Bayesian Analysis (2025) **TBA**, Number TBA, pp. 1–31

The Multiplex p_2 Model: Mixed-Effects Modeling for Multiplex Social Networks

Anni Hong* and Nynke M. D. Niezink†

Abstract. Social actors are often embedded in multiple social networks, and there is a growing interest in studying social systems from a multiplex network perspective. In this paper, we propose a mixed-effects model for cross-sectional multiplex network data that assumes dyads to be conditionally independent. Building on the uniplex p_2 model, we incorporate dependencies between different network layers via cross-layer dyadic effects and through the covariance structure among the actor random effects. These cross-layer effects model the tendencies for ties between two actors and the ties to and from the same actor to be dependent across different relational dimensions. The model can also study the effect of actor and dyad covariates. As simulation-based goodness-of-fit analyses are common practice in applied network studies, we here propose goodness-of-fit measures for multiplex network analyses. We evaluate our choice of priors and the computational faithfulness and inferential properties of the proposed method through simulation. We illustrate the utility of the multiplex p_2 model in a replication study of a toxic chemical policy network. An original study that reflects on gossip as perceived by gossip senders and gossip targets, and their differences in perspectives, based on data from 34 Hungarian elementary school classes, highlights the applicability of the proposed method. The proposed methodology is available in the R package multip2.

Keywords: social network analysis, multiplex networks, conditionally independent dyad model, Bayesian, mixed effects.

1 Introduction

The study of the social behavior of individuals and groups is inseparable from the study of social networks. Social networks can be described by graphs, consisting of nodes, representing social actors (e.g., individuals, countries, organizations), and edges, representing the connections among the actors (e.g., friendship, trade, collaboration). Social actors are often embedded in multiple social networks simultaneously, and the field of social network analysis has long recognized the importance of this multiplexity (Borgatti et al., 2009; Kadushin, 2012). For example, the development of bullying in a classroom cannot be understood without the context of friendships in that classroom (e.g., Rambaran et al., 2020) – the dynamics of positive and negative relations are dependent. Yet, in much applied work, the multiplex network perspective is not leveraged, and single networks are considered in isolation.

*Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA 15213, annihong@andrew.cmu.edu

†Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA 15213, nniezink@andrew.cmu.edu

 https://doi.org/10.1214/25-BA1527

Multiplex networks are structures representing multiple relationships observed among the same group of actors. These relationships can represent heterogeneous edge types, e.g., retweet, follows, and mentions among Twitter users (Greene and Cunningham, 2013), or co-authorship, co-citation, and co-venue relationships among academics (Hmimida and Kanawati, 2015). Multiplex networks also play a major role when studying network perspectives. Self-reported social networks might be imprecise measurements of underlying social behavior, and social actors can have different perspectives on a relationship (e.g., i may experience behavior by j as bullying but j does not think they bully i). Social desirability bias too can drive individuals to adjust their reports (e.g., i reports less gossip than i actually participated in). Tatum and Grund (2020) studied the difference in perspectives between bullies and victims in school classes by aggregating the two networks with different perspectives into a single disagreement network. While this approach allows us to study discrepancies, it does not provide an understanding of the context in which they arose. A multiplex network approach would be a more natural way to study the multiple perspectives (relationships) as well as their discrepancies.

In this paper, we develop a multiplex network model that would allow for such an analysis: a mixed-effects model assuming conditionally independent dyads (CID) for cross-sectional, directed, binary network data. Network dyads are given by pairs of nodes and the configuration of relations among them. Social networks are characterized by dependent relationships, making modeling their structure fundamentally different from that of other types of data. The various existing models for network data can be distinguished by how they handle network dependencies. In CID models, dyads are independent when conditioned on latent variables and other model components (Dabbs et al., 2020). Examples of CID models include the stochastic block model (Holland et al., 1983) and the latent space model (Hoff et al., 2002), which can capture a wide range of complex network dependencies.

We develop our model in the tradition of the p_2 modeling framework (van Duijn et al., 2004), a CID model that represents the probability of a dyad outcome in terms of baseline network density, reciprocity, and actor and dyad heterogeneity. While its predecessor, the p_1 model (Holland and Leinhardt, 1981), models actor heterogeneity with fixed effects representing actors' differential tendencies to send and receive ties (thus giving rise to a large number of parameters for large networks), the p_2 model represents these tendencies by random effects and incorporates covariates to model actor and dyadic heterogeneity. For its interpretability and simplicity, the p_2 model has been widely used, e.g., to study physician communication patterns (Keating et al., 2007) and advice seeking between public schools (Spillane et al., 2015). The multilevel model extension of the p_2 model (Zijlstra et al., 2006) enabled a large number of multi-group network studies (e.g., Veenstra et al., 2007; Vermeij et al., 2009; Tolsma et al., 2013; Smith et al., 2014; Zijlstra et al., 2008).[1]

Here, we extend the p_2 model for multiplex network analysis. The parameters and characteristics of the uniplex p_2 model remain as within-network effects in the multi-

[1] While revising this article for resubmission, we learned that a multivariate extension of the p_2 model was proposed in Chapter 4 of the PhD thesis by Zijlstra (2009).

plex model. Yet, the multiplex model introduces new cross-network density and cross-network reciprocity parameters, modeling the tendencies for ties between two actors to be dependent across different relational dimensions. Actors' random sender and receiver effects for the different relational dimensions are modeled as different, but dependent.

Bellio and Soriani (2021) proposed a maximum likelihood-based approach to p_2 model estimation. However, a robust Bayesian estimation procedure is still lacking. The `R` package `dyads` (Zijlstra, 2021) is the only existing Bayesian implementation of the p_2 model, but has a few limitations. These include its inability to include multiple actor covariates, its lack of convergence checks and goodness-of-fit tests, and its inability to handle missing network data. Accompanying this paper, we developed a Bayesian implementation of the multiplex p_2 model in the `Stan` probabilistic programming framework (Stan Development Team, 2022). This method includes the standard (uniplex) p_2 model as a special case. Our `R` package `multip2` includes multiplex goodness-of-fit measures and the implementation can handle missing network data.

In the following, we will first give a brief overview of existing methods for cross-sectional multiplex network data (Section 1.1) and then present the multiplex p_2 model (Section 2) and its estimation (Section 3). We then introduce goodness-of-fit measures for multiplex networks, which can be used in posterior predictive checks (Section 4). Section 5 presents several simulation studies in which we evaluate our choice of priors, the accuracy of our estimation procedure, and the inferential properties of the method. Sections 6 and 7 present two applications. The first one replicates the analysis by Leifeld and Schneider (2012) of how information is exchanged in policy networks, and illustrates that our multiplex approach offers additional insights into reciprocation patterns across different types of information (political and scientific). The second study analyzes the discrepancies in reports on gossip ties from the viewpoint of the gossiper and the gossip target based on data not previously studied. We find that students are not particularly perceptive when it comes to identifying individuals who gossip about them. However, they tend to gossip about those whom they believe are gossiping about them. We conclude with a discussion in Section 8.

1.1 Multiplex network methods

Several methods have been developed for the analysis of cross-sectional multiplex network data. Network regression techniques such as the multiple regression quadratic assignment procedure (MRQAP; Krackhardt, 1988; Dekker et al., 2007) allow us to model a network as a function of other networks (cf. multiple regression) but do not simultaneously model multiple networks as the outcome (cf. multivariate regression). The multiplex stochastic block model (Barbillon et al., 2017) detects communities with information from multiple network layers, and a recent development allows the number of communities to vary in different layers (Amini et al., 2024) but does not allow for borrowing of actor-specific information across layers. Moreover, these methods are tailored to assign group membership to social actors based on network data, but do not take into account the role of covariate information. The social relations model (SRM) does take this into account and supports the simultaneous analysis of multiple continuous-valued networks (Nestler, 2018). Hoff et al. (2013) includes formulations that extends

SRM binary networks and Redhead et al. (2024) applied a multiplex extension of the SRM for binary network data without covariates to study indirect reciprocity based on network-structured economic games. Multiplex latent space models (Salter-Townshend and McCormick, 2017; Sosa and Betancourt, 2022) can be used to model continuous or binary networks, but differ from the p_2 model in terms of how the probability of a tie is modeled. The p_2 model explicitly models network characteristics such as density and reciprocity while the latent space model assumes that unobserved actor attributes, as represented by latent positions, affect the connectivity pattern. Recent developments in exponential family random graph models (ERGMs) allow for the modeling of two or more binary networks as outcomes simultaneously (Wang et al., 2013, 2016; Chen, 2021). Yet, ERGM estimation can be computationally expensive and model convergence is hard to achieve in practice with increasing network dimensions. The stochastic actor-oriented model (SAOM) was developed to analyze the dynamics of network data (Snijders, 2001), and extended for the coevolution of multiple binary networks (Snijders et al., 2013). Although it is possible to study cross-sectional network data using a stationary SAOM, the applicability and properties of stationary SAOMs, even for uniplex network data, remain unexplored.

2 The multiplex p_2 model

Suppose there are T layers of networks on the same set of n actors. Let $\boldsymbol{M} \in \{0,1\}^{n \times n \times T}$ denote the directed binary multiplex network among these actors. We assume that this network has no self-loops (i.e., i can not send ties to themselves; $M_{iit} = 0$ for all i and t). We denote the p dyadic covariates (e.g., absolute age difference) by $\boldsymbol{Z} \in \mathbb{R}^{n \times n \times p}$ and the q actor covariates (e.g., gender) by $\boldsymbol{X} \in \mathbb{R}^{n \times q}$. In the following, we treat p and q as the same for all effects, without loss of generality. If we let $m_{ij}^t = 1$ denote that there is an edge from i to j in layer t, and $m_{ij}^t = 0$ if there is no such edge, we can represent the (multiplex) outcome on dyad $\{i,j\}$ by $M_{\{ij\}} = \{m_{ij}^1, m_{ji}^1, \ldots, m_{ij}^T, m_{ji}^T\}$. Since the p_2 model is a conditionally independent dyads model, we can decompose the probability of a network $\boldsymbol{M}$ into the probabilities of the dyad outcomes $M_{\{ij\}}$.

Within each layer, the multiplex p_2 model accounts for the overall propensity of ties by a density parameter, and for the likelihood of a tie being reciprocated between two actors by a reciprocity parameter. We denote the density and reciprocity parameters for layer t as μ^t and ρ^t, and let $\boldsymbol{\mu} = \{\mu^1, \ldots, \mu^T\}$ and $\boldsymbol{\rho} = \{\rho^1, \ldots, \rho^T\}$. The model also allows us to evaluate the effects of network covariates on density and reciprocity. If we denote the corresponding coefficients by $\boldsymbol{\delta}_\mu^t \in \mathbb{R}^p$ and $\boldsymbol{\delta}_\rho^t \in \mathbb{R}^p$, the total density and reciprocity effects for dyad $\{i,j\}$ in layer t are

$$\begin{aligned} \mu_{ij}^t &= \mu^t + (\boldsymbol{Z}_{ij})^\top \boldsymbol{\delta}_\mu^t, \\ \rho_{ij}^t &= \rho^t + (\boldsymbol{Z}_{ij})^\top \boldsymbol{\delta}_\rho^t. \end{aligned} \tag{1}$$

Across layers, we can model the tendency for one actor to establish multiple connections to another actor (e.g., i considers j a friend *and* a collaborator) by a cross-layer density effect. We can also capture cross-layer reciprocity – the tendency for actors to reciprocate an incoming tie of one kind (e.g., political information) with an outgoing tie of another kind (e.g., scientific information). We let $\mu_{\text{cross}}^{(t,s)}$ denote the baseline cross-layer density effect for layers s and t and $\rho_{\text{cross}}^{(t,s)}$ the baseline cross-layer reciprocity, with $\boldsymbol{\mu}_{\text{cross}}$ and $\boldsymbol{\rho}_{\text{cross}}$ containing the cross-layer density and cross-layer reciprocity parameters for all pairs of layers. The cross-network density and reciprocity too can depend on dyadic covariates, with corresponding coefficients $\boldsymbol{\delta}_{\mu_{\text{cross}}}^{(t,s)} \in \mathbb{R}^p$ and $\boldsymbol{\delta}_{\rho_{\text{cross}}}^{(t,s)} \in \mathbb{R}^p$. In the model, the total cross-layer density and reciprocity effects for dyad $\{i, j\}$ in layers s and t are thus given by

$$\begin{aligned} \mu_{\text{cross},ij}^{(t,s)} &= \mu_{\text{cross}}^{(t,s)} + (\boldsymbol{Z}_{ij})^\top \boldsymbol{\delta}_{\mu_{\text{cross}}}^{(t,s)}, \\ \rho_{\text{cross},ij}^{(t,s)} &= \rho_{\text{cross}}^{(t,s)} + (\boldsymbol{Z}_{ij})^\top \boldsymbol{\delta}_{\rho_{\text{cross}}}^{(t,s)}. \end{aligned} \tag{2}$$

The p_2 model captures actors' differential tendency to send ties by random sender effects, and their differential tendency to receive ties by random receiver effects. The tendencies for actors to send and receive ties can depend on individual actor characteristics. We define different actor sender and receiver effects for each layer, and allow actors' tendencies to send and to receive ties to be correlated across relationships. For example, if a person sends texts to a lot of people, they may also send emails to a lot of people. Yet, we assume that one actor's tendency to send or receive ties does not affect this tendency of the other actors – the random effects are correlated within but not between individuals. Formally, we let α_i^t and β_i^t denote the tendencies for actor i to send and receive ties in network layer t, respectively. The sender and receiver effects for all actors in layer t are $\boldsymbol{\alpha}^t = \{\alpha_1^t, \ldots, \alpha_n^t\}$ and $\boldsymbol{\beta}^t = \{\beta_1^t, \ldots, \beta_n^t\}$. Let $\boldsymbol{C}_i = [A_i^1, \ldots, A_i^T, B_i^1, \ldots, B_i^T]^\top \in \mathbb{R}^{2T}$ denote the random actor effects for actor i in the T network layers. We assume that $\boldsymbol{C}_1, \ldots, \boldsymbol{C}_n$ are an i.i.d. sample from $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{AB}})$. The actor sender and receiver effects are given by

$$\begin{aligned} \boldsymbol{\alpha}^t &= \boldsymbol{X}\boldsymbol{\gamma}_\alpha^t + \boldsymbol{A}^t, \\ \boldsymbol{\beta}^t &= \boldsymbol{X}\boldsymbol{\gamma}_\beta^t + \boldsymbol{B}^t, \end{aligned} \tag{3}$$

where $\boldsymbol{A}^t = [A_1^t, \ldots, A_n^t]^\top$ and $\boldsymbol{B}^t = [B_1^t, \ldots, B_n^t]^\top$. The coefficients $\boldsymbol{\gamma}_\alpha^t \in \mathbb{R}^q$ and $\boldsymbol{\gamma}_\beta^t \in \mathbb{R}^q$ represent the effects of the individual-specific covariates $\boldsymbol{X}$ on actors' tendencies to send and receive ties.

Now we can bring all the parts together and define the probability function on a dyad in a T-plex network. If we define

$$\begin{aligned} K_{ij}(M_{\{i,j\}}) = \sum_{t=1}^{T} &\Big(m_{ij}^t(\mu_{ij}^t + \alpha_i^t + \beta_j^t) + m_{ji}^t(\mu_{ji}^t + \alpha_j^t + \beta_i^t) + m_{ij}^t m_{ji}^t \rho_{ij}^t \Big) \\ &+ \sum_{t,s=1,\ldots,T\,:\,t<s} (m_{ij}^t m_{ij}^s + m_{ji}^t m_{ji}^s)\, \mu_{\text{cross},ij}^{(t,s)} \\ &+ \sum_{t,s=1,\ldots,T\,:\,t<s} (m_{ij}^t m_{ji}^s + m_{ji}^t m_{ij}^s)\, \rho_{\text{cross},ij}^{(t,s)}\,, \end{aligned} \tag{4}$$

the probability of outcome $M_{\{i,j\}}$ on dyad $\{i, j\}$ is given by

$$P(M_{\{i,j\}}) = \frac{\exp\{K_{ij}(M_{\{i,j\}})\}}{\sum_{G_{\{i,j\}} \in \{0,1\}^{2T}} \exp\{K_{ij}(G_{\{i,j\}})\}}, \tag{5}$$

where the denominator sums over all possible realizations $G_{\{i,j\}}$ of the T relationships on this dyad.

3 Model estimation

The iterated generalized least-squares procedure (van Duijn et al., 2004) was one of the earliest proposed methods for parameter estimation in the p_2 model. Subsequently, Zijlstra et al. (2009) explored several Markov chain Monte Carlo (MCMC) estimation procedures for the p_2 model, two of which used random walk proposals. More recently, Bellio and Soriani (2021) proposed a maximum likelihood estimation procedure based on the Laplace approximation.

In this study, we employ Hamiltonian Markov chain Monte Carlo (HMCMC) (Duane et al., 1987; Neal, 2011; Betancourt and Stein, 2011) for estimating the parameters of the multiplex p_2 model. HMCMC significantly improved the sampling performance compared to random walk Metropolis in situations where hierarchical structures, such as random-effect models, induce correlations between global and local parameters (Betancourt and Girolami, 2015). The advantage of HMCMC lies in its ability to effectively explore the parameter space by leveraging the local curvature of the target distribution. However, HMCMC requires careful fine-tuning, including specifying the derivative of the target distribution. To address this, we utilize the robust implementation of HMCMC provided by the `Stan` probabilistic programming language (Carpenter et al., 2017; Stan Development Team, 2022). By leveraging the HMCMC algorithm in `Stan` through the `R` package `rstan` (Stan Development Team, 2023a), we can effectively handle the structure of the multiplex p_2 model.

Next, we present the likelihood function and the prior distributions for the multiplex p_2 model. There are 2^{2T} possible outcomes on a dyad in a T-plex network. Dyadic outcome $M_{\{i,j\}}$ follows a categorical logit distribution, as defined in equation (5). We assume normal priors with mean 0 and variance 100 for the baseline within-network parameters $\boldsymbol{\mu}^t$ and $\boldsymbol{\rho}^t$ and cross-network parameters $\boldsymbol{\mu}_{cross}^{(t,s)}$ and $\boldsymbol{\rho}_{cross}^{(t,s)}$,

$$\boldsymbol{\mu}^t, \boldsymbol{\rho}^t, \boldsymbol{\mu}_{cross}^{(t,s)}, \boldsymbol{\rho}_{cross}^{(t,s)} \sim \mathcal{N}(0, 100). \tag{6}$$

This is a weakly informative prior. Since parameters in the p_2 model should be interpreted on a log odds scale, parameter values of, for example, $+10$ and -10 correspond to very large and small probabilities, respectively (Zijlstra et al., 2009). We set the priors for the fixed network and actor effects of the covariates to

$$\theta \sim \mathcal{N}\left(0, \frac{100}{\sigma(\boldsymbol{g}_\theta)^2}\right), \tag{7}$$

where $\theta \in \mathbb{R}$ denotes any entry of the parameter vectors $\boldsymbol{\delta}_\mu$, $\boldsymbol{\delta}_\rho$, $\boldsymbol{\delta}_{\mu_{\text{cross}}}$, $\boldsymbol{\delta}_{\rho_{\text{cross}}}$, $\boldsymbol{\gamma}_\alpha$, and $\boldsymbol{\gamma}_\beta$, and $\sigma(\boldsymbol{g}_\theta)^2$ denotes the sample variance of the corresponding covariate $\boldsymbol{g}_\theta$. This reflects the belief that the magnitude $|\theta \cdot g_\theta|$ of the effect of most covariates is likely to be less than 10 on the log odds scale. For a covariate with a larger sample variance, we thus set the prior variance of the associated parameter smaller to bound the combined effect. However, if this does not reflect the reality of a particular data context, analysts can freely change the prior variance of the covariate parameters (as well as the priors for other parameters) using our `multip2` package in `R`.

In the multiplex p_2 model, the random effect $\boldsymbol{C}_i \in \mathbb{R}^{2T}$ follows a multivariate Gaussian distribution. There are several potential choices for the prior of the variance-covariance matrix $\boldsymbol{\Sigma_{AB}}$. It is common to assume an inverse Wishart distribution, as was done by Zijlstra et al. (2009) for the univariate p_2 model, because of its conjugate property. However, many issues of the inverse Wishart distribution have been pointed out; most notably, the variance and the correlation parameters are correlated (e.g., Akinc and Vandebroek, 2018; Tokuda et al., 2011; Liu et al., 2016). Since `Stan` does not require the use of conjugate priors, we will use the strategy first proposed by Barnard et al. (2000) and studied by Akinc and Vandebroek (2018) and Tokuda et al. (2011). That is, we first decompose the covariance matrix $\boldsymbol{\Sigma_{AB}}$ into a vector of coefficient scales $\boldsymbol{\sigma}$ and a correlation matrix $\boldsymbol{\Omega}$,

$$\boldsymbol{\Sigma_{AB}} = \text{diag}(\boldsymbol{\sigma}) \times \boldsymbol{\Omega} \times \text{diag}(\boldsymbol{\sigma}), \tag{8}$$

and let the scale and correlation components have different, independent priors. We set the prior of the elements of the scale vector to an inverse gamma distribution, and let the correlation matrix follow an Lewandowski-Kurowicka-Joe (LKJ) distribution (Lewandowski et al., 2009) with shape parameter $\eta = 2$,

$$\begin{aligned} \sigma_i &\sim \text{InverseGamma}(\alpha = 3, \beta = 50) \\ \boldsymbol{\Omega} &\sim \text{LJKCorrelation}(\eta = 2). \end{aligned} \tag{9}$$

In the LKJ distribution, $\eta = 1$ generates a uniform correlation distribution. For $\eta > 1$, the distribution puts more mass on the unit matrix, representing less correlation. For $\eta < 1$, mass concentrates away from the unit matrix thus favoring matrices with more correlation. We let η be slightly larger than 1 (i.e., $\eta = 2$), because most of the weight should be on the unit matrix but some correlation among actors' roles across networks in terms of sending and receiving ties is expected. We picked the hyperparameters $\alpha = 3$ (shape) and $\beta = 50$ (scale) for the priors of σ_i based on the results of the prior predictive checks in Section 5.

Prior predictive checks To determine whether a model with a given set of priors on the parameters is consistent with domain expertise, we conduct prior predictive checks, examining if the simulated data generated from the priors is representative of the expected data. We analyze three key social network statistics: density, reciprocity, and transitivity. Network density refers to the proportion of potential ties that are actually present while reciprocity refers to the proportion of ties that are reciprocated. Similar to density and reciprocity, transitivity is another key network statistic, which refers to the proportion of nodes with a common connection that are also connected. Reciprocity and transitivity can capture the human tendencies to reciprocate positive ties (e.g., liking) and to operate in small groups (Robins, 2015). All three statistics are bounded between 0 and 1 and have context-dependent expected ranges. For example, friendship networks often have high reciprocity and transitivity, while networks with

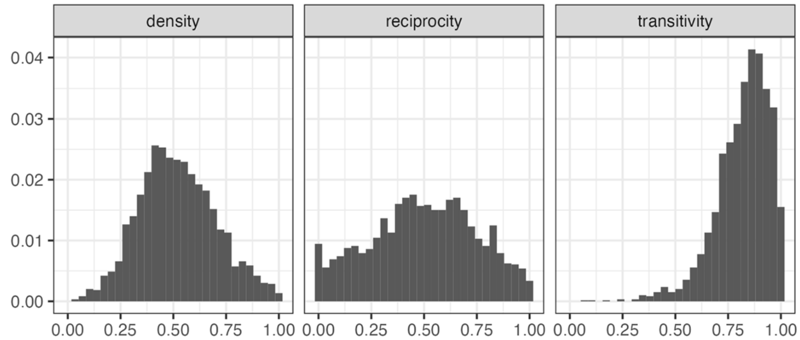


Figure 1: Prior predictive checks: density, reciprocity, and transitivity calculated on 1000 networks simulated based on the priors defined in Section 3.

implicit hierarchies, such as advice-seeking, tend to have low reciprocity. Density tends to decrease with increasing network size. We aim to choose weakly informative priors that incorporate this prior knowledge while allowing the data to drive the inference.

Figure 1 shows the prior predictive check results based on 1000 network draws from the prior defined above. The simulated density and reciprocity cover a wide range of values with peaks at 0.5, coinciding with the zero means of the priors for the density and reciprocity parameters μ and ρ. The simulated transitivity skews higher than typical social networks (0.3 to 0.6 is typical in, e.g., friendship networks), but it still covers a wide range of values. Note also that this statistic is not explicitly modeled by the p_2 model. Overall, the prior predictive checks yield adequate results, especially when we compare them to the results obtained with the inverse Wishart prior, which used to be the default for the p_2 model (Zijlstra et al., 2009). For the prior predictive check results using the inverse Wishart prior, see Figure 1 of the Supplementary Material (Hong and Niezink, 2025).

Post-sweeping of random effects Without additional constraints, the parameters in the mixed-effects model given in equation (5) are only weakly identifiable, in the sense that the parameters are correlated within MCMC chains (Ogle and Barber, 2020). In particular, a change in the within-layer density parameters μ^t can be offset by adding a constant to each of the random effects terms α_i^t and β_i^t. We resolve this issue by a procedure referred to by Ogle and Barber (2020) as "post-sweeping of random effects". This solution retains the original model parameterization, but calculates identifiable quantities after each iteration. These quantities are the ones we report in this paper and use to assess convergence. We compute the identifiable quantities as

$$\begin{aligned} \alpha_i^{t,\text{PS}} &= \alpha_i^t - \bar{\alpha}^t, \\ \beta_i^{t,\text{PS}} &= \beta_i^t - \bar{\beta}^t, \\ \mu^{t,\text{PS}} &= \mu^t + \bar{\alpha}^t + \bar{\beta}^t, \end{aligned} \tag{10}$$

where $\bar{\alpha}^t$ and $\bar{\beta}^t$ are the average random sender and receiver effects for layer t. The identifiable density parameter is obtained by adding the means of both the non-identifiable sender and receiver random effects to μ^t.

4 Goodnesss-of-fit on multiplex networks

A useful model should capture the key structure of the data-generating process such that the fitted model generates data similar to the observed data. In this section, we discuss existing methods for assessing uniplex network model fit and propose measures for multiplex goodness-of-fit analysis. We illustrate these methods in the application in Section 6.

Uniplex goodness-of-fit Simulation-based goodness-of-fit methods are widely used in social network research to evaluate if a model has captured the characteristics of the observed networks (Hunter et al., 2008), and have been adapted to the Bayesian framework in the form of posterior predictive checks (Caimo and Friel, 2011). In the latter, the observed network data are compared to a set of networks simulated from the posteriors of the parameters, on key network statistics such as the indegree and outdegree distribution, and the triad census (Leinhardt, 1971; Holland and Leinhardt, 1970). The indegree and outdegree distribution are the distribution of actors' numbers of incoming and outgoing ties, respectively. Their goodness-of-fit analyses help assess whether the model captures the tendencies of actors in a social network to send and receive ties. The triad census counts the frequency of each of the sixteen possible network configurations among three actors present in a directed network. Goodness-of-fit analysis on the triad census helps assess the extent to which triadic effects (e.g., transitivity) and local group structure are accurately represented by the model. Other potential goodness of fit measures include network density (the proportion of potential ties that are actually present), reciprocity (the proportion of ties that are reciprocated) the distribution of geodesic distances (lengths of shortest paths between actors in a network) and measures of network autocorrelation (e.g., Moran, 1948; Geary, 1954). Generally, it is good practice to align network goodness-of-fit measures with the applied research question studied. For example, if a researcher is interested in the effect of performance homophily on friendship network formation, it makes sense to evaluate a model's goodness-of-fit on network autocorrelation on performance.

Multiplex goodness-of-fit In line with this rationale, we propose statistics to describe the fit of multiplex network models. Although we illustrate these only in the context of the multiplex p_2 model, they are applicable for multiplex network models generally. Naturally, for each layer of the observed multiplex network, we can evaluate goodness-of-fit on the statistics described above for uniplex networks. Moreover, for every pair of layers, we can measure the tendency for ties to occur jointly and to be reciprocated between layers. We can measure the tendency for ties to occur in two layers simultaneously by the Jaccard index of the two layers – the proportion of ties that appear in both networks out of the total number of unique ties in the two networks. If we let

$\boldsymbol{M}^t \in \{0,1\}^{n\times n}$ denote the adjacency matrix corresponding to layer t of multiplex network $\boldsymbol{M}$, and let $X(\boldsymbol{M}^t)$ denote the set of edges in layer t, then the Jaccard distance between layers s and t is given by

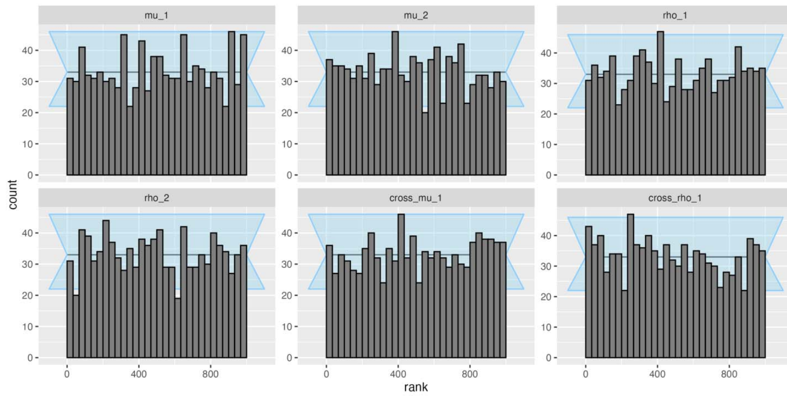
$$\text{Jaccard}(\boldsymbol{M}^t, \boldsymbol{M}^s) = \frac{|X(\boldsymbol{M}^t) \cap X(\boldsymbol{M}^s)|}{|X(\boldsymbol{M}^t) \cup X(\boldsymbol{M}^s)|}. \tag{11}$$

We can evaluate cross-layer reciprocity by $\text{Jaccard}(\boldsymbol{M}^t, (\boldsymbol{M}^s)^\top)$. The Jaccard index does not capture the joint absence of ties. In the context of social network analysis, tie presence is typically more meaningful than tie absence as social networks are often sparse. Therefore, counting similarity in non-ties would generally result in very high network similarity scores. Finally, to capture the similarities of actors' tendencies to send and receive ties in different network layers, we can use the correlations among actors' indegrees and outdegrees in all layers. Note that each of the statistics presented above aligns with an effect in the multiplex p_2 model.

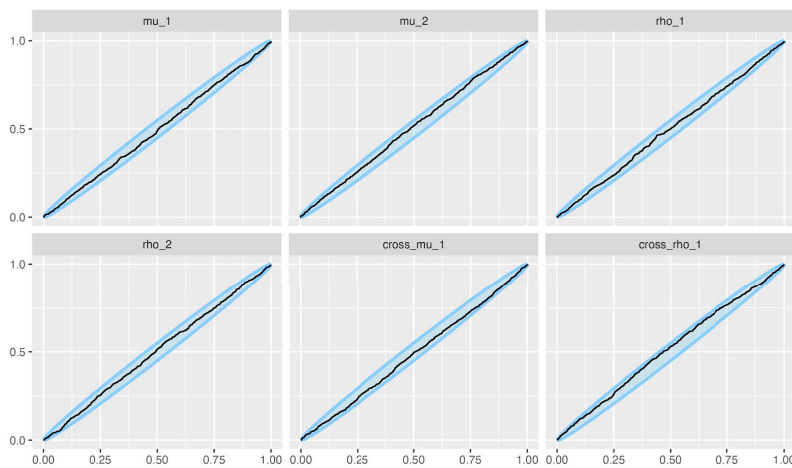
5 Simulation studies

In this section, we present the results of several simulation studies assessing the appropriateness of the selected priors for social networks, the accuracy of the estimation algorithm, and the inferential properties of the proposed method, following the workflow discussed by Gelman et al. (2020) and Schad et al. (2021). To this end, we simulate networks with a biplex structure ($T = 2$) consisting of 30 actors and thus $\binom{30}{2} = 435$ dyads. Networks of this size are typical in educational applications. The policy-related network in Section 6 is of this size as well. Our simulation model does not include covariates ($p = q = 0$) and its parameters are the within-layers densities μ^1 and μ^2, the within-layer reciprocities ρ^1 and ρ^2, the cross-layer density $\mu_{\text{cross}}^{(1,2)}$ and reciprocity $\rho_{\text{cross}}^{(1,2)}$, and the covariance $\boldsymbol{\Sigma_{AB}}$. The latter is not reported on below. The data are generated outside our estimation environment. The simulation studies were conducted on a server equipped with 64 GB of RAM and 32-core processors. We utilized the `doSNOW` (version 1.0.20) package and the internal functionality of `rstan` (version 2.32.6) to run the chains and models in parallel in R (R version 4.4.1).

Computational faithfulness We conduct simulation-based calibration (SBC) to assess the soundness of our posterior sampler (Talts et al., 2020). This procedure relies on the fact that the posterior distribution estimated from data generated from the prior should resemble the prior distribution, on average. We first draw $L = 1000$ random samples from the prior distribution, $\tilde{\boldsymbol{\theta}}_l \sim \pi(\boldsymbol{\theta})$, with $l = 1, \ldots, L$, and for each prior draw we generate a biplex network $\tilde{\boldsymbol{M}}_l$. We then fit the multiplex p_2 model to each simulated network $\tilde{\boldsymbol{M}}_l$, thus obtaining L estimated posterior distributions. If the model is well-calibrated, the sample from the prior could fall anywhere on the corresponding estimated posterior distribution. In other words, the rank of parameter $\tilde{\theta}_{lj}$ with respect to a given number of $K = 1000$ draws from the posterior $\{\theta^{'}_{lj,1}, \ldots, \theta^{'}_{lj,K}\}$ should be uniformly distributed. Figure 2 shows the histograms of the percentile statistics of the model parameters. The sampled statistics appear to be uniformly distributed, providing evidence that the model matches the generator and the algorithm is working as expected.



(a) The distribution of the observed rank statistics for the model parameters. The exact uniform distribution is given by the horizontal gray line and an approximate 95% interval of expected deviations from the uniform distribution is given in blue.



(b) The empirical cumulative distribution function (in black) where we expect to see diagonal lines with some deviations from exact uniformity. The blue ellipse outlines the 95% expected deviations.

Figure 2: Simulation-based calibration plots, created using the R package `SBC` (Modrák et al., 2023).

Model sensitivity We assess the model's adequacy for inference by measuring the bias of the posterior mean and the variance reduction from the prior to the posterior. We repeat the simulation steps from the computational faithfulness section and obtain the posterior estimates. For each parameter, the bias can be summarized by a posterior z-score, $z = (\mu_{\text{post}} - \tilde{\theta})/\sigma_{\text{post}}$, where the posterior mean μ_{post} is compared to the true parameter value $\tilde{\theta}$, scaled by the posterior uncertainty. We estimate the reduction in uncertainty by the posterior contraction, which is the variance reduction compared to

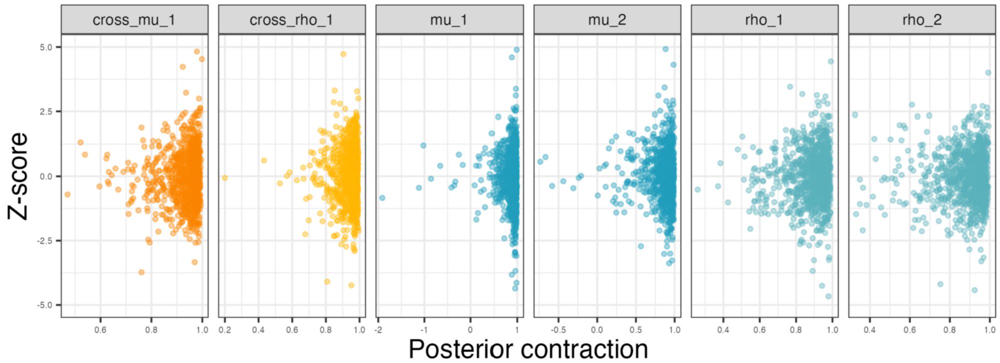


Figure 3: The model sensitivity plots of z-scores against posterior contractions calculated on 1000 simulations (26 outliers not shown). If no bias is present in the simulations, then the z-score distribution of should be centered on 0. The posterior contraction estimates how much prior uncertainty is reduced in the posterior estimation. Thus high posterior contraction and posterior z-scores close to zero reflect an ideal situation and good model fit. Here, we observe moderately low bias and low variance.

the prior variance, $s = (\sigma^2_{\text{prior}} - \sigma^2_{\text{post}})/\sigma^2_{\text{prior}}$. Schad et al. (2021) recommend plotting the posterior z-score against the posterior contraction for each parameter. Ideally, a model with low bias and low variance would have all the points concentrated around a posterior z-score of 0 and a posterior contraction of 1. Figure 3 shows that the posterior z-scores are centered around 0 with concentration at the level of posterior contraction close to 1.

We note that the μ_1 and μ_2 we report are the post-swept density parameters (see Section 3), whose prior variance incorporates the hyperprior for the random effects. Since we approximated the prior variance, the resulting posterior contraction is not exact and may be negative. In a few cases, extreme draws from the prior distribution led to uninformative networks from which the model could not recover the parameters. We have removed 26 of the 6000 points, showing extreme outliers, from Figure 3.

Computation assessment We assessed the computation time of the multiplex p_2 model on biplex networks of varying sizes: 20, 30, 45, 70, 150, and 200. Each model was estimated with 1000 warmup iterations and 1000 sampling iterations, without covariates. The setup mirrors the beginning of Section 5, with the exception of the increasing network sizes. For each network size, Figure 4 shows the total time for three chains: computation considerably increases with network size. Reducing computation time for larger networks is an important direction for future work.

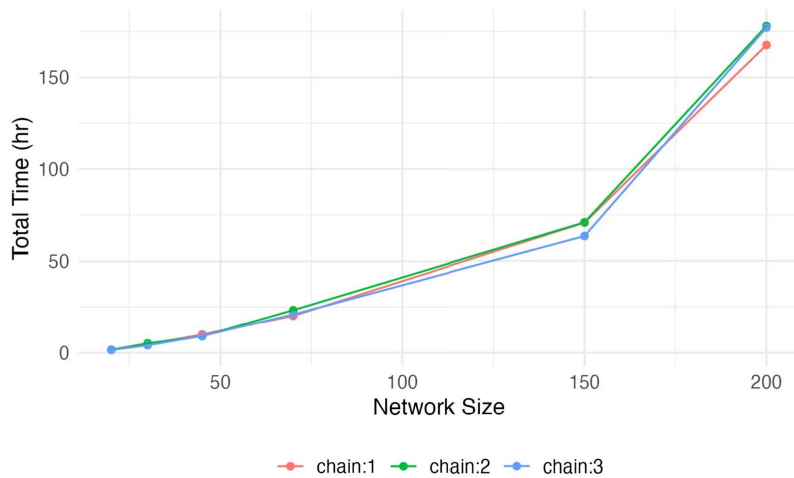


Figure 4: Computation time in hours for a biplex p_2 model without covariates for various network sizes (20, 30, 45, 70, 150, and 200). Three chains were run in parallel with 2000 iterations (warmup and sampling) for each chain and network size.

6 Application: Information exchange in a policy network

In this illustrative study, we analyze the patterns of information exchange and the perception of influence among 30 actors in the policy domain of toxic chemicals regulation in Germany in the 1980s. This (multiplex) policy network was studied previously by Leifeld and Schneider (2012), to determine how governmental and nongovernmental actors chose their potential interaction partners in this historic context, though not from a multiplex network perspective.

Leifeld and Schneider (2012) studied two types of information exchange networks and the network of organizations' perception of influence. In the *political information* network, an edge from organization i to organization j indicates that i perceived j as a partner in exchanging political information regarding chemical controls. This network was obtained by asking organizations to list the names of all organizations with whom they regularly exchanged information about affairs related to chemicals control. An edge in the *scientific information* network indicates that organization i provided scientific and technical information about toxic chemicals to organization j. In the *perception of influence* network, an edge from organization i to organization j indicates that i perceived j to be particularly influential in the policy making process. Figure 5 depicts the three networks. We refer to Schneider (1988) and Leifeld and Schneider (2012) for further details about the data collection process.

Given that power dynamics likely impact information exchange and perception of influence, we take into account the effect of governmental status (indicating whether an organization is governmental) in our analyses. An organization is also likely to perceive another organization as more influential and engage in greater information exchange if they hold similar policy positions. Therefore, we account for policy preference similarity

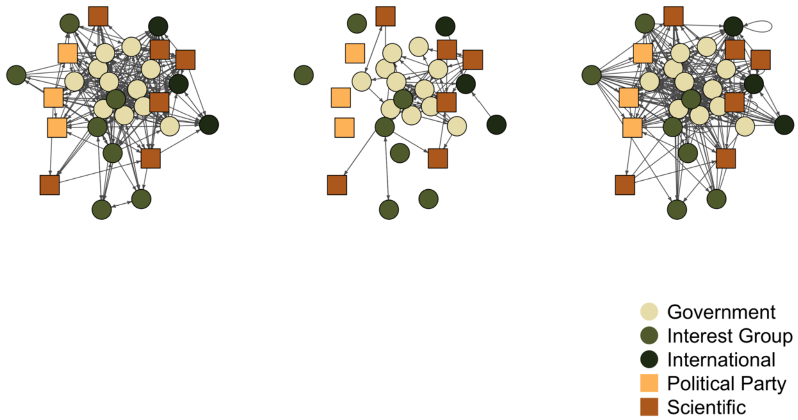


Figure 5: The information exchange and perception of influence networks of organizations in the policy domain of toxic chemicals regulation in Germany in the 1980s. Node shape and color combinations indicate institution type.

in this study, which is a pairwise similarity score reflecting how closely two actors align on six key policy issues (self-regulation, scope of the reform, control procedure, timing of the control mechanism, intensity of control, and treatment of chemicals on the market); see Leifeld and Schneider (2012) for the similarity score formula.

We address two of the main hypotheses by Leifeld and Schneider (2012) in this study. Firstly, we assess whether information exchange is more likely among organizations with pre-existing communication ties. We expect ties will co-occur in the two information networks, and to be reciprocated within and across these two network layers (Hypothesis 1). Secondly, we investigate whether institutions are more likely to exchange information with institutions they perceive as influential (Hypothesis 2), as the potential benefits gained from high-influence actors may justify the costs associated with establishing information ties.

Leifeld and Schneider (2012) tested these hypotheses by fitting separate (uniplex) exponential random graph models to the political and scientific information networks, including the other two networks as covariates to examine cross-layer effects. By contrast, here we fit a multiplex p_2 model that simultaneously models the three networks as outcomes. As both directions of cross-layer effects (that is, density *and* reciprocity) are incorporated in our model, this approach naturally extends the idea of using existing communication channels to cross-layer reciprocity, going beyond what was analyzed in the original study. Moreover, our random actor-level effects may yield insights into the association of actors' behaviors within different relational contexts.

Results Table 1 shows that the political information network and the perception of influence network are more dense than the scientific information network, while the proportion of ties that are reciprocated is highest in the political information network. Notably, there is a large overlap between the political information network and the

Network	Density	Reciprocity
Political information (PO)	0.393	0.614
Scientific information (SC)	0.072	0.349
Perception of influence (PE)	0.325	0.382

Table 1: Density and reciprocity of the three layers in the policy network.

distance to:	SC		$SC^\top$		PE		$PE^\top$	
	Obs	$\mathbb{E}(J)$	Obs	$\mathbb{E}(J)$	Obs	$\mathbb{E}(J)$	Obs	$\mathbb{E}(J)$
PO	0.174	0.061	0.157	0.061	0.398	0.182	0.308	0.182
SC					0.123	0.060	0.078	0.060

Table 2: The observed (Obs) Jaccard distance between networks X and Y reflects cross-layer network overlap in the data and is determined as in equation (11). The Jaccard distance between X and $Y^\top$ reflects cross-layer reciprocity. For reference, the $\mathbb{E}(J)$ columns show the expected Jaccard index under random edge permutation.

perception of influence network (see Table 2), meaning that organizations send political information to the organizations they deem influential. Approximately 40% of the unique ties between these networks overlap, as indicated by a Jaccard index of 0.398. Moreover, organizations tend to send information to those organizations who perceive them to be influential (Jaccard index: 0.308). However, we do not observe such strong association between the scientific information network and the perception of influence network. We note that the observed Jaccard indices are a lot higher than the expected Jaccard indices if we had permuted the edges randomly.[2] This is true for most network combinations, except for the scientific information and reversed perception of influence ties: organizations are not very likely to receive scientific information from organizations they perceive as influential.

We estimated two multiplex p_2 models on the policy network. Model 1 is a baseline model, while Model 2 includes the effect of governmental status on actors' tendencies to send and receive ties and that of policy preference similarity on within-network density. For both models, we ran four Hamiltonian MCMC chains with 4000 total iterations (2000 warm-ups) each and the priors specified in Section 3. The potential scale reduction statistics $\hat{R}$ are smaller than 1.05 for all the parameters in both models, suggesting model convergence (Gelman and Rubin, 1992). Further convergence diagnostics for Model 2 can be found in Figures 3–6 in the Supplementary Material (Hong and Niezink, 2025).

[2]The expected Jaccard index of two networks among n actors under random edge permutation, assuming m_s edges in the first network and m_t edges in the second, is given by

$$\sum_{k=\max(0,m_s+m_t-N)}^{\min(m_s,m_t)} \frac{k}{m_s+m_t-k} \frac{\binom{N}{m_s}\binom{m_s}{k}\binom{N-m_s}{m_t-k}}{\binom{N}{m_s}\binom{N}{m_t}}, \tag{12}$$

with $N = n \cdot (n-1)$ the total number of possible edges in a graph with n actors, k the number of common edges, and $\frac{k}{m_s+m_t-k}$ the corresponding Jaccard index. The numerator of the second term counts the unique ways of picking two sets of size m_s and m_t with k overlaps.

	Model 1		Model 2	
	mean	95% CI	mean	95% CI
Within-layer dyadic effects				
Density μ^{PO}	**-2.15**	$(-2.63, -1.68)$	**-3.84**	$(-5.15, -2.56)$
Preference similarity δ_{μ}^{PO}			**0.36**	(0.16, 0.57)
Density μ^{SC}	**-8.84**	$(-10.80, -7.19)$	**-9.32**	$(-12.31, -6.74)$
Preference similarity δ_{μ}^{SC}			-0.10	$(-0.44, 0.23)$
Density μ^{PE}	**-2.02**	$(-2.50, -1.56)$	**-3.19**	$(-4.77, -1.51)$
Preference similarity δ_{μ}^{PE}			0.07	$(-0.16, 0.30)$
Reciprocity ρ^{PO}	**1.95**	(1.11, 2.82)	**1.93**	(1.06, 2.81)
Reciprocity ρ^{SC}	**1.71**	(0.32, 3.21)	**1.70**	(0.32, 3.19)
Reciprocity ρ^{PE}	0.14	$(-0.67, 0.96)$	0.17	$(-0.62, 0.96)$
Cross-layer dyadic effects				
Cross-density $\mu^{\text{PO,SC}}$	**2.46**	(1.13, 3.99)	**2.50**	(1.16, 4.11)
Cross-density $\mu^{\text{PO,PE}}$	**1.13**	(0.56, 1.70)	**1.09**	(0.55, 1.66)
Cross-density $\mu^{\text{SC,PE}}$	0.31	$(-0.56, 1.21)$	0.33	$(-0.56, 1.20)$
Cross-reciprocity $\rho^{\text{PO,SC}}$	**2.50**	(1.43, 3.61)	**2.53**	(1.49, 3.65)
Cross-reciprocity $\rho^{\text{PO,PE}}$	-0.06	$(-0.65, 0.53)$	-0.05	$(-0.64, 0.53)$
Cross-reciprocity $\rho^{\text{SC,PE}}$	-0.17	$(-1.02, 0.70)$	-0.16	$(-1.04, 0.70)$
Sender effects				
Gov. status $\gamma_{\alpha}^{\text{PO}}$			0.26	$(-2.60, 3.22)$
Gov. status $\gamma_{\alpha}^{\text{SC}}$			-0.67	$(-4.16, 3.01)$
Gov. status $\gamma_{\alpha}^{\text{PE}}$			0.59	$(-1.87, 3.12)$
Receiver effects				
Gov. status $\gamma_{\beta}^{\text{PO}}$			1.93	$(-0.19, 4.06)$
Gov. status $\gamma_{\beta}^{\text{SC}}$			2.02	$(-1.05, 5.25)$
Gov. status $\gamma_{\beta}^{\text{PE}}$			2.07	$(-0.99, 5.18)$

Table 3: Posterior means and 95% credible intervals (CIs) for the multiplex p_2 parameters in the German toxic chemical policy network (PO = political information, SC = scientific information, PE = perception of influence). Posterior means with CIs above or below 0 are bolded. Cross-density and cross-reciprocity are short for cross-layer density and cross-layer reciprocity. Model 2 accounts for the effects of preference similarity on network density (δ_μ) and the actor effects of being a governmental organization ($\gamma_\alpha, \gamma_\beta$).

Table 3 summarizes the results of the two models. In line with the descriptive statistics, we find a very negative density effect for the scientific information network. Notably, information ties are likely to be reciprocated within and across network layers (log odds: 1.95 for within-layer political, 1.71 for within-layer scientific, 2.50 for cross-layer reciprocity). These findings support Hypothesis 1, suggesting that if actor i delivers any type of information to actor j, actor j is more likely to reciprocate with at least one type of information to actor i. Furthermore, we observe a positive cross-network density effect between the perception of influence and the political network (log odds: 1.13). This aligns with Hypothesis 2, indicating that actors are more likely to deliver political information to institutions they perceive as influential. In Model 2, we observe a slightly positive effect of policy preference similarity on political network density (log odds: 0.36), suggesting that organizations are more likely to send political information to others with whom they tend to agree on the core policy topics relevant to the chemical regulation process.

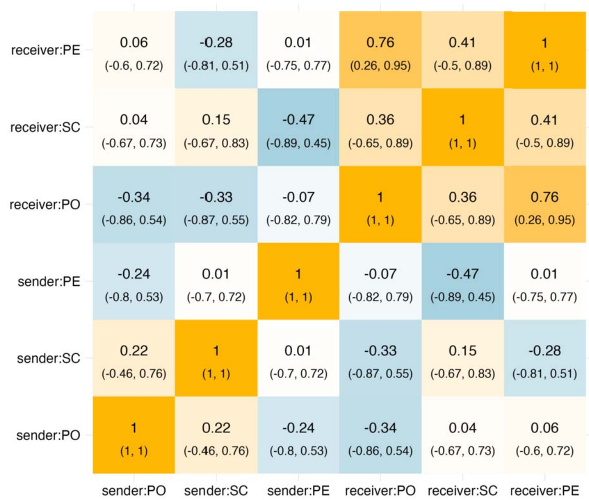


Figure 6: The correlation matrix for the actor random effects in the chemical policy networks (PO = political information, SC = scientific information, PE = perception of influence).

Figure 6 displays the estimated correlation matrix, including the 95% credible intervals. The tendency to receive political information ties is positively correlated with the tendency to be perceived as influential (posterior correlation: 0.76). While results are not conclusive, the positive correlations observed in the top-right (receiver) block suggest that the tendency to receive ties is positively correlated across network layers.

Goodness-of-fit We apply the goodness-of-fit methods described in Section 4 to Model 2. Figure 7 displays the results of the proposed multiplex goodness-of-fit methods which demonstrate an adequate fit to the multiplex network. Additional goodness-of-fit checks on the covariance among the in- and outdegrees of actors in the three networks can be found in the Figure 7 in the Supplementary Material. Figure 8 shows the results of the existing uniplex goodness-of-fit methods with implementation adapted from the `RSiena` package (Snijders et al., 2024). Overall, we observe a good fit for the indegree and outdegree distributions, as well as for the triad census, in each layer of the network individually. The adequate fit of the triad census suggests that the p_2 model is appropriate for these data, despite not explicitly modeling triadic effects. (Hong and Niezink, 2025).

7 Application: Discrepancies in perceptions of gossip

Gossip is a universal phenomenon in human groups. It can be defined as informal communication about a third, non-present person (Dores Cruz et al., 2021) and has been linked to both positive outcomes, such as promoting cooperation, and negative out-

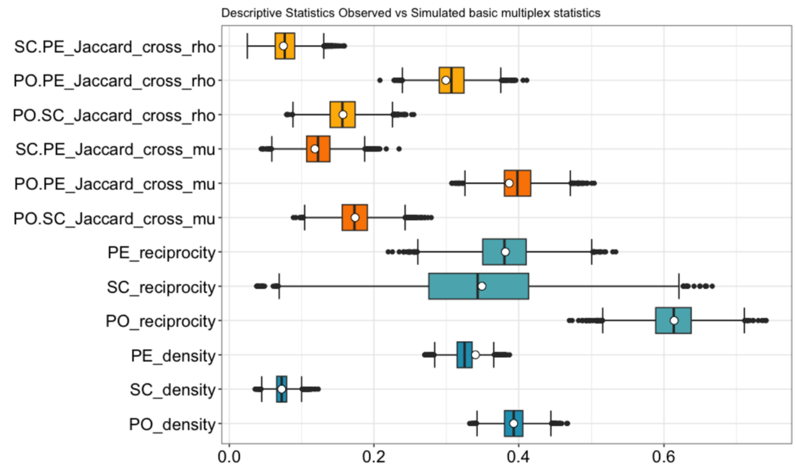


Figure 7: Multiplex goodness-of-fit measures. The white dots indicate the statistics calculated on the observed network. The box-plots are based on statistics calculated on networks simulated from 1000 posterior draws.

comes, such as the social exclusion of the target (e.g., Feinberg et al., 2014; Kisfalusi et al., 2019). In this section, we focus on gossip in the school setting. Here, gossip has been considered a form of bullying (Kisfalusi, 2018), which significantly affects children's social development and school outcomes.

Despite the important role of gossip in social interactions in schools, there are nuances to the phenomenon which are often overlooked. In particular, studies on gossip usually only consider the viewpoint of either the gossiper or the gossip target. Yet, different individuals may perceive the same event differently, leading to divergent perspectives. In the case of bullying, for example, individuals who are identified as bullies by their victims may not self-identify as such. The non-confrontational nature of gossip introduces an additional layer of complexity when gossip targets need to identify who they think is talking about them behind their back. Previous research has examined the disagreements between bullies and victims in reported bullying behavior (Veenstra et al., 2007; Tolsma et al., 2013; Kisfalusi, 2018; Tatum and Grund, 2020). In this setting, the two perspectives were highly complementary. Yet, the disagreements between gossipers and gossip targets so far remain unstudied.

In this section, we take a multiplex network perspective on the gossip relation, simultaneously studying self-reported gossip behavior and the perceptions of gossip behavior by the gossip target. Being a gossiper might affect how an individual perceives themselves as the victim of gossip, and gossip victims may feel the need to retaliate. A multiplex network perspective is necessary to understand such associations. In particular, we will focus on the following two research questions. First, are students more likely to gossip about people who they think are gossiping about them? We expect this is the case and will refer to this as *retaliation*. Second, are students likely to accurately perceive their gossipers? As discrepancies between reports by aggressors and victims

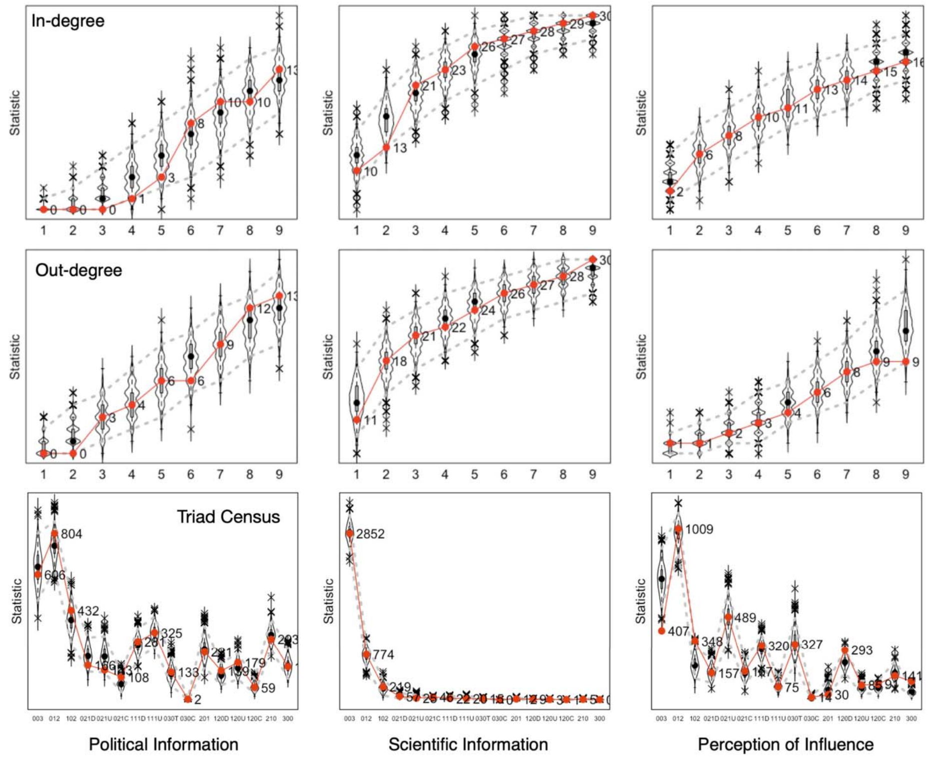
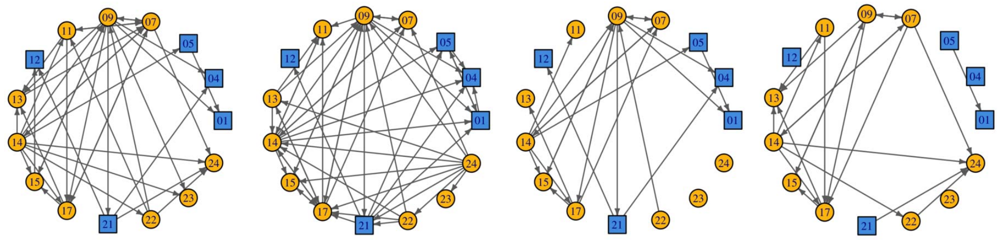


Figure 8: Simulated goodness-of-fit statistics for the policy triplex network: political information (first column), scientific information (middle), and perception of influence (last column). From top to bottom, the statistics are the cumulative indegree distribution, the cumulative outdegree distribution, and the triad census (See Figure 2 in Leinhardt (1971) for triad configurations). The figure compares the observed network's statistics (red line) to the distribution of the same statistics on networks generated from the fitted model (black dots: simulated average). The plotting functions are adopted from the R package `RSiena` (Snijders et al., 2024).

have been found in the case of bullying (Veenstra et al., 2007), we expect these to exist as well in reports of gossip. We will refer to this as *perception discrepancy* (as opposed to *perception accuracy*). In the case of retaliation, individual i both gossips about j *and* thinks that j is gossiping about them. Accurately perceived gossip ties are those where individual i gossips about j, and j thinks that i gossips about them. Figure 9 illustrates these relations.

We expect that gender moderates the above effects, such that the likelihood of retaliation and accurate perception is higher when the sender and the receiver are of the same gender. To investigate this hypothesis, we incorporate the binary indicator of whether the two actors in a dyad share the same gender as a covariate when examining cross-network density and reciprocity. Additionally, we study how gender influences



(a) **Gossip network**: $i \to j$ means i admits to gossip about j.

(b) **Perception network**: $i \to j$ means i thinks j gossips about them.

(c) **Retaliation**: $i \to j$ means i thinks j gossips about them *and* i admits to gossip about j.

(d) **Accurately perceived gossip**: $i \to j$ means i admits to gossip about j *and* j accurately perceives i as gossiping about them.

Figure 9: An example of a biplex network of self-reported gossip ties and perceived gossip ties among 15 students. There are 10 female (yellow circle) and 5 male students (blue square).

individuals' propensities to both send and receive ties by incorporating gender as an actor covariate in our analysis.

Data We will study the gossiper-target relationship based on data from the fourth wave of a six-wave panel study on Hungarian elementary school students, conducted between 2013 and 2017. The fourth wave of data collection occurred in the spring of 2015, when the students were enrolled in the sixth grade and were on average 13 years old. We dropped 9 classes with zero perception ties and analyze the remaining 34 classes containing a total of 702 students. See Kisfalusi (2018) and Kisfalusi et al. (2021) for more detailed information on the data collection. We simultaneously study self-reported gossip (students' answers to the question: 'About whom do you talk with your classmates behind his/her back?') and perceived gossip ('Who do you think is talking about you with other classmates behind your back?'). These two relations together yield a biplex network. Figure 9 illustrates the gossip data collected in one of the classrooms. An example of a retaliation-only dyad is $\{9, 21\}$, where student 9 believes that student 21 is gossiping about him, and retaliates by gossiping about 21. However, 21 neither gossips about 9 nor accurately perceives that 9 is gossiping about him.

Descriptive statistics of the 34 biplex networks are summarized in Table 4. On average, we have about 20 students present in each classroom with a balanced proportion of female to male students. On average, the gossip networks have fewer ties than the perception networks, and reciprocity is also lower in the gossip networks. Based on the Jaccard indices, there are considerably more pairs of students i and j where i gossips about j and i thinks j is gossiping about them than pairs where i gossips about j and j accurately perceives that i is gossiping about them. This descriptive finding concurs with our expectation on students' retaliating behavior.

	Mean	Median	SD
Number of students	20.7	19.5	4.96
Proportion female	0.48	0.47	0.10
Gossip density	0.09	0.08	0.06
Gossip reciprocity	0.13	0.14	0.11
Perception density	0.15	0.14	0.06
Perception reciprocity	0.23	0.23	0.11
Gossip $\times$ perception Jaccard	0.24	0.21	0.12
Gossip $\times$ perception$^\top$ Jaccard	0.11	0.10	0.07
Expected gossip $\times$ perception Jaccard	0.05	0.05	0.02

Table 4: Summary statistics of the gossip and perceived gossip networks in 34 Hungarian elementary school classrooms. The expected Jaccard distance of the gossip and perception network under random edge permutation equals that of gossip and perception$^\top$.

Plan of Analysis We fit two models to each classroom: a baseline biplex p_2 model and a model that additionally includes the effect of gender (female = 1) on students' tendencies to send and receive ties, and the (dyadic) effect of having the same gender on cross-layer density and reciprocity. For each model, we ran 4 chains with 4000 total iterations (2000 warm-ups). Posterior estimates are calculated based on 8000 posterior draws. The potential scale reduction statistic $\hat{R}$ was less than 1.05 for all the parameters in each of the 68 models, suggesting the convergence of the models.

To aggregate the results from all classrooms, we conduct a Bayesian meta-analysis by fitting a normal-normal hierarchical model to the fixed parameters of each classroom (Gelman et al., 2013, Ch. 5.5). Let $\hat{\boldsymbol{\theta}} \in \mathbb{R}^{34}$ denote the vector of the posterior means for one of the fixed model parameters. We assume that

$$\begin{aligned} \hat{\theta}_j &\sim \mathcal{N}(\theta_j, \sigma_j^2), \\ \theta_j &\sim \mathcal{N}(\mu, \tau^2), \end{aligned}$$

where μ and τ are the overall population mean and standard deviation. For the priors of the hyperparameters, we assume $\mu \sim \mathcal{N}(0, 10)$ and $\tau \sim \text{Cauchy}(0, 0.5)$ for all the fixed effects. Since correlations only take values in $[-1, 1]$, we compute the Fisher z-transformation (Fisher, 1921) of the estimated correlations and apply the normal-normal model to the transformed quantities. We used the R package `brms` (Bürkner, 2021) with 5000 iterations for the meta-analysis (Harrer et al., 2021, Ch. 13).

Results Table 5 shows the posterior means and the 95% credible intervals of the population means of the p_2 model parameters. In Model 1, we find a positive cross-density effect (estimated population mean $\hat{\mu} = 2.62$), which supports our expectation about gossip retaliation. Students are about 13 times more likely to gossip about someone if they think the person is gossiping about them (or vice versa). We do not observe a positive cross-reciprocity effect ($\hat{\mu} = -0.13$) indicating no evidence that students accurately perceive who is gossiping about them. This result is consistent with our expectation on gossip perception and the studies on the dual perspective on bullying.

	Model 1		Model 2	
	mean	95% CI	mean	95% CI
Within-layer dyadic effects				
Density μ^{G}	**−11.45**	(−13.37, −9.49)	**−12.91**	(−15.72, −10.18)
Density μ^{P}	**−6.46**	(−7.30, −5.62)	**−7.72**	(−9.00, −6.30)
Reciprocity ρ^{G}	−0.75	(−2.01, 0.57)	−0.90	(−2.44, 0.53)
Reciprocity ρ^{P}	**1.35**	(0.53, 2.07)	1.49	(0.65, 2.23)
Cross-layer dyadic effects				
Cross-density $\mu^{\text{G,P}}$	**2.62**	(1.75, 3.57)	**1.36**	(0.33, 2.42)
Same-gender $\delta^{\text{G,P}}_{\mu,\text{cross}}$			**2.05**	(0.85, 3.09)
Cross-reciprocity $\rho^{\text{G,P}}$	−0.13	(−0.98, 0.79)	−0.94	(−2.20, 0.23)
Same-gender $\delta^{\text{G,P}}_{\rho,\text{cross}}$			−0.33	(−1.89, 1.67)
Sender effects				
Female $\gamma^{\text{G}}_{\alpha}$			−0.14	(−1.99, 1.64)
Female $\gamma^{\text{P}}_{\alpha}$			0.71	(−0.50, 1.76)
Receiver effects				
Female $\gamma^{\text{G}}_{\beta}$			−0.54	(−1.53, 0.52)
Female $\gamma^{\text{P}}_{\beta}$			0.94	(−0.28, 2.05)
Correlations				
Sender$^{\text{G}}\times$ receiver$^{\text{G}}$ $\rho_{A^{\text{G}},B^{\text{G}}}$	−0.04	(−0.12, 0.05)	−0.03	(−0.11, 0.07)
Sender$^{\text{G}}\times$ sender$^{\text{P}}$ $\rho_{A^{\text{G}},A^{\text{P}}}$	**0.24**	(0.14, 0.33)	**0.22**	(0.14, 0.31)
Sender$^{\text{G}}\times$ receiver$^{\text{P}}$ $\rho_{A^{\text{G}},B^{\text{P}}}$	0.06	(−0.05, 0.17)	0.00	(−0.11, 0.10)
Sender$^{\text{P}}\times$ receiver$^{\text{G}}$ $\rho_{A^{\text{P}},B^{\text{G}}}$	0.09	(−0.05, 0.23)	0.11	(−0.02, 0.23)
Receiver$^{\text{G}}\times$ receiver$^{\text{P}}$ $\rho_{B^{\text{G}},B^{\text{P}}}$	**0.16**	(0.03, 0.29)	**0.14**	(0.03, 0.25)
Sender$^{\text{P}}\times$ receiver$^{\text{P}}$ $\rho_{A^{\text{G}},B^{\text{P}}}$	−0.11	(−0.27, 0.05)	−0.15	(−0.27, 0.00)

Table 5: Posterior means and 95% credible intervals (CIs) of the overall means of the multiplex gossip networks p_2 parameters from 34 Hungarian elementary school classrooms (G = gossip, P = perception of gossip). Posterior means with CIs above or below 0 are bolded. Cross-density and cross-reciprocity are short for cross-layer density and cross-layer reciprocity. Model 2 accounts for the effects of gender on cross-layer density ($\delta^{\text{G,P}}_{\mu,\text{cross}}$) and reciprocity ($\delta^{\text{G,P}}_{\rho,\text{cross}}$) and the actor effects of being female ($\gamma_\alpha, \gamma_\beta$).

Gossip networks are generally sparse, and we indeed find negative network-specific density effects in Model 1 for both self-reported gossip and perceived gossip. Negative ties are often reciprocated but, interestingly, we do not observe a reciprocity effect for self-reported gossip ($\hat{\mu} = -0.75$, CI $= [-2.01, 0.57]$). Yet, when it comes to perceived gossip, we do observe a positive reciprocity effect. One reason for this may be that students feel more comfortable reporting on perceived negative behavior than on their own negative behavior, because of social desirability bias. Also, if two students dislike each other, they may each suspect the other to be gossiping about them. As such, a reciprocated perceived gossip relation could indicate a more generally negative tie.

After accounting for effects of gender in Model 2, the above-mentioned findings remain the same. We find a positive same-gender cross-density effect ($\hat{\mu} = 2.05$), indicating that i is about 7.5 times more likely to gossip about j *and* to think j is gossiping about them if i and j are of the same gender. We do not find evidence that perception accuracy changes when gossiper and target are of the same gender. Also, controlling for the other effects, we do not find a differential tendency for female students to send or receive self-reported or perceived gossip ties.

Finally, we consider the estimated correlation of the actor random effects, for which the two models yield comparable results, though Model 2 shows slightly lower estimated correlations after accounting for actor effects of gender. In particular, the gossip network sender random effect is positively correlated with the perception network sender random effect and the same holds true for the receiver random effects in both networks. This means that individuals who report to gossip a lot about others also tend to think others are gossiping about them – a tendency that goes beyond the urge to retaliate among specific student pairs. And individuals who are gossiped about a lot are also suspected to be avid gossipers. Note that the first finding is a negative form of generalized reciprocity, where individuals treat others in the same way that others treated them in the past. Interestingly, we find a negative covariance between the perception network random sender and receiver effects. That is, the more students think they are gossiped about, the less likely they are accused of gossip by others. This finding contrasts the positive perception reciprocity effect we found at the dyad level.

8 Discussion

As social actors are often embedded in multiple interconnected social networks, we propose a Bayesian multiplex network model in the p_2 modeling framework. This model captures the interplay of social dynamics across different network layers by introducing cross-layer dyadic effects and actor random effects. Assuming the dyads are conditionally independent, we formulate the model as a mixed-effects multinomial logistic regression capable of handling a range of network dependencies while remaining interpretable. The proposed methodology is available in the R package `multip2`.[3] The code to replicate the simulation studies and data analyses can be found in a separate repository.[4]

Despite being widely used and simplifying model specification and estimation, the conditionally independent dyads assumption precludes modeling some more intricate network dependencies. In real networks, higher-order network effects such as transitivity (e.g., befriending a friend of a friend) can be strong drivers of tie formation, beyond the effect of reciprocity, individual and dyadic covariates, and variation in actors' tendencies to send and receive ties. Our definition of the multiplex p_2 model does not explicitly represent triadic effects. Surprisingly, however, the model adequately captures the triadic structures present in the policy network, as is illustrated in Figure 8. Regarding dyadic goodness of fit, as shown in Figure 8 of the Supplementary Material, the multiplex p_2 model captures multiplex goodness-of-fit statistics, such the overlap between different networks, much better than a uniplex approach, where a p_2 model is fitted on each of the network layers separately.

The ability of lower-order network models to capture higher-order effects (i.e., triad effects) has been documented and explored by Faust (2010). In the context of the social

[3] See GitHub: https://github.com/annihong/multip2
[4] See GitHub: https://github.com/annihong/multiplex_p2_replication_code

relations model, Minhas et al. (2019) accounted for triadic effects while preserving the conditionally independent dyads assumption by incorporating a multiplicative component in the form of a latent factor model in their model. The multiplex p_2 model could be extended similarly. Moreover, while we have proposed several multiplex goodness-of-fit measures, it still remains unexplored how to best incorporate goodness-of-fit statistics that capture multiplex triadic patterns. As the number of network layers increases, the number of possible triad configurations on a dyad grows exponentially. Therefore, the choice of which multiplex network triad configurations to include should be guided by the specific applied research context.

Additionally, it is important to note that the priors employed in our study remain invariant to increasing network dimensions and network size across key network statistics, such as density, reciprocity, and transitivity. We have included additional prior predictive checks in Figure 2 of the Supplementary Material (Hong and Niezink, 2025). However, when applying the multiplex p_2 model, practitioners should conduct separate prior predictive checks to ensure that the selected priors induce appropriate network properties.

In the gossip study (Section 7), we considered data from 34 school classes and aggregated the multiplex p_2 model results by a Bayesian meta-analysis. Nowadays, network data collection frequently involves sampling from a population of networks (e.g., school classes, households, organizational work units; Goeyvaerts et al., 2018; Lubbers, 2003). In such samples, bigger networks typically have lower density, while the mean degree is similar across the networks (Krivitsky et al., 2011) – something we did not account for in the current study. A multilevel extension of the multiplex p_2 model would enable researchers to study a sample of multiplex networks across various groups, and thus obtain more generalizable results and address group-level research questions. Moreover, the multilevel approach would estimate parameters more efficiently by partially pooling observations across the groups to estimate the global parameters. The multilevel model could also explicitly account for the effect of network size variation in a sample of networks on parameter estimates, in case that variation was large (Krivitsky et al., 2023; Niezink, 2023).

The `R` package `multip2`, associated to this paper, allows for part of the network data to be missing. This feature can be leveraged to assess, e.g., the impact of removing dyadic outliers by setting all layers of a certain dyad to missing and refitting the model for comparison. Yet, the current implementation of the package does not support missing data in specific layers – if data on a dyad in one layer are missing, all layers for that dyad are excluded. We plan to improve this functionality by retaining dyadic information in non-missing layers. We also plan to improve the computational efficiency of the estimation algorithm by utilizing Graphics processing units (GPUs) through the `OpenCl` framework and the Stan OpenCL backend implementation in `CmdStan` (Stan Development Team, 2023b). Enhancing the estimation speed will increase the applicability of the multiplex p_2 model and will be essential for its multilevel extension.

Acknowledgments

The authors would like to thank Károly Takács for sharing the data analyzed in Section 7. The authors also thank the two anonymous reviewers whose constructive comments led to a substantial improvement of the paper.

Funding

This material is based upon work supported by the National Science Foundation under Award Number SES-2020276.

Supplementary Material

Supplement to "The Multiplex p_2 Model: Mixed-Effects Modeling for Multiplex Social Networks" (DOI: 10.1214/25-BA1527SUPP; .pdf). The Supplementary Material contains additional prior predictive checks, convergence checks, and goodness-of-fit results.

References

Akinc, D. and Vandebroek, M. (2018). "Bayesian Estimation of Mixed Logit Models: Selecting an Appropriate Prior for the Covariance Matrix." *Journal of Choice Modelling*, 29: 133–151. 7

Amini, A., Paez, M., and Lin, L. (2024). "Hierarchical Stochastic Block Model for Community Detection in Multiplex Networks." *Bayesian Analysis*, 19(1): 319–345. MR4692550. doi: `https://doi.org/10.1214/22-ba1355`. 3

Barbillon, P., Donnet, S., Lazega, E., and Bar-Hen, A. (2017). "Stochastic Block Models for Multiplex Networks: An Application to a Multilevel Network of Researchers." *Journal of the Royal Statistical Society - Series A*, 295–314. MR3600512. doi: `https://doi.org/10.1111/rssa.12193`. 3

Barnard, J., McCulloch, R., and Meng, X.-L. (2000). "Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage." *Statistica Sinica*, 10(4): 1281–1311. MR1804544. 7

Bellio, R. and Soriani, N. (2021). "Maximum Likelihood Estimation Based on the Laplace Approximation for p2 Network Regression Models." *Statistica Neerlandica*, 75(1): 24–41. MR4195578. doi: `https://doi.org/10.1111/stan.12223`. 3, 6

Betancourt, M. and Girolami, M. (2015). "Hamiltonian Monte Carlo for Hierarchical Models." *Current Trends in Bayesian Methodology with Applications*, 79(30): 2–4. MR3644666. 6

Betancourt, M. and Stein, L. C. (2011). "The Geometry of Hamiltonian Monte Carlo." *arXiv preprint arXiv:1112.4118*. 6

Borgatti, S. P., Mehra, A., Brass, D. J., and Labianca, G. (2009). "Network Analysis in the Social Sciences." *Science*, 323(5916): 892–895. 1

Bürkner, P.-C. (2021). "Bayesian Item Response Modeling in R with brms and Stan." *Journal of Statistical Software*, 100(5): 1–54. MR2657265. doi: https://doi.org/10.1007/978-1-4419-0742-4. 21

Caimo, A. and Friel, N. (2011). "Bayesian Inference for Exponential Random Graph Models." *Social Networks*, 33(1): 41–55. MR2873466. 9

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). "Stan: A Probabilistic Programming Language." *Journal of Statistical Software*, 76(1): 1–32. 6

Chen, T. H. Y. (2021). "Statistical Inference for Multilayer Networks in Political Science." *Political Science Research and Methods*, 9(2): 380–397. Publisher: Cambridge University Press. 4

Dabbs, B., Adhikari, S., and Sweet, T. (2020). "Conditionally Independent Dyads (CID) network models: A latent variable approach to statistical social network analysis." *Social Networks*, 63: 122–133. 2

Dekker, D., Krackhardt, D., and Snijders, T. A. (2007). "Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions." *Psychometrika*, 72: 563–581. MR2377554. doi: https://doi.org/10.1007/s11336-007-9016-1. 3

Dores Cruz, T. D., Nieper, A. S., Testori, M., Martinescu, E., and Beersma, B. (2021). "An Integrative Definition and Framework to Study Gossip." *Group & Organization Management*, 46(2): 252–285. 17

Duane, S., Kennedy, A., Pendleton, B. J., and Roweth, D. (1987). "Hybrid Monte Carlo." *Physics Letters B*, 195(2): 216–222. MR3960671. doi: https://doi.org/10.1016/0370-2693(87)91197-x. 6

Faust, K. (2010). "A Puzzle Concerning Triads in Social Networks: Graph Constraints and the Triad Census." *Social Networks*, 32(3): 221–233. 23

Feinberg, M., Willer, R., and Schultz, M. (2014). "Gossip and Ostracism Promote Cooperation in Groups." *Psychological Science*, 25(3): 656–664. 18

Fisher, R. A. (1921). "On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample." *Metron*, 1: 3–32. 21

Geary, R. C. (1954). "The Contiguity Ratio and Statistical Mapping." *The Incorporated Statistician*, 5(3): 115–146. 9

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd edition. MR3235677. 21

Gelman, A. and Rubin, D. B. (1992). "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science*, 457–472. 15

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., and Modrák, M. (2020). "Bayesian Workflow." MR4298989. doi: https://doi.org/10.1214/20-ba1221. 10

Goeyvaerts, N., Santermans, E., Potter, G., Torneri, A., Van Kerckhove, K., Willem, L., Aerts, M., Beutels, P., and Hens, N. (2018). "Household Members Do Not Contact Each Other at Random: Implications for Infectious Disease Modeling." *Proceedings of the Royal Society B: Biological Sciences*, 285(1893): 20182201. 24

Greene, D. and Cunningham, P. (2013). "Producing a Unified Graph Representation from Multiple Social Network Views." In *Proceedings of the 5th Annual ACM Web Science Conference*, WebSci '13, 118–121. New York, NY, USA: Association for Computing Machinery. 2

Harrer, M., Cuijpers, P., A, F. T., and Ebert, D. D. (2021). *Doing Meta-Analysis With R: A Hands-On Guide*. Boca Raton, FL and London: Chapman & Hall/CRC Press, 1st edition. 21

Hmimida, M. and Kanawati, R. (2015). "Community Detection in Multiplex Networks: A Seed-Centric Approach." *Networks and Heterogeneous Media*, 10(1): 71–85. MR3359512. doi: `https://doi.org/10.3934/nhm.2015.10.71`. 2

Hoff, P., Fosdick, B., Volfovsky, A., and Stovel, K. (2013). "Likelihoods for Fixed Rank Nomination Networks." *Network Science*, 1(3): 253–277. 3

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). "Latent Space Approaches to Social Network Analysis." *Journal of the American Statistical Association*, 97(460): 1090–1098. MR1951262. doi: `https://doi.org/10.1198/016214502388618906`. 2

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). "Stochastic Blockmodels: First Steps." *Social Networks*, 5(2): 109–137. MR0718088. doi: `https://doi.org/10.1016/0378-8733(83)90021-7`. 2

Holland, P. W. and Leinhardt, S. (1970). "A Method for Detecting Structure in Sociometric Data." *American Journal of Sociology*, 76(3): 492–513. 9

Holland, P. W. and Leinhardt, S. (1981). "An Exponential Family of Probability Distributions for Directed Graphs." *Journal of the American Statistical Association*, 76(373): 33–50. 2

Hong, A. and Niezink, N. (2025). "Supplement to "The Multiplex p2 Model: Mixed-Effects Modeling for Multiplex Social Networks"." doi: `https://doi.org/10.1214/25-BA1527SUPP`. 8, 15, 17, 24

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008). "Goodness of Fit of Social Network Models." *Journal of the American Statistical Association*, 103(481): 248–258. MR2394635. doi: `https://doi.org/10.1198/016214507000000446`. 9

Kadushin, C. (2012). *Understanding Social Networks: Theories, Concepts, and Findings*. Oxford University Press. 1

Keating, N. L., Ayanian, J. Z., Cleary, P. D., and Marsden, P. V. (2007). "Factors Affecting Influential Discussions Among Physicians: A Social Network Analysis of a Primary Care Practice." *Journal of General Internal Medicine*, 22(6): 794–798. 2

Kisfalusi, D. (2018). "Bullies and Victims in Primary Schools: The Associations Between Bullying, Victimization, and Students' Ethnicity and Academic Achievement." *Intersections. East European Journal of Society and Politics*, 4(1). 18, 20

Kisfalusi, D., Janky, B., and Takács, K. (2021). "Grading in Hungarian Primary Schools: Mechanisms of Ethnic Discrimination against Roma Students." *European Sociological Review*, 37(6): 899–917. 20

Kisfalusi, D., Takács, K., and Pál, J. (2019). *Gossip and Reputation in Adolescent Networks*, 359–379. Oxford University Press. 18

Krackhardt, D. (1988). "Predicting with Networks: Nonparametric Multiple Regression Analysis of Dyadic Data." *Social Networks*, 10(4): 359–381. MR0984597. doi: https://doi.org/10.1016/0378-8733(88)90004-4. 3

Krivitsky, P. N., Coletti, P., and Hens, N. (2023). "A Tale of Two Datasets: Representativeness and Generalisability of Inference for Samples of Networks." *Journal of the American Statistical Association*, 1–21. MR4766024. doi: https://doi.org/10.1080/01621459.2024.2344624. 24

Krivitsky, P. N., Handcock, M. S., and Morris, M. (2011). "Adjusting for Network Size and Composition Effects in Exponential-Family Random Graph Models." *Statistical Methodology*, 8(4): 319–339. MR2800354. doi: https://doi.org/10.1016/j.stamet.2011.01.005. 24

Leifeld, P. and Schneider, V. (2012). "Information Exchange in Policy Networks." *American Journal of Political Science*, 56(3): 731–744. 3, 13, 14

Leinhardt, S. (1971). "The Structure of Positive Interpersonal Relations in Small Groups." In Berger, J. (ed.), *Sociological Theories in Progress*, 218–251. Hanover: Dartmouth College. 9, 19

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). "Generating Random Correlation Matrices Based on Vines and Extended Onion Method." *Journal of Multivariate Analysis*, 100(9): 1989–2001. MR2543081. doi: https://doi.org/10.1016/j.jmva.2009.04.008. 7

Liu, H., Zhang, Z., and Grimm, K. J. (2016). "Comparison of Inverse Wishart and Separation-Strategy Priors for Bayesian Estimation of Covariance Parameter Matrix in Growth Curve Analysis." *Structural Equation Modeling: A Multidisciplinary Journal*, 23(3): 354–367. MR3488827. doi: https://doi.org/10.1080/10705511.2015.1057285. 7

Lubbers, M. J. (2003). "Group Composition and Network Structure in School Classes: A Multilevel Application of the p* Model." *Social Networks*, 25(4): 309–332. 24

Minhas, S., Hoff, P. D., and Ward, M. D. (2019). "Inferential Approaches for Network Analysis: AMEN for Latent Factor Models." *Political Analysis*, 27(2): 208–222. 24

Modrák, M., Moon, A. H., Kim, S., Bürkner, P., Huurre, N., Faltejsková, K., Gelman, A., and Vehtari, A. (2023). "Simulation-based calibration checking for Bayesian computation: The choice of test quantities shapes sensitivity." *Bayesian Analysis*, Advance publication. MR4506021. doi: https://doi.org/10.1214/21-ba1287. 11

Moran, P. A. (1948). "The Interpretation of Statistical Maps." *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2): 243–251. MR0029115. 9

Neal, R. M. (2011). "MCMC using Hamiltonian Dynamics." In Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (eds.), *Handbook of Markov Chain Monte Carlo*, chapter 5. Chapman and Hall/CRC, 1st edition. MR2858447. 6

Nestler, S. (2018). "Likelihood Estimation of the Multivariate Social Relations Model." *Journal of Educational and Behavioral Statistics*, 43(4): 387–406. 3

Niezink, N. M. D. (2023). "Discussion of "A Tale of Two Datasets: Representativeness and Generalisability of Inference for Samples of Networks"." *Journal of the American Statistical Association*, 118(544): 2232–2234. 24

Ogle, K. and Barber, J. J. (2020). "Ensuring identifiability in hierarchical mixed effects Bayesian models." *Ecological Applications*, 30(7): e02159. URL `https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/eap.2159` 8

Rambaran, J. A., Dijkstra, J. K., and Veenstra, R. (2020). "Bullying as a Group Process in Childhood: A Longitudinal Social Network Analysis." *Child Development*, 91(4): 1336–1352. 1

Redhead, D., Gervais, M., Kajokaite, K., Koster, J., Hurtado Manyoma, A., Hurtado Manyoma, D., McElreath, R., and Ross, C. T. (2024). "Evidence of direct and indirect reciprocity in network-structured economic games." *Communications Psychology*, 2(1): 44. 4

Robins, G. (2015). *Doing Social Network Research: Network-Based Research Design for Social Scientists*. Los Angeles: Sage Publications Ltd. 7

Salter-Townshend, M. and McCormick, T. H. (2017). "Latent Space Models for Multiview Network Data." *The Annals of Applied Statistics*, 11(3): 1217–1244. MR3709558. doi: `https://doi.org/10.1214/16-AOAS955`. 4

Schad, D. J., Betancourt, M., and Vasishth, S. (2021). "Toward a Principled Bayesian Workflow in Cognitive Science." *Psychological Methods*, 26(1): 103–126. 10, 12

Schneider, V. (1988). *Politiknetzwerke der Chemikalienkontrolle: Eine Analyse Einer Transnationalen Politikentwicklung*, volume 10. Berlin/Boston: Walter de Gruyter GmbH. 13

Smith, S., Maas, I., and van Tubergen, F. (2014). "Ethnic Ingroup Friendships in Schools: Testing the By-Product Hypothesis in England, Germany, the Netherlands, and Sweden." *Social Networks*, 39: 33–45. 2

Snijders, T. A., Lomi, A., and Torló, V. J. (2013). "A Model for the Multiplex Dynamics of Two-Mode and One-Mode Networks, with an Application to Employment Preference, Friendship, and Advice." *Social Networks*, 35(2): 265–276. MR3074604. doi: `https://doi.org/10.1007/978-1-4614-1800-9_129`. 4

Snijders, T. A., Ripley, R., B'oda, Z., V"or"os, A., and Preciado, P. (2024). *Manual for RSiena*. University of Groningen, Groningen, The Netherlands. R package version

1.4.7.
URL https://www.stats.ox.ac.uk/~snijders/siena/ 17, 19

Snijders, T. A. B. (2001). "The Statistical Evaluation of Social Network Dynamics." *Sociological Methodology*, 31(1): 361–395. 4

Sosa, J. and Betancourt, B. (2022). "A Latent Space Model for Multilayer Network Data." *Computational Statistics & Data Analysis*, 169: 107432. MR4369145. doi: https://doi.org/10.1016/j.csda.2022.107432. 4

Spillane, J. P., Hopkins, M., and Sweet, T. M. (2015). "Intra- and Interschool Interactions about Instruction: Exploring the Conditions for Social Capital Development." *American Journal of Education*, 122(1): 71–110. 2

Stan Development Team (2022). "Stan Modelling Language Users Guide and Reference Manual v. 2.32."
URL https://mc-stan.org 3, 6

Stan Development Team (2023a). "RStan: the R Interface to Stan." R package version 2.21.8.
URL https://mc-stan.org/ 6

Stan Development Team (2023b). *Running Stan on the GPU with OpenCL*. Stan Development Team. CMDStanR Documentation. Accessed: April 23, 2024.
URL https://mc-stan.org/cmdstanr/articles/opencl.html 24

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., and Gelman, A. (2020). "Validating Bayesian Inference Algorithms with Simulation-Based Calibration." 10

Tatum, T. G. and Grund, T. U. (2020). "Accusation and Confession Discrepancies in Bullying: Dual-Perspective Networks and Individual-Level Attributes." *Social Networks*, 60: 61–70. 2, 18

Tokuda, T., Goodrich, B., Van Mechelen, I., Gelman, A., and Tuerlinckx, F. (2011). "Visualizing Distributions of Covariance Matrices." Working paper, Columbia University. 7

Tolsma, J., van Deurzen, I., Stark, T. H., and Veenstra, R. (2013). "Who Is Bullying Whom in Ethnically Diverse Primary Schools? Exploring Links Between Bullying, Ethnicity, and Ethnic Diversity in Dutch Primary Schools." *Social Networks*, 35: 51–61. 2, 18

van Duijn, M. A. J., Snijders, T. A. B., and Zijlstra, B. J. H. (2004). "p2: A Random Effects Model with Covariates for Directed Graphs." *Statistica Neerlandica*, 58: 234–254. MR2064846. doi: https://doi.org/10.1046/j.0039-0402.2003.00258.x. 2, 6

Veenstra, R., Lindenberg, S., Zijlstra, B. J. H., Winter, A. F. D., Verhulst, F. C., and Ormel, J. (2007). "The Dyadic Nature of Bullying and Victimization: Testing a Dual-Perspective Theory." *Child Development*, 78: 1843–1854. 2, 18, 19

Vermeij, L., van Duijn, M. A., and Baerveldt, C. (2009). "Ethnic Segregation in Context: Social Discrimination Among Native Dutch Pupils and Their Ethnic Minority Classmates." *Social Networks*, 31(4): 230–239. 2

Wang, P., Robins, G., Pattison, P., and Lazega, E. (2013). "Exponential Random Graph Models for Multilevel Networks." *Social Networks*, 35(1): 96–115. MR3074657. doi: https://doi.org/10.1007/978-1-4614-1800-9_182. 4

Wang, P., Robins, G., Pattison, P., and Lazega, E. (2016). "Social Selection Models for Multilevel Networks." *Social Networks*, 44: 346–362. 4

Zijlstra, B., Veenstra, R., and Duijn, M. (2008). "A Multilevel p2 Model with Covariates for the Analysis of Binary Bully-Victim Network Data in Multiple Classrooms." In Card, N., Selig, J., and Little, T. (eds.), *Modeling Dyadic and Interdependent Data in the Developmental and Behavioral Sciences*, 369–386. Routledge. 2

Zijlstra, B. J. (2021). *dyads: Dyadic Network Analysis*. R package version 1.1.4. URL https://CRAN.R-project.org/package=dyads 3

Zijlstra, B. J. H. (2009). "Random effects models for directed graphs with covariates." [Thesis fully internal (DIV), Rijksuniversiteit Groningen]. s.n. 2

Zijlstra, B. J. H., van Duijn, M. A. J., and Snijders, T. A. B. (2006). "The Multilevel p2 Model." *Methodology*, 2(1): 42–47. MR3137621. 2

Zijlstra, B. J. H., van Duijn, M. A. J., and Snijders, T. A. B. (2009). "MCMC Estimation for the p2 Network Regression Model with Crossed Random Effects." *British Journal of Mathematical and Statistical Psychology*, 62(1): 143–166. MR2518398. doi: https://doi.org/10.1348/000711007X255336. 6, 7, 8