij}) = \frac{P(Y_{ij} = 1)}{P(Y_{ij} = 0)}.
\]

(6)

The reciprocity parameter in the $j_1$ model, $r_{ij}$, is similar to the expression for $\rho_{ij}$ in the $p_1$ model (Equation (4)),

\[
\exp(r_{ij}) = \frac{P(Y_{ij} = 0, Y_{ji} = 0)P(Y_{ij} = 1, Y_{ji} = 1)}{P(Y_{ij} = 1, Y_{ji} = 0)P(Y_{ij} = 0, Y_{ji} = 1)}.
\]

(7)

The values for $c_{ij}$ in Equation (5) and thus the probabilities for the dyadic outcomes can be determined after solving $r_{ij}$ for $c_{ij}$ as
\[
c_{ij} = \left( -\frac{1 + \exp(m_{ij} + m_{ji}) + \exp(m_{ij} + r_j) + \exp(m_{ji} + r_j) - \exp(m_{ij} + m_{ji})}{2(\exp(r_j) - 1)} \right) \]

In the \( j_1 \) model, \( P(Y_{ij} = 1) \) only depends on \( m_{ij} \) and consequently is independent of \( c_{ij} \). Any differences in \( c_{ij} \) as a result of differences in \( r_j \) are subsequently also independent of \( m_{ij} \). In summary, the \( j_1 \) model has a density parameter \( m_{ij} \) for the log odds of \( Y_{ij} \) and an independent reciprocity parameter \( r_j \) for the log odds ratio (Equation (7)).

### 2.3 The \( j_2 \) model

The \( p_2 \) model (Van Duijn 1995; Van Duijn et al., 2004) took a major step forward by extending the \( p_1 \) model (Holland and Leinhart, 1981) to allow predictor variables for the sender and receiver effects and for density and reciprocity. This was achieved by introducing random sender and receiver effects, freeing up parameter space for the coefficients of the predictors. The \( j_2 \) model extends on the \( j_1 \) model analogously to the extension of the \( p_2 \) model on the \( p_1 \) model.

The density parameter \( m_{ij} \) is substituted by random sender and receiver effects \( A_i \) and \( B_j \) and \( m \) an overall parameter for the log odds of \( Y_{ij} \). The random effects are assumed to have a bivariate normal distribution with variances \( \sigma_A^2, \sigma_B^2 \) and covariance \( \sigma_{AB} \). The density parameter \( m_{ij} \) is further substituted by the regression parameters \( \gamma_1, \gamma_2, \) and \( \gamma_3 \) for sender, receiver, and density predictors \( X_{1i}, X_{2j}, \) and \( X_{3ij} \), respectively,

\[
m_{ij} = A_i + B_j + m + \gamma_1 X_{1i} + \gamma_2 X_{2j} + \gamma_3 X_{3ij}.
\]

The sender and receiver predictors are variables at the actor level and the density predictors are variables at the dyad level.

The parameter for reciprocity \( r_j \) (Equation (7)) is substituted by an overall estimate of reciprocity \( r \) and regression coefficients \( \gamma_4 \) for symmetric dyadic predictors, \( X_{4ij}, \)

\[
r_{ij} = r + \gamma_4 X_{4ij}.
\]

### 3. Estimation and prior distributions for the \( j_2 \) model

In logistic regression, estimates can become highly unstable due to a phenomenon called separation (Albert and Anderson 1984). This stems from sparse data being (nearly) identified for one of the outcomes. This is also observed in the \( j_1 \) and \( j_2 \) models, in particular with the extended specification of the reciprocity parameters \( r \) and \( \gamma_4 \) in the \( j_2 \) model. Since these are independent of the probability of a tie \( P(Y_{ij} = 1) \), they are relatively unstable. One way to deal with this problem is to apply an appropriate prior distribution in Bayesian simulation methods. For this reason the \( j_2 \) model is estimated with an Markov chain Monte Carlo (MCMC) algorithm (see e.g. Gelman et al. 2014), subsequently sampling the random effects \((A,B)\), their covariance matrix \( \Sigma \), and the regression parameters \( m, r, \gamma_1, \gamma_2, \gamma_3, \gamma_4 \). The prior distribution applied to \( \Sigma \) is an inverse Wishart distribution. Since the \( j_2 \) model closely resembles logistic regression for the parameters \( m, \gamma_1, \gamma_2, \) and \( \gamma_3 \) in \( m_{ij} \), their prior distribution is based on assuming half an observation for each outcome in logistic regression, as suggested by Gelman et al. (2008). This results in the likelihood

\[
\left( \frac{\sigma^2}{1 + \sigma^2} \right)^{\frac{1}{2}} \left( \frac{1}{1 + \sigma^2} \right)^{\frac{1}{2}} = \frac{\sigma^2}{1 + \sigma^2},
\]

closely resembling a \( t \)-distribution with 7 degrees of freedom and scale 2.5 (see Figure 2), which was taken as the actual prior distribution. In a similar manner, a prior distribution for the parameters \( r \)
and $\gamma_4$ in $r_{ij}$ was derived. However, the prior based on assuming half an observation for all dyads still resulted in unstable estimates. Therefore, the prior distribution was based on the likelihood assuming one observation for each dyad (and $m_{ij} = 0$),

$$P(Y_{ij} = 0, Y_{ji} = 0)P(Y_{ij} = 1, Y_{ji} = 1) = \frac{e^r}{16(1 + e^r)^4}.$$  

This distribution is close to a $t$-distribution with 7 degrees of freedom with scale 2 (see Figure 2), which was applied as prior distribution.

For the regression parameters $\gamma_1$, $\gamma_2$, $\gamma_3$, and $\gamma_4$, the above priors were further adjusted to the variation in the predictors through dividing the scale of the $t$-distributions by the observed standard deviation of $X_1$, $X_2$, $X_3$, and $X_4$. Following Raftery (1996) and Gelman (2008), the predictors were centered to an average of zero to assure the center of the prior distributions can be set at zero as well.

4. Simulations

To verify the effectiveness of the prior distributions against separation, simulations were run. Data were simulated according to model 3 from Zijlstra et al. (2009), containing predictors at the actor level (for receiving) as well as at the dyad level (for density and reciprocity). The variance of the random sender effects ($\sigma_A^2 = 1.5$) was taken twice as large as the variance of the random receiver effects ($\sigma_B^2 = .75$) with a negative covariance ($\sigma_{AB} = -.5$). The density and reciprocity were $m = -1.43$ and $r = 2$, corresponding to $P(Y_{ij} = 1) = .193$. A receiver effect with coefficient $\gamma_2 = -.1$ was added for a predictor drawn from a binary (0,1) distribution with equal probabilities. There are two density predictors; $\gamma_{31} = .2$ is the coefficient for a predictor drawn from a multinomial distribution with five outcomes (1, 2, 3, 4, 5) with equal probability and $\gamma_{32} = .5$ is the coefficient for a random digraph with $P(Y = 1) = .20$. Finally, the coefficient for the predictor for reciprocity, $\gamma_4$, represents the effect of the absolute differences of the multinomial outcomes used as a predictor for the density. Different values for $\gamma_4$ were applied. A very small value of .05 was based on Zijlstra et al. (2009). To have a more complete test for separation, a substantial value of $\gamma_4 = 1$, implying a predicted range of 4 for the log of Equation (4), and a large value $\gamma_4 = 2$, implying a predicted range of 8 for the log of Equation (4), were used as well. All models were simulated 1,000 times for relatively small networks of 20 actors and medium-sized networks of 40 actors. The $t$-distribution with $df = 7$ was used as prior, with scales as described in Section 3. Results are presented in Table 1. Here, only $\gamma_4$ is reported for the models with simulated values $\gamma_4 = 1$ and $\gamma_4 = 2$ because results for the other parameters were highly similar to those in the model with $\gamma_4 = .05$.

For the regression parameters, all coverage rates in Table 1 are relatively close to the nominal values of 95 and 99. This indicates that even for small networks of 20 actors the prior distributions for the regression parameters yield acceptable intervals for inference, avoiding separation. However, for the small networks the average point estimates for $r$ and $\gamma_4$ are relatively far from their simulated
values. This appears to be mainly due to the large standard deviations for these parameters. For the medium-sized networks of 40 actors, the standard deviations are smaller and the point estimates are relatively close to their simulated values.

5. Application: reported emotional support between Dutch high school pupils

To further illustrate the $j_2$ model, it was applied to a network from the Dutch Social Behavior Study (Baerveldt and Snijders 1994; Baerveldt et al. 2004). The network was measured on a year group of 62 high school students with 37 boys and 27 girls of ages between 16 and 18 years old. Each student was asked from whom they had received emotional support when they were depressed, for instance, after the end of a relationship or in a conflict with someone else. Two models were estimated. In model 1, a sender and receiver effect for being a boy is included as well as a density effect for (dyads of) different gender. In model 2, a reciprocity effect for different gender is added.

Results are presented in Table 2. For the $j_2$ model, a negative receiver effect is found for being a boy. This indicates that the probability of reporting having received emotional support from boys is smaller compared to girls. There is an additional negative density effect for “different gender,” indicating a smaller probability of receiving emotional support in dyads with different gender. Thus, given the receiver effect of “boy,” receiving emotional support from someone of similar gender is more likely. In model 2, no further effect of different gender on the reciprocity is found.

For $p_2$, estimates for analogue models are presented in Table 2 as well. A similar receiver effect of “boy” and density effect of “different gender” was found on the odds of an asymmetric versus a null dyad. However, in model 2, the density effect of different gender is no longer significant, while there is now a significant negative reciprocity effect (assuming a significance level of 0.05). This is unexpected since the observed log odds ratio for different gender (Equation (2)) is actually higher (6.14 compared to 5.03 for same gender). In the $p_2$ model, such effects are possible since the sender, receiver, density, and reciprocity effects are all dependent on each other and jointly model the observed data (see also Figure 1).

Comparing between models, the standard deviation for the coefficient of the predictor for reciprocity, $\gamma_{41}$, is particularly large in the $j_2$ model. This originates from the reciprocity parameters $\rho$ and $\gamma_4$ in $p_2$ partly modeling the probability $P(Y = 1)$. In contrast, in $j_2$ the reciprocity parameters are independent of $P(Y = 1)$, resulting in less stable estimates.

6. Concluding remarks

With $j_2$, a new model for regression of predictors on dichotomous dyadic network data was introduced. This was compared to the most similar model, the $p_2$ model. Both can be preferred in
different situations. An advantage of \( j_2 \) is that the density and reciprocity parameters are independent, avoiding ambiguous estimates as seen for the \( p_2 \) model in the application in Section 5. Besides, the parameter for density in the \( j_2 \) model has a familiar interpretation as the log odds of a tie, comparable to logistic regression. On the other hand, when there is theoretical interest in non-reciprocated ties, the \( p_2 \) model seems to be most appropriate because the density parameter in \( p_2 \) estimates the odds of asymmetric dyads.

A property of the \( j_2 \) model that requires further investigation is that the estimation of the reciprocity parameter is unstable because of the independence from the density parameter. With the introduction of specific prior distributions, a practical solution to this problem was presented. However, an analytical solution may further improve the estimates.

For the \( p_1 \) and \( p_2 \) model, the dependence between the density and reciprocity also requires further investigation. Even though it applies the same expression for reciprocity (Equation (4)) as the \( j_1 \) and \( j_2 \) model (Equation (7)), it is not yet clear how specifically the interpretation is affected by its dependence on density. Because the reciprocity parameter in the \( p_1 \) and \( p_2 \) model is related to the probability of a single tie, the frequency of symmetric dyads, that is, (1,1), is modeled to a greater extent by the reciprocity parameter than the frequency of null dyads, that is, (0,0). On the contrary, given the probability of a single tie, the \( j_1 \) and \( j_2 \) model assign equal weight to both null and symmetric dyads. These models may therefore be of particular value in cases where there is substantial theoretical interest in the extent to which both null and symmetric dyads are observed beyond chance.

Software for estimating \( j_2 \) models is freely available from the corresponding author. The R-package (R Core Team 2014) dyads is expected to become available in the near future.

**ORCID**

Bonne J.H. Zijlstra [http://orcid.org/0000-0001-9924-4387](http://orcid.org/0000-0001-9924-4387)

**References**


