Exponential-Family Random Graph Models for Multi-Layer Networks

Pavel N. Krivitsky[a,*], Laura M. Koehly[b], Christopher Steven Marcum[b]

[a]*Department of Statistics, School of Mathematics and Statistics, UNSW Sydney, NSW, Australia*
[b]*National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA*

Abstract

Multi-layer networks arise when more than one type of relation is observed on a common set of actors. Modeling such networks within the exponential-family random graph (ERG) framework has been previously limited to special cases and, in particular, to dependence arising from just two layers. Extensions to ERGMs are introduced to address these limitations: Conway–Maxwell-Binomial distribution to model the marginal dependence among multiple layers; a "layer logic" language to translate familiar ERGM effects to substantively meaningful interactions of observed layers; and non-degenerate triadic and degree effects. The developments are demonstrated on two previously published data sets.

Keywords: ERGM Multi-Layer Multi-Relational Multiplexity Conway–Maxwell-Binomial

1. Background

Studying social processes unfolding in networks consisting of a common set of actors incident on multiple types of relationships has long been of interest across the social sciences. Variously called *multivariate* or *multiplex*, *multi-item*, and *multi-relational* networks (P. E. Pattison 1982; S. E. Fienberg, Meyer, and

---

*Corresponding Author
Email address:* `p.krivitsky@unsw.edu.au` (Pavel N. Krivitsky)

Wasserman 1980; Frank and Shafie 2016; Jeub et al. 2017; Voros and Snijders 2017), these social structures typically involve connections among actors on more than one type of binary relationship, which are conceptually "stacked" into layers (Magnani and Wasserman 2017). This type of layered network data can be used to address research questions about social structure and processes that may unfold both within and between layers. Multi-layer models would allow one to consider questions such as: what is the likelihood of forming a friendship tie given the presence of a trust tie between two actors? Do friendships tend to form between actors who consider the same people to be trustworthy? Previous network analytic approaches to multi-layer networks include a combination of covariance based measures (i.e., correlation, singular value decomposition, canonical correlation), descriptive measures (i.e., joint density, multiplexity, edit distance) and model-based inference (i.e., entailment analysis, stochastic block-models, network regression and two-layer exponential-family random graph models). These important foundational techniques are limited in that they provide few tools for analyzing cross-layer structural effects that are often of substantive interest to researchers. Moreover, methods that handle association across many network layers jointly while considering structural properties both within and between layers are not currently available. This paper addresses these gaps by building upon the foundations of multi-layer network analysis.

The balance of this manuscript is organized as follows. First, a brief literature review of modeling multiple layers using the exponential-family random graph family of models (ERGMs) is provided. Second, three novel advances for the statistical analysis of an arbitrary number of layers in a network using this framework is described: section 2 proposes the Conway–Maxwell-Binomial distribution for modeling marginal dependence across layers; section 3, introduces *layer logic*—a language for modeling joint and conditional dependence across layers;

and section 4 introduces non-degenerate cross-layer triadic effects. Combined, these contributions provide methods to address research questions about how social structure unfolds within and between multiple layers of relationships, while controlling for the underlying marginal association among the layers. Finally, in section 5, two empirical analyses using previously published data (Lazega's Lawyers and a subset of Knecht's Students) are presented.

1.1. Previous Work on Multi-Layer Exponential-Family Random Graph Models

There is a long history of random graph approaches for modeling both local and global structures within observed network systems, the majority of which focus on single binary relations (Erdős and Rényi 1959; Frank and Strauss 1986; Holland and Leinhardt 1981; S. Fienberg and Wasserman 1981; Wasserman and Pattison 1996). However, network processes, particularly when considering relationships that span great lengths of time between actors, often involve multiple relations. To investigate associations among those multiple relations, methods used include relational algebras and block modeling approaches (D. R. White and Reitz 1983; H. C. White, Boorman, and Breiger 1976; Boorman and White 1976; Barbillon et al. 2017; P. E. Pattison 1982), correlational approaches (Krackardt 1987; Wasserman, Faust, and Galaskiewicz 1990), and comparisons to null distributions, such as multivariate uniform graph distributions (S. Fienberg and Wasserman 1981). Extensions have been developed that considered local structural characteristics within a stochastic modeling framework were initiated over 30 years ago where the dyadic independence model was augmented to incorporate multiple relations (Wasserman 1987). The stochastic blockmodels of Nowicki and Snijders (2001) incorporated multilayered relations. In the 1990s, Wasserman and Pattison (1996) adapted random graph models of Frank and Strauss (1986) to larger network systems with more complicated dependence structures and made innovations in parameter estimation. With these newly

specified p^* models, or ERGMs, came the potential to move beyond dyadic independence and to hypothesize and fit models with more complex dependence structures. In so doing, the door was opened to consider the potential for such hypothesized dependencies to be characterized by interdependent relations of different types (Koehly and Pattison 2005).

In the late 1990s, ERGMs were applied to several multivariate network datasets, most notably resource exchanges among lawyers in Lazega's classic dataset (Lazega and Pattison 1999). These models were also redefined to consider temporal networks, with longitudinal measurements on the same relations (G. Robins and Pattison 2001). Lazega and Pattison (1999), in particular, proposed an initial taxonomy of possible triad configurations for directed networks. Unfortunately, the type of triad-based specification used has since been found to induce poorly behaved models in a phenomenon that has come to be called *degeneracy* (Handcock 2003; Schweinberger 2011). In the paper of Lazega and Pattison (1999), it may have been masked by their use of maximum pseudo-likelihood estimation (MPLE) (Strauss and Ikeda 1990) as opposed to the more accurate MCMC-based methods (Snijders 2002; Hunter and Handcock 2006) that were developed a few years later. (Schweinberger et al. (2020) provides a more detailed historical account.) A more recent chapter by Wang (2012) described a number of specifications for two-layer networks, reviewing the models introduced by Lazega and Pattison (1999) and introducing two-layer extensions of the non-degenerate specifications of (Snijders et al. 2006), implemented in the `XPNet` software.

Ultimately, applications that consider multi-layer networks are still limited, due, largely, to the increased complexity of and inability to fit such multi-layer models with more than a few relations of interest. Given the complex social worlds within which we live and the importance of interpersonal mechanisms on health,

well-being, and mortality (Holt-Lunstad, Smith, and Layton 2010), development of models to understand how varying types of social ties are structured has important implications. As such, there has been significant effort within the community towards identifying methodological approaches to characterize these complex social systems. Moreover, one of the advantages of estimating joint models with many layers—rather than simply fitting separate ERGMs for each layer—is that within-model contrasts between differences in social processes at play in the separate layers can be evaluated.

These methodological advances bring with them some challenges to implementation. First, as discussed by Voros and Snijders (2017), the number of structural parameters that can characterise a structural model increases exponentially as the number of layers defining the system increases. As such, there is a plethora of interpersonal mechanisms that may explain potential cross-layer dependencies requiring model specification to be hypothesis driven and grounded in theory. Second, logistical and computational aspects related to model specification and estimation become more complicated in multi-layer contexts, particularly when one is aiming to move beyond correlation between layers. There has been success in implementing such models with just a few relational layers, including among one-mode multi-layer networks and between two-mode and one-mode networks (Huitsing et al. 2012; Snijders, Lomi, and Torló 2013; Wang 2012). But, extensions moving beyond just a few layers are scarce and most of those extensions are limited in their ability to map to hypothesised structural characteristics that might be threaded across layers. For example, Frank and Shafie (2016) and Shafie (2015) consider the use of entropy to identify interdependencies among a set of network variables derived from multigraphs. Salter-Townshend and McCormick (2017) propose latent space analysis as a way to understand dependencies across layers conditional on within layer structure. Others have used data reduction

approaches, such as cluster analysis, to reduce the dimensionality of layers and, thus, simplify the multigraph problem (Voros and Snijders 2017). And, De Domenico et al. (2013) demonstrate the use of tensor algebra for investigating structural aspects of multi-layer networks that can account for both multiplex relations and temporal dynamics.

These developments notwithstanding, there exists a considerable gap in the theory and estimation of ERGMs for multi-layer networks. In what follows, our contribution builds upon this history from within the ERGM framework. We begin with some notation for multi-layer networks.

1.2. Notation

Let $\boldsymbol{Y}$ be a random multi-layer network among a common set of actors $N = \{1, \ldots, n\}$, on L binary relations where $Y_{i,j,l}$ indexes the presence ($= 1$) or absence ($= 0$) of the l-th relation between actors i and j ($i \neq j \, \forall \, l$). Let $\boldsymbol{y}$ be a realization of $\boldsymbol{Y}$. This multi-layer network can be regarded as a composition of L binary networks, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_L$ or as a multivariate network, where each dyad $\vec{y}_{i,j}$ is a binary L-vector $[y_{i,j,l}]_{l=1}^{L}$.

1.3. Multilayer ERGM

To specify an ERGM on such a network, define the set of potential relations within each layer: $\mathbb{Y}_l$, $l = 1, \ldots, L$. For a directed layer, $\mathbb{Y}_l \subseteq \{(i,j) : (i,j) \in N \times N\}$, and for an undirected layer, $\mathbb{Y}_l \subseteq \{\{i,j\} : (i,j) \in N \times N\}$. For convenience, also define $\mathbb{Y}_. \equiv \mathbb{Y}_1 \cup \cdots \cup \mathbb{Y}_L$: the set of dyads (ordered or unordered) for which a relation may exist in at least one layer. In the case where both directed and undirected layers are modeled jointly, it is convenient to treat $\mathbb{Y}_.$ as directed, with an undirected $\mathbb{Y}_l$ being treated as $\{(i,j), (j,i) : \{i,j\} \in \mathbb{Y}_l\}$ for the purpose of taking this union. The sample space for the ERGM is the set of possible subsets of potential relations of interest: $\mathcal{Y} \subseteq 2^{\mathbb{Y}_1 \times \cdots \times \mathbb{Y}_L}$.

An ERGM is specified by its vector function $\boldsymbol{g}(\boldsymbol{Y})$, sufficient statistics, operationalizing the multilayer network's features of interest—including those of its constituent layers individually or, as discussed later, various combinations thereof—and, in some formulations (Hunter and Handcock 2006; Krivitsky 2017), a mapping $\boldsymbol{\eta}(\boldsymbol{\theta})$ of model parameter vector $\boldsymbol{\theta}$ to the natural parameters. Then, write $\boldsymbol{Y} \sim \mathrm{ERGM}_{\mathcal{Y},\boldsymbol{g},\eta}(\boldsymbol{\theta})$ if for $\boldsymbol{y} \in \mathcal{Y}$,

$$\Pr(\boldsymbol{Y} = \boldsymbol{y}) = \exp\{\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \boldsymbol{g}(\boldsymbol{y})\}/\kappa(\boldsymbol{\theta}),$$

where $\kappa(\boldsymbol{\theta}) = \sum_{\boldsymbol{y}' \in \mathcal{Y}} \exp\{\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \boldsymbol{g}(\boldsymbol{y}')\}$ is the normalizing constant (Hunter and Handcock 2006).

1.4. Homogeneous and Heterogeneous Layer Effects

Before introducing models for dependence between layers, it is helpful to consider how a model may "couple" layers even under the assumption of conditional independence. Observe that an ERGM may be specified for each layer individually, that is,

$$\Pr(\boldsymbol{Y}_l = \boldsymbol{y}_l) = \exp\{\boldsymbol{\eta}_l(\boldsymbol{\theta}_l) \cdot \boldsymbol{g}_l(\boldsymbol{y}_l)\}/\kappa_l(\boldsymbol{\theta}_l),$$

with the likelihood that results being

$$\prod_{l=1}^{L} \Pr(\boldsymbol{Y}_l = \boldsymbol{y}_l) = \exp\{\sum_{l=1}^{L} \boldsymbol{\eta}_l(\boldsymbol{\theta}_l) \cdot \boldsymbol{g}_l(\boldsymbol{y}_l)\}/\prod_{l=1}^{L} \kappa_l(\boldsymbol{\theta}_l).$$

This is a single ERGM, whose sufficient statistic $\boldsymbol{g}(\boldsymbol{y}) \equiv [\boldsymbol{g}_1(\boldsymbol{y}_1)^\top, \ldots, \boldsymbol{g}_L(\boldsymbol{y}_L)^\top]^\top$, parameter vector $\boldsymbol{\eta}(\boldsymbol{\theta}) \equiv [\boldsymbol{\eta}(\boldsymbol{\theta})_1^\top, \ldots, \boldsymbol{\eta}(\boldsymbol{\theta})_L^\top]^\top$, and $\kappa(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L) \equiv \prod_{l=1}^{L} \kappa_l(\boldsymbol{\theta}_l)$.

Even if forms of corresponding $g_k(\cdot)$ for two layers are the same, no assumption is made about whether the social forces they embody have the same strength and direction—they are heterogeneous. However, one can also consider models

in which the strength and the direction are assumed to be the same—net of other social forces—even though the layers themselves are still stochastically independent. This can provide better parsimony and more statistical power, and heterogeneity of forces may be tested using likelihood-ratio testing or score testing. That is, if effect $g_k(\cdot)$ is assumed to be homogeneous among the layers with coefficient $\eta_k(\boldsymbol{\theta})$, it may be expressed through a joint model

$$
\begin{aligned}
\Pr(\boldsymbol{Y}_1 = \boldsymbol{y}_1, \ldots, \boldsymbol{Y}_L = \boldsymbol{y}_L) &= \prod_{l=1}^{L} \exp\{\boldsymbol{\eta}_l(\boldsymbol{\theta}_l) \cdot \boldsymbol{g}_l(\boldsymbol{y}_l) + \eta_k(\boldsymbol{\theta}) \cdot g_k(\boldsymbol{y}_l)\} \big/ \kappa_l(\boldsymbol{\theta}_l) \\
&= \exp\{\sum_{l=1}^{L} \boldsymbol{\eta}_l(\boldsymbol{\theta}_l) \cdot \boldsymbol{g}_l(\boldsymbol{y}_l) + \eta_k(\boldsymbol{\theta}) \cdot \sum_{l=1}^{L} g_k(\boldsymbol{y}_l)\} \big/ \kappa(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L),
\end{aligned}
$$

with a sufficient statistic simply summing the layer-specific statistics. Homogeneous effects can also be applied to subsets of layers.

2. The Conway–Maxwell-Binomial Model

Layers, which are assumed are composed of binary random variables, may or may not be independent and multi-layer ERGM model terms may depend on linear combinations of edges across layers. This is particularly the case when the research question involves multi-layer networks comprising some mixture of underlying relational constructs, such as cohesion and conflict identified by Koehly and Marcum (2016). Thus, special consideration should be given to the distribution of edges across a set of L binary relations in a multi-layer network. One of the principal challenges in modeling associated Bernoulli random variables is the asymmetry between positive and negative pairwise correlations. Kadane (2016) proposes an approach to this problem via the *Conway–Maxwell-binomial* (CMB) distribution. Here, this approach is used to model the marginal dependence among all dyads across layers in which they have a tie.

A random variable $X \in \{0, \ldots, m\}$ is CMB if, up to a constant,[1]

$$\Pr(X = x; \pi, \gamma) \propto \binom{m}{x}^{1-\gamma} \pi^x (1-\pi)^{m-x}. \tag{1}$$

Then, if $\gamma = 0$, X is binomial. When $\gamma > 0$, the denominator of the binomial coefficient (i.e., $x!(m-x)!$) is attenuated, thus increasing the variance for a given π; and, conversely, when $\gamma < 0$, the binomial coefficient is amplified, inducing underdispersion relative to the binomial distribution with a given π. Viewed as a sum of Bernoulli random variables, higher variance (for a given expectation) corresponds to a positive correlation among the variables, and a lower variance corresponds to a negative correlation.

CMB is an exponential family, and CMB distribution for the number of layers within a dyad can be induced with $g_{\mathrm{CMB}}(\boldsymbol{y}) = -\sum_{(i,j)\in\mathbb{Y}_{\cdot}} \log\binom{L}{y_{i,j,\cdot}}$, where $y_{i,j,\cdot} = \sum_{l=1}^{L} y_{i,j,l}$, the number of relations between i and j across all layers. Such ties may be directed or undirected, and in the latter case only the upper or lower triangle of the layer's associated adjacency matrix need be evaluated. Observe that in an ERGM whose only element of $\boldsymbol{g}(\cdot)$ is the total number of edges across all layers ($g_{\mathrm{edge}}(\boldsymbol{y}) = \sum_{(i,j)\in\mathbb{Y}_{\cdot}} y_{i,j,\cdot}$), for any given (i,j), $Y_{i,j,\cdot} \sim \mathrm{Binomial}\{L, \mathrm{logit}^{-1}(\theta_{\mathrm{edge}})\}$. Then, adding $g_{\mathrm{CMB}}(\boldsymbol{y})$ to the model would turn this distribution into the CMB specification above, with $\gamma \equiv \theta_{\mathrm{CMB}}$. The CMB statistic can also be computed across only a subset of layers. For example, a positive association may be expected among layers measuring relations with positive valence and among those with negative valence, but likely not between the two subsets. Practically, although CMB has an intractable normalizing constant, in an ERGM context it is absorbed into $\kappa(\boldsymbol{\theta})$, requiring no additional computational effort (Krivitsky 2012).

---

[1]This parametrization is slightly different from Kadane's.

Other coefficients held constant, the coefficients associated with the CMB term in an ERGM therefore have a convenient associational interpretation. The sign of θ_{CMB} determines whether on average, the dependence across the relations is positive or negative. As a corollary, this result may be useful for simulating underlying association between graphs from fitted models employing CMB terms, thus providing one mechanism for drawing reference distributions for multi-layer graph correlations (akin to Kadane's intent for simulating binomial random variates with known correlation).

A limitation of this approach is that it entails a homogeneity assumption of sorts: strictly speaking, the CMB form in (1) implies that success probabilities are equal for all constituent binary observations, which in our context implies that all layers evaluated by a CMB term have equal densities. When this does not hold, the distribution will not be strictly CMB—but the effect of the sign of θ_{CMB} is not affected.

3. Layer Logic

Multi-layer network data may comprise a large number of layers which may be substantively associated in a variety of ways. That is, layers may depend upon one another. This dependence can be modeled marginally using CMB, but one may also wish to model some specific structural characteristics that describe cross-layer dependencies in a conditional and joint manner. Such conditional dependencies may take the form of cross-layer (i.e., cross-relation) effects where the presence or absence of a tie in one layer conditionally affects the presence or absence of ties in another. As discussed in section 1, there is considerable interest in these types of cross-layer effects in the published literature. However, descriptions of such effects are distributed throughout the literature and no single source lingua franca currently exists. Thus, this work introduces a general

language for such interactions across layers. This language is based on Boolean logic and so we call it *layer logic*.

Perhaps the most well-studied conditionally dependent cross-layer effect is entailment (D. R. White 1996). The presence of a trust relation between two actors, i and j, for instance, may be necessary for the presence of an exchange relation between them. Or, the absence of conflict and disrespect relations between i and j may be required for a friendship to form between them. While such dyadwise cross-layer effects are interesting, one may also be interested in higher-order tuples of actors across relations (such as triads). Moreover, substantively, the relations represented by observed layers in the data may not be the same ones as those one might wish to model. For example, suppose that one wishes to model a scenario in which an individual i's next-door neighbor k recommends laborer j, whom k regularly hires to perform yard-work, and i subsequently also hires j. Here, i and k have a "live close together" relationship, and k and j have an "engage in commerce" relationship, and the triad is closed by i and j also establishing an "engage in commerce" relationship. However, one would expect a similar effect if i and k were coworkers or personal friends or some other form of acquaintances, instead of, or in addition to, being neighbors. One might thus want to construct a *logical layer* "acquaintanceship", in which i and k have a relation if they are neighbors, coworkers, or friends (or any combination thereof). One could then formulate a triadic closure term of these layers, such as that if i and k are acquainted, and k and j engage in commerce, that increases the probability that i and j will engage in commerce. Thus, layer logic represents a flexible system for theorizing, representing, and modeling multiplex relationships between actors.

In the balance of this section, a layer logic is formalized and described that encompasses, generally, arbitrary combinations of multi-layer effects including

those described above. As the most simple cases in any Boolean logic involve relations between two binary elements, our descriptions involve two hypothetical relations "A" and "B" (and employing a third "C" where necessary). However, this layer logic applies to any number of relations, and our technical implementation is general enough to model conditional dependence across layers in an arbitrary fashion, provided that the researcher has a theoretical reason to do so.

3.1. Elements of Layer Logic

Elements of layer logic include the following:

Conjunction $A \wedge B$**:** A logical layer $A \wedge B$ is considered to have a relationship between i and j if there are relationships between them in both layers A and B. A simple example where this form may be useful is for a network which includes familial relations ("A"), coworker relations ("B"), and advising/mentoring relations ("C"). Substantively, advice and mentoring in the context of one's employment is likely to be different in its character and content from advice and mentoring in the context of one's family, so one may distinguish $A \wedge C$ from $B \wedge C$. In the literature on social networks, this effect is referred to as "entrainment" by Lusher and Robins (2013) and is often induced by an intersection-rule of two or more relations as described by Koehly and Marcum (2016).

Disjunction $A \vee B$**:** A logical layer $A \vee B$ is considered to have a relationship between i and j if there are relationships between them in one or the other layer, or both. Substantively, this represents substitutability: either A or B can, for example, create the two-path needed to set up triadic closure. In the literature, this logical network is often induced by a union-rule of two or more relations.

Exclusive disjunction $A \oplus B$**:** A logical layer $A \oplus B$ is considered to have a relationship between i and j if there are relationships between them in one or the other layer, but not if it's in both. For example, suppose that one is interested in a set of family relationships where the layer logic returns a network

12

of kinship types meant to exclude half-related ties. In this case, since the network boundary is limited to family members, i and j can either be related in the logical layer network if they are consanguineous or social kin but are not considered related if the nature of their family relationship is a mixture of the two. Another example where exclusive disjunction may be of use is in the analysis of strictly non-ambiguous relationships between actors where a mixture of pro- and anti-social ties have been measured. For example, friendship and bullying in a high school, where the research question asks about joint structure between non-multiplex dyads.

Negation and other logical operations can also be used, and different elements can be combined. We call a given layer logic *sparse* if, in the absence of ties in all of the observed layers, there are no ties in the resulting logical layer: an empty observed network produces an empty logical layer. Similarly, a layer logic is *dense* if an empty observed network produces a complete logical layer. Naturally, logical layers that result from a layer logic on at least two layers will induce edges according to the truth-table (e.g., the matrix of possible binary outcomes) of the underlying logic. That is, logical layers inherit the properties of their logical operators, the outcome of which may depend on the number of layers employed. For example, an exclusive disjunction of layers $A \oplus B \oplus C$ results in a logical layer where the resulting network draws an edge between i and j in one or all layers, but an exclusive disjunction of layers $A \oplus B \oplus C \oplus D$ results in a more complicated logical layer where the resulting network has an edge between i and j when an edge is present in one or three of the layers but not otherwise. Thus, care must be taken on the part of the researcher when applying layer logic on multi-layer networks to ensure that the desired logical layers represent sensible relations.

In addition, layer operations can be used to model specific forms of dependence among the layers. The following examples provide a substantive interpretation and a description of their conditional effects on the model.

Conjunction: For a model with $L = 2$, Consider an ERGM of the form

$$\Pr(\boldsymbol{Y} = \boldsymbol{y}) \propto \exp\{\theta_1 \sum_{(i,j)\in\mathbb{Y}} y_{i,j,A} + \theta_2 \sum_{(i,j)\in\mathbb{Y}} y_{i,j,B} + \theta_3 \sum_{(i,j)\in\mathbb{Y}} y_{i,j,A\wedge B}\}. \quad (2)$$

one can express the third statistic as $\sum_{(i,j)\in\mathbb{Y}}(y_{i,j,A} \times y_{i,j,B})$; then, the conditional probability of $y_{i,j,A}$ can be expressed as $\text{logit}^{-1}(\theta_1 + \theta_3 y_{i,j,B})$. This means that $\theta_3 > 0$ induces a positive association between the two layers: presence of a relationship in layer B increases the probability of one in layer A and *vice versa*.

Disjunction: Consider instead a model of the form

$$\Pr(\boldsymbol{Y} = \boldsymbol{y}) \propto \exp\{\theta_1 \sum_{(i,j)\in\mathbb{Y}} y_{i,j,A} + \theta_2 \sum_{(i,j)\in\mathbb{Y}} y_{i,j,B} + \theta_3 \sum_{(i,j)\in\mathbb{Y}} y_{i,j,A\vee B}\}.$$

Reexpressing the last term as $\sum_{(i,j)\in\mathbb{Y}} \max(y_{i,j,A}, y_{i,j,B})$ gives the conditional probability of $y_{i,j,A}$ as $\text{logit}^{-1}\{\theta_1 + \theta_3(1 - y_{i,j,B})\}$ which, for $\theta_3 > 0$, induces a negative association, since absence of a relationship in layer B increases its probability in A.

Exclusive Disjunction: Lastly, a model of the form

$$\Pr(\boldsymbol{Y} = \boldsymbol{y}) \propto \exp\{\theta_1 \sum_{(i,j)\in\mathbb{Y}} y_{i,j,A} + \theta_2 \sum_{(i,j)\in\mathbb{Y}} y_{i,j,B} + \theta_3 \sum_{(i,j)\in\mathbb{Y}} y_{i,j,A\oplus B}\},$$

gives the conditional probability of $y_{i,j,A}$ as $\text{logit}^{-1}\{\theta_1 + \theta_3(-1)^{y_{i,j,B}}\}$: for $\theta_3 > 0$, presence of a relationship in layer B decreases its probability in A. This effect

can therefore be used to model mutual exclusivity between relationship types.

The three models specified above all have three free parameters and model a system with four possible states, whose probabilities sum to 1. It can be easily seen that they are reparametrizations of the same distribution of networks: as is the case with static network analysis in general, the model itself cannot resolve direction of causality, so care must be taken when specifying layer logic constructs that they reflect likely substantive effects.

On the other hand, it is important to distinguish an ERGM fit to a logical layer of a multilayer network from an ERGM fit to a network constructed by evaluating the same logical expression. This is because while the sufficient statistic may be the same, the normalizing constant in the denominator may be different. For example, using model (2) as a starting point, if one were to estimate θ_3 fixing $\theta_1 = \theta_2 = 0$, the likelihood would be

$$\Pr(\boldsymbol{Y} = \boldsymbol{y}) \propto \exp\{\theta_3 \sum_{(i,j)\in\mathbb{Y}} y_{i,j,A\wedge B}\}\Big/ \prod_{(i,j)\in\mathbb{Y}} \{3 + \exp(\theta_3)\}.$$

This is because one must consider four possible states of each dyad: one in which an edge is present in both A and B and three in which it is absent from at least one of the layers. On the other hand, if one models the intersection of A and B, i.e.,

$$\Pr(\boldsymbol{Y}_{A\wedge B} = \boldsymbol{y}_{A\wedge B}) \propto \exp\{\theta_3 \sum_{(i,j)\in\mathbb{Y}} y_{i,j,A\wedge B}\}\Big/ \prod_{(i,j)\in\mathbb{Y}} \{1 + \exp(\theta_3)\},$$

each dyad only has two states. This underscores the importance of considering sample spaces in ERGMs very carefully.

15

3.3. Directedness

Here, we briefly consider additional issues raised by directed layers. Firstly, some directed relations may be more meaningful (together or in logical layer combinations) with their ties reversed. For example, a researcher may consider reports of being bullied by someone to be more reliable than reports of bullying someone, but modelling bullying may be more substantively interpretable than modelling being bullied. Thus, the tie-reversal operation can be a useful part of the layer logic toolkit.

Secondly, it is also possible for observed layers to be heterogeneous in their directedness. For example, Nowicki and Snijders (2001) analyzed data collected by Kapferer (1972), in which two workers working together was an undirected relation, whereas friendship nominations between workers were directed. In the context of layer logic, the most natural approach to an operation on a directed and an undirected layer is to treat each edge in the undirected layer as a pair of directed reciprocating edges. (When the layer logic does not call for an operation between layers with different directedness, this conversion is not needed.)

4. Non-Degenerate Multi-Layer Triadic and Degree Effects

Layer logic permits flexible modeling of degree and triadic effects within a single layer—observed or logical—as such combinations of layers produce single relational structures on which any arbitrary graph statistic may be calculated. That is, given an observed or logical layer, commonly used ERGM terms may be used to model social process and structure on those networks. However, it may also be desirable to explicitly consider cross-layer triadic effects and degree effects in an ERGM. For example, in a directed network measuring relations "considers to be a friend" and "trusts", one may wish to distinguish between modeling whether i is more likely to trust j if i's friends are friends with j, if

16

i's friends trust j, and if people i trusts are friends with j. Moreover, one may wish to model the joint degree distribution resulting from the combination of ties accumulated in the trust and friendship layers. Here, we recognize that these cross-layer terms may be of special importance to researchers studying multi-layer networks.

As discussed in the introduction, the triadic specifications proposed by Lazega and Pattison (1999) may lead to degenerate models, and we therefore extend the more recent developments of Snijders et al. (2006) and Hunter (2007) to the multi-layer paradigm for both triadic and degree effects. We note that we are not the first to do so: Wang (2012) extended Snijders et al. (2006) formulation to two-layer networks. Our work differs in three ways: we extend the specification beyond two layers, we allow a more fine-grained specification of the triadic structure, and we reformulate the effects in the geometrically weighted terms of Hunter (2007).

4.1. Triadic Effects

In the parametrization of Hunter (2007 eq. (25)), the model term "Geometrically Weighted Edgewise Shared Partners" (GWESP) to represent friend-of-a-friend effects is

$$v(\boldsymbol{y}; \theta_1, \theta_2) = \theta_1 \, \mathrm{e}^{\theta_2} \sum_{a=1}^{n-2} \{1 - (1 - \mathrm{e}^{-\theta_2})^a\} \, \mathrm{EP}_a(\boldsymbol{y}),$$

where $\mathrm{EP}_a(\boldsymbol{y})$ is the number of extant relations (i, j) such that the number of 2-paths between i and j in the network is exactly a. Then θ_1 controls the overall strength and direction of the triad closure or anti-closure effect, and θ_2 controls the effect of each additional shared partner. The latter parameter can either be fixed as a tuning parameter, or (with some difficulty (Hunter and Handcock 2006; Stewart et al. 2019)) estimated from the data. $\mathrm{EP}_a(\cdot)$ can be

further decomposed into a sum of indicators that test whether there are exactly a 2-paths between i and j:

$$
\begin{aligned}
v(\boldsymbol{y}; \theta_1, \theta_2) &= \theta_1 \, \mathrm{e}^{\theta_2} \sum_{a=1}^{n-2} \{1 - (1 - \mathrm{e}^{-\theta_2})^a\} \sum_{(i,j) \in \mathbb{Y}} y_{i,j} \mathbb{I} \left\{ \sum_{k=1}^{n} y_{i,k} y_{k,j} = a \right\} \\
&= \theta_1 \, \mathrm{e}^{\theta_2} \sum_{(i,j) \in \mathbb{Y}} y_{i,j} \sum_{a=1}^{n-2} \{1 - (1 - \mathrm{e}^{-\theta_2})^a\} \mathbb{I} \left\{ \sum_{k=1}^{n} y_{i,k} y_{k,j} = a \right\} \\
&= \theta_1 \, \mathrm{e}^{\theta_2} \sum_{(i,j) \in \mathbb{Y}} y_{i,j} \{1 - (1 - \mathrm{e}^{-\theta_2})^{\mathrm{TP}_{i,j}(\boldsymbol{y})}\}
\end{aligned}
\tag{3}
$$

where $\mathrm{TP}_{i,j}(\boldsymbol{y}) = \sum_{k=1}^{n} y_{i,k} y_{k,j}$ is the number of two-paths between i and j, and for the case of $\theta_2 = 0$, 0^0 is defined to equal 1. "Geometrically Weighted Dyadwise Shared Partner (GWDSP)" (Hunter 2007, eq. (26)) can be constructed from (3) by omitting "$y_{i,j}$", and "Geometrically Weighted Non-edgewise Shared Partner" (GWNSP) can be constructed by replacing it with "$(1 - y_{i,j})$".

For directed networks, one may further contrast (to use the nomenclature of Butts (2008)) the *Outgoing Two-Paths* ($\mathrm{OTP}_{i,j}(\boldsymbol{y}) = \sum_{k=1}^{n} y_{i,k} y_{k,j}$) with *Incoming Two-Paths* ($\mathrm{ITP}_{i,j}(\boldsymbol{y}) = \sum_{k=1}^{n} y_{j,k} y_{k,i}$), *Outgoing Shared Partners* ($\mathrm{OSP}_{i,j}(\boldsymbol{y}) = \sum_{k=1}^{n} y_{i,k} y_{j,k}$), and *Incoming Shared Partners* ($\mathrm{ISP}_{i,j}(\boldsymbol{y}) = \sum_{k=1}^{n} y_{k,i} y_{k,j}$).

From this formulation, we propose to extend GWESP to multi-layer modeling by specifying two additional items: the layer in which the "$y_{i,j}$" is measured, and the layers in which the two-path may lie. In particular, one may specify in which (possibly different) layers the segments of the two-path or shared partnership lie and the ordering of layers in the directed two-path. (Recall the earlier example distinguishing the count of friends of i who trust j from the count of those whom i trusts who consider j a friend.) For OSP and ISP variants, one may similarly distinguish between counting friends of i whom j trusts from friends of j whom

i trusts, but only if presence or absence of $y_{i,j}$ is involved. Table 1 provides a taxonomy of the terms, as well as indication whether their direction and ordering is significant or whether it produces isomorphic structures.

In practice, the principle of marginality requires that each component sub-structure of any given triad be specified in a model aimed at identifying such triadic effects. For single-layer networks, this is relatively straightforward: the sub-structure terms one would introduce to model general triadic effects with a GWESP term in an undirected network under the principle of marginality include edges and two-stars. However, for multi-layer triadic effects, one must also consider sub-structure terms across all layers included in the triadic term. For instance, a three-layer undirected two-path GWESP term—such as that shown in the first row and column of table 1—one would need to specify edges terms for all three layers and three two-star terms reflecting the three combinations of layer segments that are centred about each vertex (i.e., in the table, the (i, k) red to (k, j) blue star, the (i, k) red to (i, j) green star, and (j, k) blue to (j, i) green star). Thus, care must be taken in model specification to ensure that the desired multi-layer triadic effect is appropriately identified. By extension, one may consider cross-layer terms for each pair of segments along a two-path or two-star centred on the topmost node described in table 1.

4.2. Degree Effects

The cross-layer degree and degree distribution effects can be modelled in our framework in a number of ways. Firstly, a common degree effect such as degree frequency, alternating-k-star (Snijders et al. 2006), or the geometrically weighted degree of Hunter (2007 eq. (14)) statistic can be evaluated on each layer of interest and summed to produce a homogeneous (per section 1.4) model. Secondly, a substantively meaningful logical layer can be constructed and a degree statistic of interest evaluated on it. For example, the cross-layer alternating-k-

Table 1: Triadic configurations of multi-layer networks.

	Undirected	Directed			
	TP	OTP	ITP	OSP	ISP
Edgewise					
Dyadwise					
Nonedgewise					
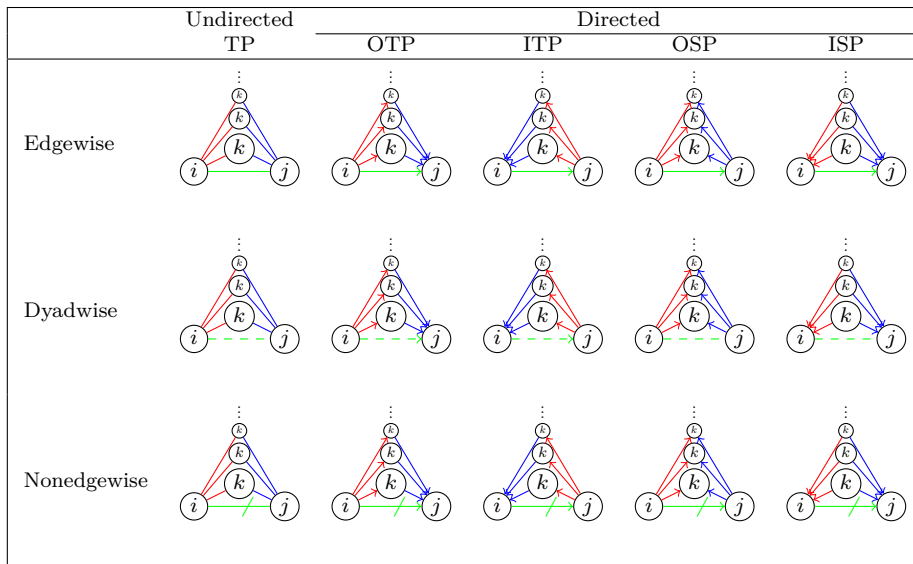
Layer: ▮ first segment ▮ second segment ▮ base

Base: ——— present ——/ absent – – – either

Order: ——— any ——→ specified

Note: Segments along any of the depicted triadic configurations may represent edges in up to three different layers (and thus, up to three different relations such as friendship, advice, and trust). Edgewise triadic configurations conditions the sufficient statistic associated with these terms on the presence of a base tie between i and j, nonedgewise conditions on the absence of a base tie between i and j, and dyadwise considers either case.

star effect described by Wang (2012, fig. 10.4) can be represented as an ordinary alternating-k-star effect evaluated on a disjunction. As a variant of the second approach, one may instead count a tie between i and j in both layer A and layer B as two separate ties for the purposes of degree.

Lastly, an important special case, identified by both Lazega and Pattison (1999) and Wang (2012, fig. 10.2), is cross-layer two-stars and two-paths. These should be used with some caution, as they have an unbounded change statistic (change due to adding or removing an edge) and may thus be prone to degeneracy (Schweinberger 2011), but are nonetheless very useful for representing a variety of phenomena.

In a directed context the modeller must also decide whether a configuration containing a tie from i to j in layer A and a tie from j to i in layer B—*exchange*—is a special case of a cross-layer two-path.

5. Empirical Examples

5.1. Data and Methods

We draw upon two previously published datasets for the empirical example component of this paper: Knecht's Netherlands Classroom dataset and Lazega's Lawyers dataset. We use two datasets to illustrate different aspects of the multi-layer ERGM developments we've proposed here. We describe each network and their limitations by turns below; briefly, we use eight of the relations captured by Knecht to model the marginal distribution of edges across many layers using the CMB parameterization described above and we use three of the relations captured by Lazega to demonstrate the utility of the layer-logic in modeling conditional effects that are traditionally sought after in the ERG framework.

Parameter estimation was accomplished using an experimental package `ergm.multi` extending the `ergm` package (Hunter et al. 2008) in the Statnet

suite of packages for R, which implements the above-mentioned MCMC-based estimation (Hunter and Handcock 2006). A benchmark study of the software is included in the supplementary materials. Where of interest, we use linear hypothesis tests to evaluate contrasts between coefficients representing similar social processes unfolding in different layers in a single model.[2] Model fit was evaluated by comparing observed values of graph statistics to those under simulation of the fitted model.

For convenience, we group the data description, model specification, and set of results separately by data set in the following subsections.

5.2. Knecht's Classroom Data

This data set was collected by Knecht (2008) and consists of a rich source of longitudinal, multi-layer, multi-site, network data on relationships between children in Dutch schools. For our purposes, we extract one of the largest classrooms from Wave 1, consisting of 32 students. The network data were collected in a classroom census via questionnaire, in which each student could nominate up to twelve of their classmates in response to each network question. We chose eight layers to employ in our empirical example of CMB: current classmates respondent has been friends with at primary school (basis layer), respondent's best-friends nominations (friend layer), classmates who provide emotional support to the respondent (emosup layer), classmates who provide practical support to the respondent (pracsup layer), classmates whom respondents would lend 25 Euro to (lend layer), classmates whom respondents would not lend 25 Euro to (notlen layer), classmates respondents dislike (dislik layer), and classmates respondents have been bullied by (bully layer). We also make use of several actor-level covariates available in the dataset that are *a priori* thought

---

[2]That is, to test a null hypothesis $\delta = 0$ for $\delta \equiv \theta_a - \theta_b$, write $\hat{\delta} = \hat{\theta}_a - \hat{\theta}_b$, with $\mathrm{var}(\hat{\delta}) = \mathrm{var}(\hat{\theta}_a) + \mathrm{var}(\hat{\theta}_b) - 2\,\mathrm{cov}(\hat{\theta}_a, \hat{\theta}_b)$, and then use a *z*-test with $z = \hat{\delta}/\sqrt{\mathrm{var}(\hat{\delta})}$.

to be important in the formation of one or more of the relations indexed in the multi-layer network. These include the gender of the student, how much pocket money the student has per month, the number of siblings the student has, and the extent to which the student self-reports having engaged in kissing, lying, and skipping class.

Figure 1 presents illustrations of the Knecht multi-layer network and the associations between layers. Vertices in the Panel A network projections are colored by gender and scaled by the amount of self-reported monthly allowance (a rough indicator of social class of the student's family). The edges are weighted in proportion to the count of layers in which vertices i and j have a tie present, respectively among prosocial and antisocial layers. There is apparently a stronger force of gender homophily for prosocial affiliation than there is for antisocial affiliation among these students, as the aggregated prosocial layers shown in Panel A exhibits clear separation between boys and girls while the antisocial aggregated layers exhibit relatively more apparent gender mixing. By visual inspection, there does not appear to be a strong effect of social class on edge formation in these networks. From Panel B, which plots the raw graph pairwise correlations, the anti-correlation is very low but positive correlations within pairs of layers corresponding to either prosocial or antisocial components are strong. This result is further illuminated in Panel C, the biplot of the first two principal components of the vectorized layers (Koehly and Marcum 2016), which demonstrates strong association within, but orthogonality between, components. Given these descriptive results, we would expect that a two-component CMB model may be well-suited to model the marginal association across layers.

5.2.1. Model Specification

As an exposition on how to model marginal and conditional association across layers, a series of baseline multi-layer ERGMs was fit to the Knecht
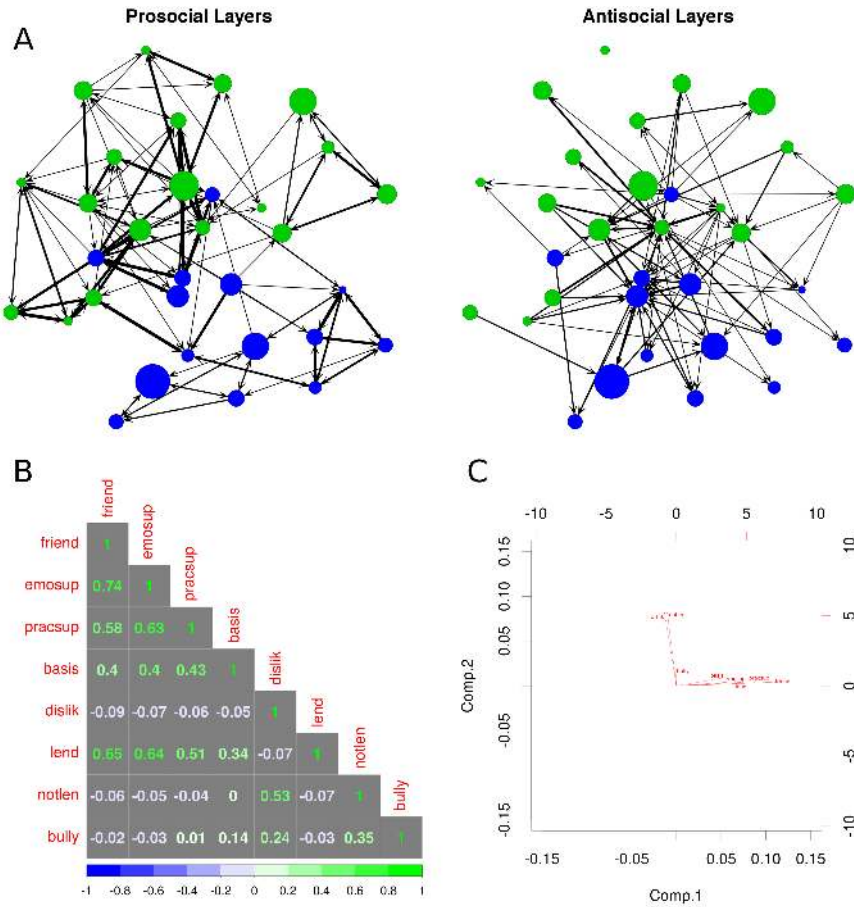
Figure 1: The Knecht multi-layer classroom. Panel A visualises two aggregate perspectives on the Knecht classroom multi-layer network. The left network represents the resulting network from the sum of the five prosocial layers (basis, friend, emosup, pracsup, and lend) with edge values scaled in proportion to the value of that network and the right network represents the same aggregation of the antisocial layers (dislik, notlen, and bully). Vertices are scaled in proportion to the amount of monthly allowance the student reports and are coloured by gender. The layout was determined via the Fruchterman-Reingold algorithm. Panels B and C illustrate two perspectives on the association between layers in this dataset; Panel B reports the raw Pearson product-moment pairwise correlations and Panel C is a biplot of the first two components from the principal component analysis of the vectorized layers

dataset. Specifically, we ask the question of how well can observed association among layers be recovered under various models employing CMB and layer logic model specifications. After these models were estimated, 1,000 realizations were simulated from each one and the simulated pairwise distribution of graph correlations (Pearson product-moment) was compared to the observed values. We specified four baseline models in this exercise: a pair of models (Models 1 and 3) with edge terms for pairwise conjunction of logical layers not resulting in sparse networks to model conditional association (one without covariates and one with covariates, respectively) and a pair of models (Models 2 and 4) with two CMB terms for the prosocial and antisocial component layers (again, one with and one without covariates, respectively). All models contain layer-specific edge terms and all covariate models assume homogeneity within prosocial and antisocial component layers, respectively.

5.2.2. Results

Table 2 reports terms, coefficients, standard errors, and p-values from the multi-layer ERGM fits of Knecht's classroom data. Controlling for differences in layer-specific densities, strong positive edge count terms for logical-layer conjunctions of most prosocial pairs of layers, on the one hand, and most antisocial pairs of layers on the other, are indicative of positive conditional association. Consistent with observing little negative association in figure 1, only a single pairwise conjunction had a significant negative association; that between baseline friendship and disliking (-2.68, $p = 0.03$). On average, the coefficients for both layer-specific edge count terms and logical-layer conjunction edge count terms appear to be attenuated slightly by the presence of terms for covariates in Model 3. Similarly, the two component CMB terms in Models 2 and 4 are both positive and significant (2.06, $p < 0.001$; 2.59, $p < 0.001$; and, 1.94, $p < 0.001$; 2.54, $p < 0.001$, respectively), indicating positive marginal association across

layers in both antisocial and prosocial components as expected. As the CMB terms model the distribution of edges across all layers in each component, there is some expected collinearity with the layer-specific edges terms which is evident in the relatively small edges terms estimates when compared to Models 1 and 3. Net of all else, only the positive homogeneous effect of gender homophily on prosocial layers and the negative homogeneous effect of having similarly reported "kissing" statuses on antisocial layers were significant in Models 3 and 4.

Notably, among the four models, Akaike's Information Criterion (AIC) selects Model 3 whereas the BIC—which tends to value parsimony more—selects Model 4; similarly, the more parsimonious CMB Models 2 and 4 are preferred over the pairwise conjunction Models 1 and 3 by BIC but not by AIC.

Figure 2 illustrates how well each baseline approach models the observed correlation structure among the layers by overlaying observed values in pairwise correlations (top panels) and underlying variance components (bottom panels) on the boxplots of 10,000 simulated values from each of those statistics under the respective models. Clearly, modeling conditional dependence using pairwise logical layer conjunctions in Model 1 better represents the respective observed pairwise product-moment correlations than the two-component marginal CMB in Model 2 does, on average. However, the two-component CMB approach in Model 2 appears to do at least as well replicating the underlying variance components from the observed principal components analysis as Model 1. These results highlight a trade-off from a modeling perspective: pairwise logical layer conjunctions perform better as a model of correlation between two layers in a multi-layer network, but with a greater number of parameters to estimate, and CMB performs at least as well as a model of underlying covariance among many layers with fewer parameters. Thus, when one wishes to model the association across many layers and the specific pairwise association between those layers is
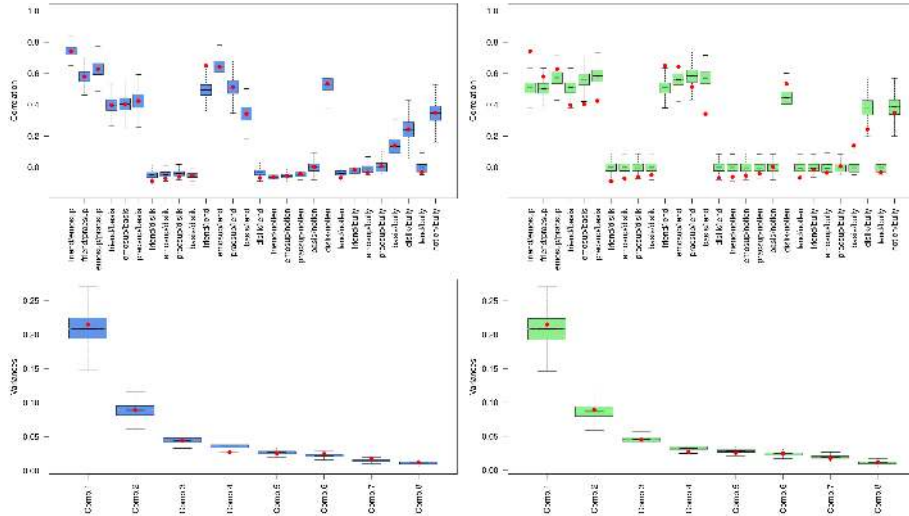
Figure 2: Simulations of correlation structure under Models 1 and 2 of the multi-layer ERGM fits to the Knecht classroom data. Boxplots in the top row of panels represent the distribution of pairwise product-moment correlations as simulated under Models 1 (in blue) and 2 (in green) for all pairs of layers (number of simulations = 10,000 multi-layer networks) with observed values indicated by red dots. Likewise, the bottom two panels represent the distribution of variance components from the principal components analysis of multi-layer networks as simulated under Models 1 and 2 also with observed values indicated by red dots.

unimportant, CMB is likely to be a better approach, particularly when parsimony is valued over complexity.

5.3. Lazega's Lawyers Data

This dataset comes from a network study of corporate law partnership that was carried out in a Northeastern US corporate law firm, referred to as SG&R, 1988-1991 in New England by Lazega (2001). Measurements of the networks among the 71 attorneys (partners and associates) of this firm include (among others) whether the ego has received alter's assistance in preparing documents (coworker layer), received alter's advice (advice layer), and whether the ego considers the alter a friend outside of work (friendship layer). Various lawyers' attributes are also part of the dataset including: age and seniority, formal social status, office in which they work, gender, law school attended, individual

27

Table 2: Marginal and conditional dependence ERGM model results for Knecht's classroom data

	Model 1	Model 2	Model 3	Model 4
Layer Specific Edge Effects				
edge count on layer friend	−3.339 (0.188)***	0.321 (0.152)*	−3.538 (0.211)***	−0.098 (0.200)
edge count on layer emosup	−5.969 (0.528)***	−0.639 (0.173)***	−6.143 (0.536)***	−1.061 (0.216)***
edge count on layer pracsup edge count on layer basis	−3.660 (0.220)***	−0.902 (0.181)***	−3.861 (0.241)***	−1.322 (0.224)***
edge count on layer lend	−3.930 (0.235)***	−0.862 (0.179)***	−4.143 (0.257)***	−1.282 (0.221)***
edge count on layer notlen	−3.610 (0.217)***	−0.548 (0.210)**	−3.312 (0.266)***	−0.309 (0.254)
edge count on layer dislik	−3.458 (0.196)***	−0.518 (0.210)*	−3.166 (0.249)***	−0.277 (0.254)
edge count on layer bully	−6.252 (0.697)***	−3.336 (0.339)***	−5.954 (0.716)***	−3.095 (0.370)***
Conditional Layer Dependence via Pairwise Conjunction (non-sparse layers)				
edge count on layer friend&emosup	5.080 (0.565)***		4.916 (0.567)***	
edge count on layer friend&pracsup	2.888 (0.589)***		2.753 (0.588)***	
edge count on layer emosup&pracsup	1.229 (0.589)*		1.157 (0.584)*	
edge count on layer friend&basis	1.842 (0.480)***		1.712 (0.481)***	
edge count on layer emosup&basis	0.270 (0.594)		0.198 (0.590)	
edge count on layer pracsup&basis	1.360 (0.465)**		1.296 (0.463)**	
edge count on layer basis&dislik	−2.685 (1.249)*		−2.742 (1.244)*	
edge count on layer emosup&lend	3.639 (0.418)***		3.453 (0.417)***	
edge count on layer pracsup&lend	1.128 (0.503)*		1.046 (0.498)*	
edge count on layer basis&lend	0.705 (0.493)		0.628 (0.487)	
edge count on layer friend&notlen	−0.676 (0.948)		−0.730 (0.945)	
edge count on layer emosup&notlen	−0.690 (1.341)		−0.737 (1.338)	
edge count on layer pracsup&notlen	−0.405 (1.271)		−0.445 (1.267)	
edge count on layer basis&notlen	0.373 (0.825)		0.353 (0.825)	
edge count on layer dislik&notlen	3.649 (0.346)***		3.590 (0.347)***	
edge count on layer friend&bully	−1.752 (1.610)		−1.783 (1.597)	
edge count on layer pracsup&bully	0.578 (1.636)		0.526 (1.624)	
edge count on layer basis&bully	3.623 (0.881)***		3.584 (0.881)***	
edge count on layer dislik&bully	1.530 (0.835)		1.479 (0.834)	
edge count on layer notlen&bully	3.423 (0.842)***		3.353 (0.845)***	
Marginal Layer Dependence via Two Component CMB Effects				
CMB over layers friend, emosup, pracsup, basis, and lend		2.057 (0.066)***		1.940 (0.071)***
CMB over layers notlen, dislik, and bully		2.593 (0.201)***		2.541 (0.204)***
Two-Component Homogeneous Covariate Effects				
matching sex summed over layers friend, emosup, pracsup, basis, and lend			0.717 (0.110)***	0.730 (0.116)***
absolute difference in age summed over layers friend, emosup, pracsup, basis, and lend			−0.098 (0.053)	−0.095 (0.053)
absolute difference in siblings summed over layers friend, emosup, pracsup, basis, and lend			−0.050 (0.033)	−0.046 (0.032)
absolute difference in money summed over layers friend, emosup, pracsup, basis, and lend			−0.004 (0.004)	−0.004 (0.004)
matching actkiss summed over layers friend, emosup, pracsup, basis, and lend			−0.079 (0.069)	−0.064 (0.068)
matching sex summed over layers notlen, dislik, and bully			0.131 (0.136)	0.080 (0.135)
absolute difference in age summed over layers notlen, dislik, and bully			−0.086 (0.099)	−0.077 (0.100)
absolute difference in siblings summed over layers notlen, dislik, and bully			−0.064 (0.070)	−0.057 (0.070)
absolute difference in money summed over layers notlen, dislik, and bully			−0.001 (0.007)	−0.001 (0.007)
matching actkiss summed over layers notlen, dislik, and bully			−0.477 (0.149)**	−0.472 (0.150)**
AIC	2596.964	2660.361	2526.594	2588.098
BIC	2792.381	2730.152	2791.803	2727.682
Log Likelihood	−1270.482	−1320.180	−1225.297	−1274.049

Notes: CMB=Conway Maxwell Binomial; AIC=Akaike Information Criterion; BIC=Bayesian Information Criterion
Signif. codes: *** < 0.001 <** < 0.01 <* < 0.05 <' < 0.1

performance measurements (hours worked, billable fees), and even attitudes concerning various management policy options. All layers in this network are directed and lack loops.

These data were previously analysed as a two-layer network using the ERGM framework by Lazega and Pattison (1999), which represents a foundational study that motivates our work here. Three different views of the multi-layer network are presented in figure 3.

5.3.1. Model Specification

We estimate four models of the Lazega data. The first (Model 1) represents a baseline that includes only within-layer structural terms and covariates. Specifically, we add within-layer structural effects for edge counts, reciprocated dyads, and geometrically weighted two-paths closed transitively with decay fixed at 0.25. We also control within each layer for various mixing and covariate effects based on actor-level attributes for lawyer gender, type of legal practice, office department, status within the firm (associate vs. partner), and years at the firm.

Model 2 builds upon the baseline by adding cross-layer and multi-layer structural terms. These include a complex cross-layer effect for geometrically weighted transitive two-paths (with a fixed decay parameter of 0.25) specified to capture the effect of seeking friendship ties from lawyers whom one's coworkers seek advice from: that is, the base relation in the corresponding diagram in table 1 would be friendship (the focal tie that transitively closes the two-path) and the out-bound two-path follows i to k along a coworker tie and k to j along an advice tie. While primarily included as a demonstration, one might argue that being a source of advice for one's coworkers is an attractive friendship quality, especially within a profession where advice is highly valued. Similarly, we include as a control a term for a cross-layer outbound two-path effect, likewise following i to k along a coworker tie and k to j along an advice tie. This effect is
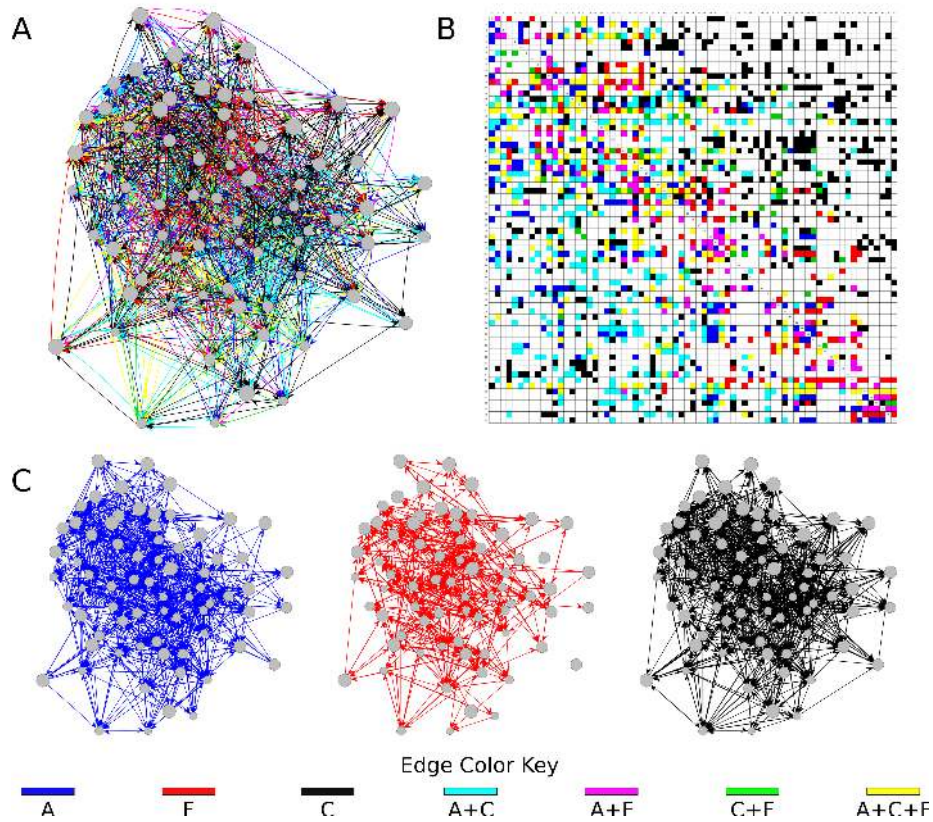
29

Figure 3: Three visualisations of the Lazega's Lawyers multi-layer network. Edges between vertices are coloured in Panel A to represent all seven possible combinations of advice (A), friendship (F), and coworker (C), relations (with the absent eighth combination representing no edge between two vertices i and j). A sociomatrix projection in Panel B illustrates an alternative visualisation of the same multi-layer network and demonstrates the heterogeneity across edges layer combinations. Panel C plots the three layers separately. Layouts for projections in Panels A and C were determined using the same coordinate output from the Fruchterman-Reingold force-directed algorithm and vertices were scaled in proportion to their relative seniority in the firm.

intended to capture the lower-order two-path component that is marginal to the cross-layer GWESP term. Also in Model 2, two terms modeling the multi-layer geometrically weighted in- and out- degree distributions (with decay parameters fixed at 0.40) are added, which all other things equal results in declining weight to ties among higher degree nodes. With Model 3, we add edge and reciprocated dyads in all pairwise logical conjunction layers, reciprocated ties in each pair of layers (ties sent in one layer and reciprocated in either of the other layers in the pair), and the number of instances of an individual receiving advice from one alter and helping another alter prepare documents (i.e., a cross-layer two-path). Finally, in Model 4 layer-specific cyclically closed two-paths within each layer are added.

5.3.2. Results

The results from the multi-layer ERGM fit of the Lazega lawyers data are reported in table 3. The best fitting model by BIC was the most complex Model 4 ($BIC = 9161.197$), which included all within-layer covariate terms, cross-layer terms, multi-layer terms, and additionally employed more complex within-layer GW terms. Due to the number of parameters (64), we are wary of the multiple testing problem, and therefore we only consider results with significance level of 0.01 and, ideally, even stronger. In the following, the coefficients reported are all from Model 4.

We summarise the within-layer covariate effects first, followed by the within-layer structural terms, and finally the multi- and cross-layer effects. Within-layer reciprocation of ties appears to be present in all layers (c/w: 2.46, $p < 0.001$; fr.: 2.55, $p < 0.001$; adv.: 0.66, $p = 0.004$), though the reciprocity appears to be weaker in the advice layer than in the other layers (adv. vs. c/w: $\hat{\delta} = -1.80$, $p < 0.001$; adv. vs. fr.: $\hat{\delta} = -1.90$, $p < 0.001$; fr. vs. c/w: $\hat{\delta} = 0.09$, $p = 0.7$).

We next consider the actor covariate effects within layers. Although there

appears to be differential gender covariate effects in some of the layers, some such effects appear to be moderated by multi-layer structural terms in the more complex models. For instance, while gender homophily appears to be weak or absent in the coworker layer across all models (via the non-significant gender matching coefficients), friendship (and to a lesser extent advice) layers have significant positive coefficients for gender matching in the first two models, which lose both magnitude and significance in Models 3 and 4 with the presence of conjunction logical layer and cross-layer reciprocation terms. Moreover, across all models, women are less likely to be the senders of coworker and friendship ties, more likely to be receivers of friendship ties. Lawyers co-located in the same office also tend to interact in all layers more with each other, as indicated by the respective significant positive matching effects.

There are differential covariate effects for legal practice in each layer. Litigators, for example, are less likely than their corporate counterparts to receive coworker ties (coefficients -0.31, $p = 0.002$), and more likely to receive (0.36, $p = 0.002$) but possibly less likely to send friendship ties (-0.31, $p = 0.007$). Evidence of legal practice homophily is present in all models for coworker and advice layers, but the matching practice coefficients in Models 1 and 2 switch signs from positive to negative for the friendship layer in Models 3 and 4 suggesting that friendship homophily between litigators depends on some structure involved in having joint ties across multiple layers.

The covariate effects of years of tenure in a firm are likewise mixed. Net of partner/associate status, the longer a lawyer has been at a firm, the less likely they are to send ties of all types (though the respective coefficients are non-significant in Models 3 and 4 in the friendship layer) and the more likely they are to receive friendship and advice ties from their colleagues. Recalling that coworker and advice layers concern *receiving* help and advice, this means

that more experienced lawyers tend to advise others and require less advice themselves. Increasing the absolute difference in years at the firm between two lawyers has no detectable effect on their likelihood of forming coworker ties, but significantly decreases their chances of being friends (log-odds-ratio per year difference, -0.05, $p < 0.001$). This suggests a cohort effect on friendship—lawyers who joined the firm around the same time are more likely to be friends—but not necessarily more likely to have professional interactions. (We have also considered the effects of age, but we found them universally non-significant net of other effects in the model.)

The effect of being a partner (as opposed to an associate) in a firm appears to increase sending and receiving coworker ties to and from colleagues. The significant positive receiver effects evident in Models 1 and 2, however, may be explained by the logical and cross-layer terms in Models 3 and 4, as they lose significance in those more complex models. There do not appear to be substantial effects of partnership status on the sending or receiving of ties in the friendship layer and, in the most complex Model 4, partners are more likely than associates to be both targets and reporters of advice ties (coefficients are (0.37, $p = 0.003$ and 0.32, $p = 0.02$). Lawyers display status heterophily in the coworker layer (-0.61, $p < 0.001$) and homophily in advice (0.78, $p < 0.001$). This confirms the finding of the original analysis of Lazega (2001): when a partner in a firm brings in a client, he or she recruits one or more associates (as opposed to other partners) to work together on the client's case, while occasionally seeking advice from other partners.

The within-layer triadic effects are consistently positive and highly significant for transitive configurations and negative and slightly weaker (both in effect size and statistical significance) for cyclical configurations. The exception being layer friendship, which has a weaker cyclical effect, particularly compared to advice

(contrasts fr. vs. c/w: $\hat{\delta} = 0.25$, $p = 0.1$; fr. vs. adv.: $\hat{\delta} = 0.44$, $p < 0.001$), suggesting that friendship relations are less hierarchical than the professional relations.

Lastly, we discuss the results of the cross- and multi-layer terms. The cross-layer GWESP term modeling the distribution of out-bound two-paths leading first along advice and then along coworking ties to be closed transitively by friendship ties has a significant negative coefficient in Models 2 through 4 (e.g. -0.39, $p < 0.001$ in Model 4). Net of everything else, it appears that others who are sought after for advice by one's coworkers are not, apparently, acquiring a quality desirous in forming one's friendship ties. This result suggests that under simulation, our model would lead to multi-layer networks with fewer number of triads involving cross-layer out-bound two-paths of this nature closed transitively by friendship. This does not appear to be the result of having a dearth of marginal two-paths however, as the coefficient for the distinct out-bound two-path starting along coworker and continuing to advice layers is significant and positive in Model 2 and has no detectable effect in Model 4 ($\hat{\delta} = 0.44$, $p < 0.001$). This suggests that, at least marginally, this model would produce networks with a greater number of out-bound two-paths that follow along advice to coworker ties that are either closed cyclically by friendship or not closed at all. Similarly, the multi-layer geometrically weighted out-degree distribution model term coefficient is negative and significant in Models 2 and 3 (coefficients of about -0.92, $p < 0.05$) and drops out in Model 4, while the corresponding positive coefficient for the multi-layer in-degree distribution is not significant in any model.

There is an interesting pattern involving the edge count and reciprocated dyads coefficients associated with the logical layer terms: in Models 3 and 4 the conjunction logical layer edge count coefficients are all positive, the corresponding

reciprocated dyads coefficients are negative, and the cross-layer reciprocation effects are positive. In Model 4 all of the conjunction logical layer effects are significant while only the coworker/advice cross-layer reciprocated dyads effect is significant. These results suggest that, net of all else in Model 4, ties sent from one lawyer to another in one of the layers are very likely to also be sent in the other layers but that reciprocation of those same ties in specific dyads is unlikely. The exception to this is a special case when ties sent in the coworker layer are reciprocated in specific dyads in a cross-layer fashion from the advice layer (or vice versa).

The goodness-of-fit for the Lazega example was assessed by simulating 10,000 draws from the fitted Model 4 and evaluating the extent to which observed sufficient statistics were covered by the distribution of those same statistics under simulation. Briefly, there is no evidence that the model is degenerate. Dyadwise and edgewise partner distributions are captured reasonably well, as are lower order component two-path effects. Finally, the model does a moderately good job capturing the shape and scale of the multi-layer degree distributions. A more detailed discussion can be found in the supplement.

6. Conclusion

This paper, introduces new approaches for the analysis of multi-layer networks using ERGMs. Previous work focused largely on modeling just a few layers of multi-relational networks; our contribution applies to an arbitrary number of layers. A new approach for modeling marginal dependence between a set of correlated Bernoulli variables using the Conway-Maxwell Binomial (CMB) distribution was adapted to the case of ERGMs for multi-layer networks. Additionally, both a *layer logic* and novel ERGM terms (including non-degenerate shared partners and degree distributions, as well as two-star and two-path terms

Table 3: Multi-layer ERGM model results for Lazega's lawyers data

	Model 1	Model 2	Model 3	Model 4
edge count on layer coworker	-5.373 (0.440)***	-5.492 (0.443)***	-5.040 (0.432)***	-5.128 (0.447)***
reciprocated dyads on layer coworker	2.554 (0.130)***	2.553 (0.130)***	2.391 (0.155)***	2.460 (0.159)***
receiver effect of having gender woman on layer coworker	0.141 (0.128)	0.168 (0.129)	0.112 (0.139)	0.088 (0.137)
sender effect of having gender woman on layer coworker	-0.574 (0.127)***	-0.548 (0.123)***	-0.523 (0.132)***	-0.507 (0.132)***
matching gender on layer coworker	0.029 (0.096)	0.034 (0.096)	-0.070 (0.110)	-0.072 (0.110)
matching office on layer coworker	1.118 (0.076)***	1.104 (0.076)***	0.664 (0.084)***	0.645 (0.085)***
receiver effect of having practice litigation on layer coworker	-0.290 (0.101)**	-0.286 (0.101)**	-0.333 (0.104)**	-0.310 (0.098)**
sender effect of having practice litigation on layer coworker	-0.064 (0.102)	-0.076 (0.100)	-0.017 (0.102)	-0.021 (0.095)
matching practice on layer coworker	1.027 (0.072)***	1.019 (0.073)***	0.676 (0.081)***	0.685 (0.081)***
receiver effect of years at firm on layer coworker	0.001 (0.008)	0.005 (0.008)	-0.013 (0.009)	-0.011 (0.008)
sender effect of years at firm on layer coworker	-0.028 (0.007)***	-0.028 (0.007)***	-0.023 (0.008)**	-0.021 (0.008)**
absolute difference in years at firm on layer coworker	-0.005 (0.005)	-0.005 (0.005)	0.012 (0.006)	0.012 (0.006)
GW two-path with decay 0.25 closed transitively on layer coworker	1.521 (0.317)***	1.509 (0.314)***	1.441 (0.302)***	1.736 (0.321)***
receiver effect of having status partner on layer coworker	0.452 (0.136)***	0.369 (0.140)**	0.223 (0.150)	0.232 (0.146)
sender effect of having status partner on layer coworker	0.549 (0.134)***	0.514 (0.133)***	0.380 (0.142)**	0.406 (0.140)**
matching status on layer coworker	-0.206 (0.081)*	-0.173 (0.081)*	-0.609 (0.100)***	-0.608 (0.099)***
edge count on layer friendship	-5.702 (0.225)***	-5.474 (0.259)***	-5.297 (0.259)***	-5.488 (0.265)***
reciprocated dyads on layer friendship	2.386 (0.175)***	2.396 (0.176)***	2.469 (0.217)***	2.553 (0.230)***
receiver effect of having gender woman on layer friendship	0.511 (0.135)***	0.489 (0.136)***	0.547 (0.142)***	0.540 (0.142)***
sender effect of having gender woman on layer friendship	-0.400 (0.134)**	-0.409 (0.129)**	-0.327 (0.134)*	-0.276 (0.138)*
matching gender on layer friendship	0.243 (0.091)**	0.236 (0.092)*	0.181 (0.100)	0.184 (0.100)
matching office on layer friendship	0.881 (0.093)***	0.947 (0.095)***	0.474 (0.100)***	0.417 (0.102)***
receiver effect of having practice litigation on layer friendship	0.298 (0.111)**	0.325 (0.113)**	0.397 (0.117)***	0.364 (0.117)**
sender effect of having practice litigation on layer friendship	-0.440 (0.111)***	-0.443 (0.106)***	-0.335 (0.112)**	-0.309 (0.115)**
matching practice on layer friendship	0.374 (0.075)***	0.429 (0.077)***	-0.164 (0.097)	-0.207 (0.096)*
receiver effect of years at firm on layer friendship	0.034 (0.009)***	0.034 (0.009)***	0.031 (0.010)**	0.030 (0.010)**
sender effect of years at firm on layer friendship	-0.026 (0.009)**	-0.024 (0.009)**	-0.006 (0.009)	-0.004 (0.010)
absolute difference in years at firm on layer friendship	-0.054 (0.007)***	-0.055 (0.007)***	-0.049 (0.008)***	-0.049 (0.008)***
GW two-path with decay 0.25 closed transitively on layer friendship	1.502 (0.153)***	1.478 (0.169)***	1.531 (0.167)***	1.630 (0.183)***
receiver effect of having status partner on layer friendship	0.154 (0.163)	0.283 (0.170)	-0.036 (0.178)	-0.130 (0.172)
sender effect of having status partner on layer friendship	0.309 (0.156)*	0.293 (0.153)	0.118 (0.160)	0.114 (0.159)
matching status on layer friendship	0.342 (0.085)***	0.330 (0.085)***	0.162 (0.097)	0.160 (0.098)
edge count on layer advice	-5.716 (0.277)***	-5.829 (0.298)***	-6.087 (0.316)***	-5.768 (0.317)***
reciprocated dyads on layer advice	0.757 (0.148)***	0.760 (0.148)***	0.202 (0.217)	0.657 (0.229)**
receiver effect of having gender woman on layer advice	-0.043 (0.118)	0.008 (0.113)	-0.004 (0.124)	-0.092 (0.118)
sender effect of having gender woman on layer advice	-0.151 (0.116)	-0.130 (0.115)	0.031 (0.133)	0.054 (0.124)
matching gender on layer advice	0.212 (0.107)*	0.219 (0.107)*	0.161 (0.132)	0.181 (0.133)
matching office on layer advice	1.367 (0.098)***	1.359 (0.097)***	0.663 (0.112)***	0.777 (0.116)***
receiver effect of having practice litigation on layer advice	0.061 (0.093)	0.066 (0.091)	0.139 (0.097)	0.011 (0.084)
sender effect of having practice litigation on layer advice	-0.345 (0.094)***	-0.340 (0.092)***	-0.224 (0.101)*	-0.110 (0.088)
matching practice on layer advice	1.211 (0.089)***	1.205 (0.089)***	0.777 (0.108)***	0.889 (0.113)***
receiver effect of years at firm on layer advice	0.042 (0.008)***	0.041 (0.007)***	0.043 (0.008)***	0.018 (0.007)**
sender effect of years at firm on layer advice	-0.056 (0.008)***	-0.054 (0.008)***	-0.042 (0.008)***	-0.022 (0.008)**
absolute difference in years at firm on layer advice	-0.044 (0.007)***	-0.043 (0.007)***	-0.030 (0.009)***	-0.019 (0.008)*
GW two-path with decay 0.25 closed transitively on layer advice	1.280 (0.179)***	1.277 (0.181)***	1.249 (0.180)***	1.399 (0.168)***
receiver effect of having status partner on layer advice	1.057 (0.140)***	0.998 (0.139)***	0.614 (0.149)***	0.367 (0.122)**
sender effect of having status partner on layer advice	0.217 (0.133)	0.133 (0.141)	-0.024 (0.154)	0.322 (0.135)*
matching status on layer advice	0.631 (0.096)***	0.617 (0.097)***	0.902 (0.124)***	0.781 (0.118)***
GW coworker-advice two-path in specified order with decay 0.25 closed transitively by friendship		-0.318 (0.092)***	-0.669 (0.100)***	-0.385 (0.102)***
GW indegree with decay 0.4 summed over all layers		-0.582 (0.447)	-0.452 (0.430)	-0.365 (0.397)
GW outdegree with decay 0.4 summed over all layers		-0.917 (0.396)*	-0.934 (0.365)*	-0.353 (0.352)
receiving coworker tie and sending advice tie to/from distinct actors		0.015 (0.007)*	0.002 (0.007)	0.008 (0.006)
edge count on layer coworker&friendship			0.582 (0.186)**	0.564 (0.185)**
reciprocated dyads on layer coworker&friendship			-0.585 (0.275)*	-0.572 (0.274)*
edge count on layer coworker&advice			2.199 (0.128)***	2.190 (0.127)***
reciprocated dyads on layer coworker&advice			-1.195 (0.221)***	-1.257 (0.220)***
edge count on layer friendship&advice			1.715 (0.159)***	1.719 (0.159)***
reciprocated dyads on layer friendship&advice			-0.435 (0.264)	-0.518 (0.264)*
edges in layer coworker reciprocated in layer friendship or vice versa			0.062 (0.181)	0.041 (0.181)
edges in layer coworker reciprocated in layer advice or vice versa			0.704 (0.136)***	0.707 (0.137)***
edges in layer friendship reciprocated in layer advice or vice versa			0.267 (0.166)	0.256 (0.167)
GW two-path with decay 0.25 closed cyclically on layer coworker				-0.352 (0.130)**
GW two-path with decay 0.25 closed cyclically on layer friendship				-0.106 (0.106)
GW two-path with decay 0.25 closed cyclically on layer advice				-0.542 (0.045)***
AIC	9680.616	9667.652	8783.239	8674.170
BIC	10045.886	10063.361	9247.437	9161.197
Log Likelihood	-4792.308	-4781.826	-4330.620	-4273.085

Notes: GW=Geometrically Weighted;
AIC=Akaike Information Criterion; BIC=Bayesian Information Criterion; Signif. codes: *** < 0.001 ** < 0.01 * < 0.05 . < 0.1

for multi-layer applications) were introduced, facilitating modeling conditional dependence across—or between—layers. Two empirical examples, that consisted of networks with eight and three layers, were presented to highlight different aspects of our advances. With these examples, the paper also demonstrated how fitting joint ERGMs to multiple layers—logical or observed—facilitates the use of within-model contrasts to test for differences in the magnitude of coefficients modeling the same social processes across layers.

The results of our empirical example using the Knecht classroom data in section 5.2.2 demonstrated that there exists a trade-off between using two-layer logical *intersection* layer specifications (which model well bivariate correlations with a greater number of parameters) and using the CMB specifications (which are better at modeling underlying dependence structure across many layers with fewer parameters). We also demonstrated that our approach can be useful in revealing new features of social structure and process vis-à-vis cross-layer ERGM effects. Specifically, using the Lazega lawyers data in section 5.3.2, we showed that triadic processes involving two-paths that cross two different layers in each segment may be important for suppressing ties in a third layer, controlling for other within- and cross-layer effects.

It is important to note that, as presented, these multi-layer ERGMs are best suited for theoretically informed, hypothesis driven analyses. The number of structural features that can be constructed from the layer logic and through the CMB increase exponentially with the number of layers comprising the network system. Thus, it is important for users of multi-layer ERGMs to be thoughtful about model specification, such that conceptually appropriate and hypothesized structural characteristics define the estimated models. For example, in some scenarios the CMB may represent a theoretically important structural feature of the multi-layer system. These might include the case when multiple raters

37

are informing on a single relation (Krackhardt 1987; Koehly and Pattison 2005) or multi-item construct networks (Koehly and Marcum 2016). However, in other scenarios, the layers may be distinct, with totally different meanings (e.g., positive and negative ties); in these scenarios, the CMB may not be meaningful, requiring the use of the layer logic to characterize hypothesized associations across layers.

More generally, while CMB effects behave well in our analyses, more study—both theoretical and simulation—is needed. For example, in Knecht Model 4, the prosocial CMB coefficient is significantly smaller than its antisocial counterpart ($\hat{\delta} = -0.60$, $p = 0.005$), but does that mean that there is greater positive association among the antisocial layers? In particular, the number of layers differ between the prosocial and antisocial sets, which may mean that they cannot be compared on an interval scale. And, as we note in section 2, the effects of heterogeneous layer densities on the CMB statistic require further study as well.

Innovations in multi-layer ERGMs that we've introduced here make a clear contribution to advance the field of statistical social network analysis. However, we would be remiss to not mention several limitations and opportunities for future work. First, while in principle our approach is appropriate for an arbitrary number of layers from a theoretical perspective, there are computational and practical limitations that may make estimation infeasible for very complex models with a large number of layers. The burden is on the researcher to ensure that their multi-layer model makes sense for the social processes and structures investigated by the research question at hand. Second, while we address how to apply the principle of marginality to cross-layer triadic effects, realizing carefully written models to that end remains a challenge in maximum likelihood estimation. In part, this is because two-path and two-star effects, on which many triadic closure processes are conditioned, often lead to ERGM degeneracy in general—cross-layer

specifications of such terms are apparently not spared from this result based on our supplemental analysis. An alternative to specifying such ideal models is to evaluate how well one's candidate model replicates features from omitted terms.

For example, in our cross-layer triadic closure model using the Lazega data, we simulated cross-layer two-paths from fitted models and compared the resulting distribution to corresponding two-path values of the observed network. This procedure helps to diagnose how well a higher-order cross-layer triadic term in a model recovers its lower order constituent two-path/two-star terms under simulation. Improvements in estimation for such degeneracy-inducing terms are needed for ERGMs in general and specifically for multi-layer ERGMs where the nascence of these models may inspire new innovations. Second, we've limited our treatment of this family of models to the cross-sectional multi-layer case where we've discussed each layer as representing different social relations. As both Koehly and Pattison (2005) and Snijders, Lomi, and Torló (2013) point out, however—by adopting strong theoretical and measurement approaches—the multi-layer ERGM methods can be adapted to multiple rater, or multiple time, assessments of a single relation. Finally, the models we choose to estimate and feature in this paper for our empirical examples may not be ideal from a social theoretic perspective and may suffer from omitted variable biases. Rather than strive for such theoretical niceties, we choose to use the empirical examples to highlight novel features for estimating multi-layer ERGMs that we've introduced here—we at least believe the models to be sensible, if not ideal. We leave ample room for new research questions to be explored using multi-layer ERGMs on these publicly available datasets. Thus, our contributions towards the development of ERGMs for multi-layer networks augments the toolkit from which network scientists may draw in their own research.

7. Acknowledgements

References

Barbillon, Pierre, Sophie Donnet, Emmanuel Lazega, and Avner Bar-Hen. 2017. "Stochastic Block Models for Multiplex Networks: An Application to a Multilevel Network of Researchers." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 180 (1): 295–314. doi:10.1111/rssa.12193.

Boorman, Scott A., and Harrison C. White. 1976. "Social Structure from Multiple Networks. Ii. Role Structures." *American Journal of Sociology* 81 (6): 1384–1446. doi:10.1086/226228.

Butts, Carter T. 2008. "A Relational Event Framework for Social Action." *Sociological Methodology* 38 (1): 155–200.

De Domenico, Manlio, Albert Solè-Ribalta, Emanuele Cozzo, Mikko Kivelà, Yamir Moreno, Mason A. Porter, Sergio Gòmez, and Alex Arenas. 2013. "Mathematical Formulation of Multilayer Networks." *Physical Review X* 3.4.

doi:10.1103/PhysRevX.3.041022.

Erdős, Paul, and Alfred Rényi. 1959. "On Random Graphs." *Publicationes Mathematicae Debrecen* 6: 290–97.

Fienberg, Stephen E., Michael M. Meyer, and Stanley S. Wasserman. 1980. "Analyzing Data from Multivariate Directed Graphs: An Application to Social Networks." In *Interpreting Multivariate Data*, edited by V Barnett. London: John Wiley.

Fienberg, Stephen, and Stanley S. Wasserman. 1981. "Categorical Data Analysis of Single Sociometric Relations." *Sociological Methodology* 12: 156–92.

Frank, Ove, and Termeh Shafie. 2016. "Multivariate Entropy Analysis of Network Data." *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 129 (1): 45–63. doi:10.1177/0759106315615511.

Frank, Ove, and David Strauss. 1986. "Markov Graphs." *Journal of the American Statistical Association* 81 (395): 832–42.

Handcock, Mark S. 2003. "Assessing Degeneracy in Statistical Models of Social Networks." Working Paper 39. Seattle, WA: Center for Statistics; the Social Sciences, University of Washington. http://www.csss.washington.edu/Papers/.

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

Holt-Lunstad, J., T.B. Smith, and J.B. Layton. 2010. "Social Relationships and Mortality Risk: A Meta-Analytic Review." *PloS Medicine* 7 (7): e1000316. doi:10.1371/journal.pmed.1000316.

Huitsing, Gijs, Marijtje A.J. Van Duijn, Tom A.B. Snijders, Peng Wang, Miia Sainio, Christina Salmivalli, and René Veenstra. 2012. "Univariate and Multivariate Models of Positive and Negative Networks: Liking, Dis-

liking, and Bully–Victim Relationships." *Social Networks* 34 (4): 645–57. doi:10.1016/j.socnet.2012.08.001.

Hunter, David R. 2007. "Curved Exponential Family Models for Social Networks." *Social Networks* 29: 216–30. doi:10.1016/j.socnet.2006.08.005.

Hunter, David R., and Mark S. Handcock. 2006. "Inference in Curved Exponential Family Models for Networks." *Journal of Computational and Graphical Statistics* 15 (3): 565–83. doi:10.1198/106186006X133069.

Hunter, David R., Mark S. Handcock, Carter T. Butts, Steven M. Goodreau, and Martina Morris. 2008. "Ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks." *Journal of Statistical Software* 24 (3): 1–29. http://www.jstatsoft.org/v24/i03.

Jeub, Lucas G. S., Michael W. Mahoney, Peter J. Mucha, and Mason A. Porter. 2017. "A Local Perspective on Community Structure in Multilayer Networks." *Network Science* 5 (2): 144–63. doi:10.1017/nws.2016.22.

Kadane, Joseph B. 2016. "Sums of Possibly Associated Bernoulli Variables: The Conway–Maxwell-Binomial Distribution." *Bayesian Analysis* 11 (2): 403–20. doi:10.1214/15-BA955.

Kapferer, Bruce. 1972. *Strategy and Transaction in an African Factory: African Workers and Indian Management in a Zambian Town.* Manchester University Press.

Knecht, Andrea B. 2008. *Friendship Selection and Friends' Influence. Dynamics of Networks and Actor Attributes in Early Adolescence.* Utrecht University.

Koehly, Laura M., and Christopher Steven Marcum. 2016. "Multi-Relational Measurement for Latent Construct Networks." *Psychological Methods* 21 (4): 1–35.

Koehly, Laura M., and Phillippa Pattison. 2005. "Random Graph Models for Social Networks: Multiple Relations or Multiple Raters." In *Models and*

Methods in Social Network Analysis, edited by PJ Carrington, J Scott, and Stanley Wasserman, 162–91. Cambridge University Press Cambridge, UK. doi:10.1017/CBO9780511811395.009.

Krackardt, David. 1987. "QAP Partialling as a Test of Spuriousness." *Social Networks* 9 (2): 171–86. doi:10.1016/0022-2496(82)90008-6.

Krackhardt, David. 1987. "Cognitive Social Structures." *Social Networks* 9 (2). Elsevier: 109–34.

Krivitsky, Pavel N. 2012. "Exponential-Family Random Graph Models for Valued Networks." *Electronic Journal of Statistics* 6: 1100–1128. doi:10.1214/12-EJS696.

———. 2017. "Using Contrastive Divergence to Seed Monte Carlo Mle for Exponential-Family Random Graph Models." *Computational Statistics & Data Analysis* 107: 149–61.

Lazega, Emmanuel. 2001. *The Collegial Phenomenon: The Social Mechanisms of Cooperation Among Peers in a Corporate Law Partnership.* New York: Oxford University Press.

Lazega, Emmanuel, and Philippa E. Pattison. 1999. "Multiplexity, Generalized Exchange and Cooperation in Organizations: A Case Study." *Social Networks* 21 (1): 67–90.

Lusher, Dean, and Garry Robins. 2013. "Example Exponential Random Graph Models." In *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*, edited by Dean Lusher, Johan Koskinen, and Garry Robins, 37–46. Cambridge: Cambridge University Press.

Magnani, Matteo, and Stanley S. Wasserman. 2017. "Introduction to the Special Issue on Multilayer Networks." *Network Science* 5 (2): 141–43. doi:10.1017/nws.2017.15.

Nowicki, Krzysztof, and Tom A.B. Snijders. 2001. "Estimation and Prediction

for Stochastic Blockstructures." *Journal of the American Statistical Association* 96 (455): 1077–87.

Pattison, Philippa E. 1982. "The Analysis of Semigroups of Multirelational Systems." *Journal of Mathematical Psychology* 25 (2): 87–118. doi:10.1016/0022-2496(82)90008-6.

Robins, Garry, and Philippa E. Pattison. 2001. "Random Graph Models for Temporal Processes in Social Networks." *Journal of Mathematical Sociology* 25: 5–41.

Salter-Townshend, Michale, and Tyler H. McCormick. 2017. "Latent Space Models for Multiview Network Data." *Annals of Applied Statistics* 11 (3): 1217–44. doi:10.1016/j.socnet.2017.01.002.

Schweinberger, Michael. 2011. "Instability, Sensitivity, and Degeneracy of Discrete Exponential Families." *Journal of the American Statistical Association* 106 (496): 1361–70. doi:10.1198/jasa.2011.tm10747.

Schweinberger, Michael, Pavel N. Krivitsky, Carter T. Butts, and Jonathan Stewart. 2020. "Exponential-Family Models of Random Graphs: Inference in Finite-, Super-, and Infinite Population Scenarios." *Statistical Science* To appear (March). http://arxiv.org/abs/1707.04800v4.

Shafie, Termeh. 2015. "A Multigraph Approach to Social Network Analysis." *Journal of Social Structure* 16 (1): 1–21. https://www.cmu.edu/joss/content/articles/volume16/Shafie.pdf.

Snijders, Tom A.B. 2002. "Markov chain Monte Carlo Estimation of Exponential Random Graph Models." *Journal of Social Structure* 3 (2).

Snijders, Tom A.B., Alessandro Lomi, and Vanina Jasmine Torló. 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): 265–76. doi:10.1016/j.socnet.2012.05.005.

Snijders, Tom A.B., Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock. 2006. "New Specifications for Exponential Random Graph Models." *Sociological Methodology* 36 (1): 99–153.

Stewart, Jonathan, Michael Schweinberger, Michal Bojanowski, and Martina Morris. 2019. "Multilevel Network Data Facilitate Statistical Inference for Curved ERGMs with Geometrically Weighted Terms." *Social Networks* 59: 98–119. doi:10.1016/j.socnet.2018.11.003.

Strauss, David, and Michael Ikeda. 1990. "Pseudolikelihood Estimation for Social Networks." *Journal of the American Statistical Association* 85 (409): 204–12.

Voros, Andras, and Tom A.B. Snijders. 2017. "Cluster Analysis of Multiplex Networks: Defining Composite Network Measures." *Social Networks* 49: 93–112.

Wang, Peng. 2012. "Exponential Random Graph Model Extensions: Models for Multiple Networks and Bipartite Networks." In *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*, edited by Dean Lusher, Johan Koskinen, and Garry Robins, 115–29. Structural Analysis in the Social Sciences. Cambridge University Press. doi:10.1017/CBO9780511894701.012.

Wasserman, Stanley S. 1987. "Conformity of Two Sociometric Relations." *Psychometrika* 52 (1): 3–18. doi:10.1007/BF02293953.

Wasserman, Stanley S., and Philippa E. Pattison. 1996. "Logit Models and Logistic Regressions for Social Networks: I. an Introduction to Markov Graphs and P*." *Psychometrika* 61 (3): 401–25.

Wasserman, Stanley S., Katherine Faust, and Joseph Galaskiewicz. 1990. "Correspondence and Canonical Analysis of Relational Data." *Journal of Mathematical Sociology* 15 (1): 11–64. doi:10.1080/0022250X.1990.9990060.

White, Douglas R. 1996. "Statistical Entailments and the Galois Lattice."

Social Networks 18 (3): 201–15.

White, Douglas R., and Karl P. Reitz. 1983. "Graph and Semigroup Homomorphisms on Networks of Relations." doi:10.1016/0378-8733(83)90025-4.

White, Harrison C., Scott A. Boorman, and Ronald L. Breiger. 1976. "Social Structure from Multiple Networks. I. Blockmodels of Roles and Positions." *American Journal of Sociology* 81 (4): 730–80. doi:10.1086/226141.