

# Scalable Signed Exponential Random Graph Models under Local Dependence

Marc Schalberger\*, Freie Universität Berlin

Cornelius Fritz, Trinity College Dublin

## Abstract

Traditional network analysis focuses on binary edges, while real-world relationships are more nuanced, encompassing cooperation, neutrality, and conflict. The rise of negative edges in social media discussions spurred interest in analyzing signed interactions, especially in polarized debates. However, the vast data generated by digital networks presents challenges for traditional methods like Stochastic Block Models (SBM) and Exponential Family Random Graph Models (ERGM), particularly due to the homogeneity assumption and global dependence, which become increasingly unrealistic as network size grows. To address this, we propose a novel method that combines the strengths of SBM and ERGM while mitigating their weaknesses by incorporating local dependence based on non-overlapping blocks. Our approach involves a two-step process: first, decomposing the network into sub-networks using SBM approximation, and then estimating parameters using ERGM methods. We validate our method on large synthetic networks and apply it to a signed Wikipedia network of thousands of editors. Through the use of local dependence, we find patterns consistent with structural balance theory.

*Keywords:* Exponential Random Graph Models, Signed Networks, Local Dependence, Large-Scale Networks

---

\*Corresponding author: m.schalberger@fu-berlin.de. Postal address: Freie Universität Berlin, Garystr. 21, Berlin 14195, Germany

# 1 Introduction

Network analysis methods are traditionally constrained to binary edges, where relationships are either positive (e.g., alliances, friendships) or negative (e.g., rivalries, conflicts). However, these two types of relationships can occur simultaneously in many real-world scenarios. For instance, Fritz et al. (2025) examined interstate conflict and cooperation between 2000 and 2010. Similarly, Leskovec et al. (2010) studied in the setting of social media networks how interaction patterns are affected by both positive and negative connections between users. Historically speaking, signed networks were the basis of one of the first network theories, being theories where the dependence between nodes is embedded into the theory (Wasserman and Faust, 1994): Heider (1946) proposed the structural balance theory giving rise to the aphorism that “*the enemy of my enemy is my friend*”. The underlying theory divides triads into balanced and unbalanced. Balanced triads are considered to be more stable and will, therefore, last longer than unbalanced triads.

However, the analysis of signed networks was hindered by limited computational resources to estimate more complicated models and by expensive measurements that make their observation possible. This situation has changed dramatically with the rise of social media platforms and the availability of digital trace data (Lazer et al., 2009). Online interactions, especially in polarized settings, often contain positive and negative edges. Incorporating negative relations provides a richer understanding of coalitions, conflicts, and divisions within social systems (Leskovec et al., 2010).

Nevertheless, the modeling of signed networks remains challenging. On the one hand, the class of models specifically designed for signed networks is relatively limited, since most models were developed for unsigned networks. On the other hand, the massive size of modern networks poses separate challenges, as even for unsigned networks many models

struggle to scale computationally. At the intersection of these two issues, the need to model both signed relations and large networks, there are little to no methods available.

Two of the most prominent models for network analysis are Stochastic Block Models (SBM, Nowicki and Snijders, 2001) and Exponential Family Random Graph Models (ERGM, Lusher et al., 2013). SBMs are a latent variable model in which all dependence between edges is accounted for by latent group membership. ERGMs, on the other hand, allow for more complex local dependencies through so-called sufficient statistics (De Nicola et al., 2023). Signed network extensions exist for both SBM and ERGM (Jiang, 2015; Fritz et al., 2025) with inherent shortcomings: While the SBM can represent cohesive subgroups among the nodes, it implies conditional independence for forming edges between nodes. This assumption is unrealistic for networks exhibiting more dependence among edges. ERGMs, on the other hand, offer a flexible framework to analyze general network properties. For a parsimonious model, these properties are commonly assumed to be homogeneous across the entire network. The induced global dependence structure, where each edge may technically depend on every other edge in the same way, gives rise to undesirable behavior specifically for large networks (Schweinberger, 2011; Handcock, 2003). Moreover, the estimation process, when based on Markov Chain Monte Carlo (MCMC) methods, often fails due to the complexity and size of the data. Larger networks may imply degenerate models with poor sampling properties and longer mixing times (Bhamidi et al., 2011).

In Section 3, we fill the resulting methodological gap by combining the strengths of both models while simultaneously addressing their respective weaknesses, introducing local dependence based on non-overlapping blocks. Our model assumes that each node is only aware of the activities within its block. Consequently, the formation of within-block edges can be characterized by a more complex ERGM, while between-block edges are not affected

by endogenous effects (i.e., internal and structural influences within a network) that rely on knowledge of neighboring areas. For example, in a social network, individuals may form friendships within their immediate circle without being influenced by the entire network. Snijders (2007) considered these models that combine latent space and exponential random graph models as ‘the next generation of social network models’, capable of capturing both cohesive structures and subgraph patterns in larger networks.

To enable large-scale estimation of our novel model extension, we rely on a two-step estimation approach proposed by Babkin et al. (2020), which we detail in Section 4. In the first step, the network is decomposed into sub-networks by approximating the likelihood of the model with an SBM for signed networks. This step is carried out with the help of a variational approximation and computationally fast MM updates (Vu et al., 2013). We build upon the efficient algorithms introduced by Dahbura et al. (2021) by extending them to handle signed networks. As outlined in Section 4.4, we incorporate uncertainty quantification into this step to assess the reliability of the estimated block structure. In the second step, the parameters are estimated given the previously estimated block structure with known methods for ERGMs. Thereby, we enable the estimation of signed ERGMs under local dependencies for large signed networks encompassing thousands of nodes. Additionally, the local dependence assumption speeds up simulation, as the between-block edges are characterized by a less complex model, and enables parallelization, as blocks are conditionally independent given the block membership. To demonstrate the computational advantage of this method, we apply it to both synthetic and real data with up to 5,000 nodes. In Section 6, we consider a large signed network of independent Wikipedia editors (Lerner and Lomi, 2019), where users interact either by undoing (negative edge) or redoing (positive edge) each other’s contributions (Lerner and Lomi, 2017). This network is par-

ticularly suitable for a model that assumes local dependency because editors are likely to have knowledge limited to their focus areas, and through our analysis we have successfully identified these focus areas.

Our contributions are both methodological and computational:

1. We introduce a novel model of signed ERGMs with local dependence, enabling the analysis of large signed networks where dyadic dependence is confined to within-block structures.
2. We extend the efficient computing algorithms proposed in Dahbura et al. (2021) to signed networks.
3. We account for the uncertainty of the inferred block structure in a principled manner.
4. We develop the open-source R package **bigsergm**, which implements our full estimation procedure and is publicly available in the following GitHub repository [mschalberger/bigsergm](https://github.com/mschalberger/bigsergm).
5. We demonstrate the method’s practical value by analyzing a large Wikipedia editor network. We find evidence consistent with predictions of structural balance theory in a large-scale, real-world setting.

Throughout this paper, we consider a signed network with  $N \in \mathbb{N}$  nodes. The adjacency matrix for this network is denoted as  $\mathbf{y} = (y_{i,j}) \in \mathcal{S}^{N \times N} := \mathcal{Y}$ , with  $\mathcal{S} := \{“-”, “0”, “+”\}$ , where  $\mathcal{Y}$  is the set of all observable signed adjacency matrices among  $N$  fixed nodes. Therefore,  $y_{i,j} = “-”$  indicates a negative edge,  $y_{i,j} = “+”$  represents a positive edge, and  $y_{i,j} = “0”$  signifies the absence of an edge between nodes  $i$  and  $j$ . In the undirected case, we have  $y_{i,j} = y_{j,i}$ , implying that the adjacency matrix is symmetric. For this paper, we assume that the network is undirected and has no self-loops, so  $y_{i,i} = “0”$ . However,

extensions to the directed case follow naturally. By  $\mathbf{y}_{(-ij)}$ , we denote  $\mathbf{y}$  excluding  $y_{i,j}$ . The set of nodes is partitioned into  $K \in \mathbb{N}$  disjoint blocks. Let  $N_k \in \mathbb{N}$  denote the number of nodes in block  $k$ , for  $k = 1, \dots, K$ , such that  $\sum_{k=1}^K N_k = N$ . The unobservable vector  $\mathbf{z}_i = (z_{i1}, \dots, z_{iK})$  denotes the block membership of each node, where  $z_{i,k} = 1$  if node  $i$  belongs to block  $k$ , and  $z_{i,k} = 0$  otherwise. Let  $\mathbf{y}_{k,l} \in \mathcal{S}^{N_k \times N_l}$  denote the submatrix of  $\mathbf{y}$  that contains all edges between nodes in block  $k$  and nodes in block  $l$ , for  $k, l = 1, \dots, K$ . Throughout this paper, capital letters refer to random variables, lowercase letters to their realizations, and bold letters denote vectors or matrices.

## 2 Existing Models for Signed Networks

Statistical models provide a principled framework for analyzing signed networks, with the most prominent examples being ERGMs and SBMs.

**Signed Exponential Random Graph Model.** The Signed Exponential Random Graph Model (SERGM, Fritz et al., 2025) posits that an observed network can be fully characterized through a set of sufficient statistics, which capture both endogenous dependencies, which are structural features of the network, and exogenous influences, relating to information such as covariate effects independent of the network (Lusher et al., 2013). Fritz et al. (2025) defined the probability of observing  $\mathbf{y} \in \mathcal{Y}$  is given by

$$\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y}) = \frac{\exp(\boldsymbol{\theta}^\top \mathbf{s}(\mathbf{y}))}{\kappa(\boldsymbol{\theta})}, \quad (1)$$

where  $\mathbf{s}(\mathbf{y})$  denotes a vector of sufficient statistics, given by a function  $\mathbf{s} : \mathcal{Y} \rightarrow \mathbb{R}^p$ , weighted by the estimated coefficients  $\boldsymbol{\theta} \in \mathbb{R}^p$  and the normalizing constant  $\kappa(\boldsymbol{\theta}) := \sum_{\tilde{\mathbf{y}} \in \mathcal{Y}} \exp(\boldsymbol{\theta}^\top \mathbf{s}(\tilde{\mathbf{y}}))$  guarantees that (1) sums up to 1. When facing large networks, this framework faces several challenges: First, evaluating the normalization constant  $\kappa(\boldsymbol{\theta})$  is

intractable for almost all networks. Statistical inference generally relies on a MCMC approximations of the likelihood function, which often fails due to the complexity and size of the large networks. Second, the sufficient statistics induce global dependence by being calculated over the entire network. The inclusion of these statistics, therefore, implies knowledge about the entire network, which is unrealistic for networks with more than a few thousand nodes.

**Signed Stochastic Block Model.** One approach to restrict the nodes' knowledge locally is to cluster into blocks. A seemingly simple model along these lines assumes that nodes exhibit structural equivalence within their respective block (Fienberg and Wasserman, 1981). Assuming that the block assignments are latent random variables that we learn from observed networks leads to the Stochastic Block Model (SBM, Snijders and Nowicki, 1997), where the probability of any observed network  $\mathbf{y} \in \mathcal{Y}$  given a latent clustering  $\mathbf{Z}$  is:

$$\mathbb{P}_{\boldsymbol{\pi}}(\mathbf{Y} = \mathbf{y} \mid \mathbf{Z} = \mathbf{z}) = \sum_{i < j} \mathbb{P}_{\boldsymbol{\pi}}(Y_{i,j} = y_{i,j} \mid Z_i = k, Z_j = l),$$

where the sum ranges over all unordered pairs of distinct nodes  $(i, j)$  in the network. The conditional probability to observe  $Y_{i,j} = y_{i,j}$  is assumed to be the probability mass function of a Bernoulli-distributed random variable:

$$\mathbb{P}_{\boldsymbol{\pi}}(Y_{i,j} = y_{i,j} \mid Z_i = k, Z_j = l) = \pi_{k,l}^{y_{i,j}} (1 - \pi_{k,l})^{1-y_{i,j}}, \quad (2)$$

where the block-specific connection probability between blocks  $k$  and  $l$  is given by  $\pi_{k,l} \in [0, 1]$ .

By substituting the conditional Bernoulli distribution in (2) with a multinomial distribution, we can accommodate for signed edges  $y_{i,j} \in \mathcal{S}$ :

$$\mathbb{P}_{\boldsymbol{\pi}}(Y_{i,j} = y_{i,j} \mid Z_i = k, Z_j = l) = \prod_{y \in \mathcal{S}} \pi_{k,l}(y)^{\mathbb{I}(y_{i,j}=y)}$$

where  $\pi_{k,l}(y)$  with  $y \in \mathcal{S}$  denotes the probability of  $Y_{i,j} = y$  with  $\sum_{y \in \mathcal{Y}_{i,j}} \pi_{k,l}(y) = 1$  and  $\boldsymbol{\pi} = (\pi_{k,l}(y))$  (Li et al., 2023). A shortcoming of this model is its simplicity: The probability to observe a negative, neutral, or positive edge between nodes  $i$  and  $j$  solely depends on the block memberships of the involved nodes. This dependence structure is not able to capture foundational theories about signed networks, such as the structural balance theory.

### 3 Exponential Random Graph Model for Signed Networks under Local Dependence

Since the assumption of global dependence among edges is generally infeasible for large networks, both from a modeling and interpretational perspective, it seems more plausible that edge dependencies are local, meaning that they depend only on a subset of other edges. This is because dependence implies awareness or knowledge of the activities of other nodes. It is therefore reasonable to assume that this knowledge exists only about nodes within the same block, and not across the entire network. This structure is captured through a latent block membership  $\mathbf{Z}$ , which groups nodes into unobserved communities that govern local interaction patterns. Given this block structure  $\mathbf{Z}$ , the probability of observing a network  $\mathbf{y}$  can be decomposed into two components: one that captures potentially complex dependencies within each block, and another that describes interactions between blocks that are independent of each other. This decomposition is attractive not only because it reflects how many real-world networks are organized, but also because it reduces computational complexity. By limiting complex dependencies to within-block interactions, and by allowing parallel computation, since blocks are conditionally independent given  $\mathbf{Z}$ .

The overall probability of observing the network  $\mathbf{y}$ , given the block structure  $\mathbf{z}$ , is



the product of the probabilities for the within-block and between-block sub-networks, the former running over all individual blocks, the latter over all unordered pairs of distinct blocks. This leads to the following expression for the network  $\mathbf{y} \in \mathcal{Y}$ :

$$\begin{aligned} \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y} \mid \mathbf{Z} = \mathbf{z}) &= \left( \prod_{k=1}^K \mathbb{P}_{\boldsymbol{\theta}_{k,k}}(\mathbf{Y}_{k,k} = \mathbf{y}_{k,k} \mid \mathbf{Z} = \mathbf{z}) \right) \\ &\times \left( \prod_{k < l} \mathbb{P}_{\boldsymbol{\theta}_{k,l}}(\mathbf{Y}_{k,l} = \mathbf{y}_{k,l} \mid \mathbf{Z} = \mathbf{z}) \right), \end{aligned} \quad (3)$$

where  $\boldsymbol{\theta} := \text{vec}(\boldsymbol{\theta}_{1,1}, \dots, \boldsymbol{\theta}_{K,K}, \boldsymbol{\theta}_{1,2}, \dots, \boldsymbol{\theta}_{K-1,K})$  denotes the vector of both between-block and within-block parameters with  $\text{vec}$  defining a function that stacks its arguments vertically. Unlike with stochastic block models, this model does not assume that edges within and between blocks are independent. Rather, only the independence of between block edges is guaranteed, while within block edges can be strongly dependent.

For nodes within the same block  $k$ , the probability of observing the sub-network  $\mathbf{y}_{k,k}$  is given by a complex signed ERGM:

$$\mathbb{P}_{\boldsymbol{\theta}_{k,k}}(\mathbf{Y}_{k,k} = \mathbf{y}_{k,k} \mid \mathbf{Z} = \mathbf{z}) = \frac{\exp(\boldsymbol{\theta}_{k,k}^\top \mathbf{s}(\mathbf{y}_{k,k}))}{\kappa(\boldsymbol{\theta}_{k,k})}, \quad (4)$$

where  $\boldsymbol{\theta}_{k,k} \in \mathbb{R}^p$  denotes the vector of within-block parameters for block  $k$ , and  $\mathbf{s} : \mathcal{Y}_{k,k} \rightarrow \mathbb{R}^p$  is a function calculating the sufficient statistics. This function includes both dyad-independent and dyad-dependent variables, since we assume that nodes are aware of the activities within their block. The number of sufficient statistics used to model within-block networks  $p_w \in \mathbb{N}$  is assumed to be constant across blocks.

The between-block edges are characterized by a signed SBM, where the probability of observing the sub-network  $\mathbf{y}_{k,l}$ , conditional on the block assignments  $\mathbf{z}$ , is given by:

$$\mathbb{P}_{\boldsymbol{\theta}_{k,l}}(\mathbf{Y}_{k,l} = \mathbf{y}_{k,l} \mid \mathbf{Z} = \mathbf{z}) = \frac{\exp(\boldsymbol{\theta}_{k,l}^\top \mathbf{h}(\mathbf{y}_{k,l}))}{\kappa(\boldsymbol{\theta}_{k,l})},$$

with  $\boldsymbol{\theta}_{k,l} \in \mathbb{R}^q$  denoting the between-block parameters. The sufficient statistics  $\mathbf{h} : \mathcal{Y}_{k,l} \rightarrow \mathbb{R}^q$  (with  $l \neq k$ ) include only dyad-independent variables, since dyad-dependency would

require knowledge of the activity in other blocks:

$$\mathbf{h}_{k,l}(\mathbf{y}_{k,l}) = \sum_{i,j} z_{i,k} z_{j,l} (\mathbb{I}(y_{i,j} = "-") \mathbf{x}_{i,j,-} + \mathbb{I}(y_{i,j} = "+") \mathbf{x}_{i,j,+}),$$

where  $\mathbf{x}_{i,j,+}$  and  $\mathbf{x}_{i,j,-}$  are covariate vectors associated with positive and negative edges, respectively. Because only dyad-independent terms are included, the normalizing constant becomes tractable and factorizes over dyads:

$$\kappa_{k,l}(\boldsymbol{\theta}_{k,l}) = \prod_{i,j: z_{i,k} z_{j,l} = 1} (1 + \exp(\mathbf{x}_{i,j,+}^\top \boldsymbol{\theta}_{k,l,+}) + \exp(\mathbf{x}_{i,j,-}^\top \boldsymbol{\theta}_{k,l,-})).$$

In practice, we are interested in parsimonious models to ease interpretation, hence letting  $\boldsymbol{\theta}_{k,l}$  vary freely over  $k, l = 1, \dots, K$  is unreasonable. Extending Slaughter and Koehly (2016) and Krivitsky et al. (2023), we represent different homogeneity assumptions on these coefficients by linear combinations of population-level coefficients for within and between blocks,  $\boldsymbol{\beta}_w \in \mathbb{R}^{s \times p}$  and  $\boldsymbol{\beta}_b \in \mathbb{R}^{t \times q}$ , and block-specific covariates,  $\mathbf{v}_k \in \mathbb{R}^s$  and  $\mathbf{u}_{k,l} \in \mathbb{R}^t$ :

$$\boldsymbol{\theta}_{k,k} := (\mathbf{v}_k^\top \boldsymbol{\beta}_w)^\top \text{ and } \boldsymbol{\theta}_{k,l} := (\mathbf{u}_{k,l}^\top \boldsymbol{\beta}_b)^\top. \quad (5)$$

The model would constitute a curved ERGM with local dependence if the relationship between the population-level coefficients and the block-specific covariates in equation (5) were non-linear (Hunter and Handcock, 2006). Block-specific covariates  $\mathbf{v}_k$ , such as block size  $N_k$  or  $\log(N_k)$ , further enable size-dependent parametrizations (see, e.g., Butts and Almquist, 2015 and Krivitsky et al., 2011). Larger blocks will typically have a lower density, however using non-size dependent parameters would lead to preservation of the density. Let  $\beta_{w,a,b}$ ,  $\theta_{k,k,a}$ , and  $v_{k,b}$  denote the  $(a,b)$ th entry of  $\boldsymbol{\beta}_w$ , the within-block effect of the  $a$ th statistic in  $\boldsymbol{\theta}_{k,k}$ , and value of the  $b$ th entry of  $\mathbf{v}_i$ , respectively. Then  $\theta_{k,k,b}$  is given by  $\beta_{w,a,b} \mathbf{v}_k$  and  $\beta_{w,a,b}$  can be interpreted as the increase of  $\theta_{k,k,b}$  when  $v_{k,b}$  is increase by one unit. Covariates  $\mathbf{u}_{k,l}$  typically only include a constant, dummy variables for specific pairs of blocks, or functions of  $\mathbf{v}_k$  and  $\mathbf{v}_l$ .

**Example 1: Signed SBM.** To start edges  $y_{i,j}$  are assumed to be conditionally independent given the block memberships in Section 2. Let  $\mathbb{I}(y_{i,j} = y) = a_{i,j,y}$  be an indicator for observing edge value  $y \in \{“+”, “-”\}$  between node  $i$  and  $j$ . The vector of sufficient statistics for the within-block model  $\mathbf{s}(\mathbf{y}_{k,k}) \in \mathbb{R}^2$  and the between-block model  $\mathbf{h}(\mathbf{y}_{k,l}) \in \mathbb{R}^2$  then include the count of positive and negative edges for each block or block pair, respectively. Define the edge terms as:

$$\text{Edges}^+(\mathbf{y}) := \sum_{i < j} a_{i,j,+}, \quad \text{Edges}^-(\mathbf{y}) := \sum_{i < j} a_{i,j,-}.$$

Then the vectors of sufficient statistics is given by:

$$\mathbf{s}(\mathbf{y}_{k,k}) = \mathbf{h}(\mathbf{y}_{k,l}) = (\text{Edges}^+(\cdot), \text{Edges}^-(\cdot))^\top.$$

These statistics correspond to parameters  $\boldsymbol{\theta}_{k,l} = (\theta_{k,l,+}, \theta_{k,l,-})$ , where we define  $\theta_{k,l,0} = 0$ .

The multinomial probabilities coming up in (2) are then recovered as follows:

$$\pi_{k,l}(y) = \frac{\exp(\theta_{k,l,y})}{\sum_{y^* \in \mathcal{S}} \exp(\theta_{k,l,y^*})}.$$

Thus,  $\theta_{k,l,+}$  and  $\theta_{k,l,-}$  represent log-odds of observing a positive and negative edge between blocks  $k$  and  $l$  compared to observing no edge. In this form, each  $\mathbf{s}(\mathbf{y}_{k,k})$  and  $\mathbf{h}(\mathbf{y}_{k,l})$  acts like an intercept term: it aggregates the number of edges of a particular sign, and its corresponding parameter controls the baseline probability of that edge type appearing between nodes with the given block memberships. For the within-block model, we can also represent this in the linear predictor structure from Equation (5). In this case each  $\mathbf{v}_k \in \mathbb{R}^K$  is a one-hot vector with a 1 in the  $k$ th position and 0 elsewhere. Similarly, for the between-block model,  $\mathbf{u}_{k,l} \in \mathbb{R}^K$  is a vector with 1s in position  $k$  and  $l$ . To reduce model complexity and avoid estimating block-pair-specific parameters, we assume homogeneous parameters across all between-block pairs. In this case, we define  $\mathbf{u}_{k,l} = \mathbf{1}$ , so all  $\boldsymbol{\theta}_{k,l} = \boldsymbol{\beta}_b \in \mathbb{R}^{1 \times 2}$ . In

the following examples, this parameter vector remains fixed, as no dyad-independent terms are added to the between-block model.

**Example 2: Signed Model with Triadic Terms.** To relax the assumption of the SSBM that edges within and between blocks are independent, we can incorporate triadic terms that reflect structural balance theory. Specifically, the model may include triadic configurations reflecting the principles “*friend of my friend is my friend*” and “*enemy of my enemy is my friend*” (Fritz et al., 2025). However, the inclusion of triadic terms often leads to model degeneracy (Schweinberger, 2011). Schweinberger and Stewart (2020) showed that defining triadic statistics by:

$$\begin{aligned}\text{Triad}^{+++}(\mathbf{y}_{k,k}) &:= \sum_{i < j} a_{i,j,+} \mathbb{I} \left( \sum_{h \neq i,j} a_{i,h,+} a_{h,j,+} > 0 \right), \\ \text{Triad}^{+--}(\mathbf{y}_{k,k}) &:= \sum_{i < j} a_{i,j,+} \mathbb{I} \left( \sum_{h \neq i,j} a_{i,h,-} a_{h,j,-} > 0 \right)\end{aligned}$$

yields better-behaved distributions. The vector of sufficient statistics for the within-block model is given by:

$$\mathbf{s}(\mathbf{y}_{k,k}) = \left( \text{Edges}^+(\mathbf{y}_{k,k}), \text{Edges}^-(\mathbf{y}_{k,k}), \text{Triad}^{+++}(\mathbf{y}_{k,k}), \text{Triad}^{+--}(\mathbf{y}_{k,k}) \right)^\top,$$

where the triadic terms count the number of positive edges that are connected through at least one mutual friend, in the case of  $\text{Triad}^{+++}(\mathbf{y}_{k,k})$ , or at least one mutual enemy, in the case of  $\text{Triad}^{+--}(\mathbf{y}_{k,k})$ .

**Example 3: Signed Model with Structural Terms.** To capture more complex local structures beyond count variables such as edge and triad counts, we include geometrically weighted positive/negative degree and edgewise shared partner statistics, specifically the geometrically weighted “*enemy of my enemy is my friend*” triad. These geometrically weighted statistics reduce the risk of model degeneracy by reducing the weight of high degree nodes or high shared partner counts, thereby improving model stability. The decay

parameter  $\omega$  controls this discounting: smaller values of  $\omega$  place more weight on higher-degree nodes or triads with many shared partners, while larger values of  $\omega$  reduce their influence. For a more detailed overview of geometrically weighted statistics, see Hunter (2007).

Define the geometrically weighted positive degree statistic as:

$$\text{GWD}^y(\mathbf{y}_{k,k}, \omega) := \exp(\omega) \sum_{d=1}^{N_k-1} (1 - (1 - \exp(-\omega))^d) \deg_{k,d}^y,$$

where

$$\deg_{k,d}^y := \sum_{i=1}^{N_k} \mathbb{I} \left( \sum_{j \neq i} a_{i,j,y} = d \right)$$

counts nodes in block  $k$  with positive or negative degree exactly  $d$ .

Define the geometrically weighted edgewise shared partner statistic as:

$$\begin{aligned} \text{GWESE}^y(\mathbf{y}_{k,k}, \omega) &:= \exp(\omega) \sum_{d=1}^{N_k-2} (1 - (1 - \exp(-\omega))^d) \text{ESE}_d^y, \\ \text{GWESF}^y(\mathbf{y}_{k,k}, \omega) &:= \exp(\omega) \sum_{d=1}^{N_k-2} (1 - (1 - \exp(-\omega))^d) \text{ESF}_d^y, \end{aligned}$$

where

$$\text{ESE}_d^y := \sum_{i < j} a_{i,j,y} \mathbb{I} \left( \sum_{h \neq i,j} a_{i,h,-} a_{j,h,-} = d \right), \quad \text{ESF}_d^y := \sum_{i < j} a_{i,j,y} \mathbb{I} \left( \sum_{h \neq i,j} a_{i,h,+} a_{j,h,+} = d \right).$$

Here,  $\text{ESE}_d^y$  counts edges of sign  $y$  sharing exactly  $d$  common enemies, and  $\text{ESF}_d^y$  counts edges of sign  $y$  sharing exactly  $d$  common friends.

The vector of sufficient statistics for the within-block model is then defined as:

$$\begin{aligned} \mathbf{s}(\mathbf{y}_{k,k}) &= (\text{Edges}^+(\mathbf{y}_{k,k}), \text{Edges}^-(\mathbf{y}_{k,k}), \text{GWD}^+(\mathbf{y}_{k,k}, \omega), \\ &\quad \text{GWD}^-(\mathbf{y}_{k,k}, \omega), \text{GWESE}^+(\mathbf{y}_{k,k}, \omega))^\top. \end{aligned}$$

## 4 Scalable Estimation

The block allocation  $\mathbf{Z}$ , a multinomial variable taking values in  $\{1, \dots, K\}$ , is often unobserved and must be learned from the observed network  $\mathbf{y}$ . Assuming that the number

of blocks  $K$  is fixed, we model  $\mathbf{Z}$  under a multinomial prior with relative group-sizes  $\gamma_1, \dots, \gamma_K$ :

$$Z_i \stackrel{\text{iid}}{\sim} \text{Multinomial}(1; \gamma_1, \dots, \gamma_K), \quad (6)$$

following the practice for SSBMs (Li et al., 2023). Combining (6) with the conditional model  $\mathbf{Y} \mid \mathbf{Z} = \mathbf{z}$  (defined in (3)) yields a latent variable model (see, David et al., 2011 for an introduction to the field). The log-likelihood for parameters  $\boldsymbol{\theta}$  and  $\boldsymbol{\gamma}$  is

$$\ell(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \log(\mathbb{P}_{\boldsymbol{\theta}, \boldsymbol{\gamma}}(\mathbf{Y} = \mathbf{y})) = \log\left(\sum_{\mathbf{z} \in \mathcal{Z}} \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y} \mid \mathbf{Z} = \mathbf{z}) \mathbb{P}_{\boldsymbol{\gamma}}(\mathbf{Z} = \mathbf{z})\right) \quad (7)$$

where  $\mathcal{Z} := \{1, \dots, K\}^N$  is the discrete space of block assignments.

Plugging (3) in to  $\ell(\boldsymbol{\theta}, \boldsymbol{\gamma})$ , shows that direct maximization of (7) would involve the evaluation of nested intractable sums: the normalization constants of the within networks (4) and the discrete space of block assignments, both of which grow superexponentially with  $N$ .

To address this, we introduce an auxiliary distribution  $A$  over  $\mathcal{Z}$  and apply Jensen's inequality:

$$\ell(\boldsymbol{\theta}, \boldsymbol{\gamma}) \geq \mathbb{E}_A(\log(\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y} \mid \mathbf{Z} = \mathbf{z}) \mathbb{P}_{\boldsymbol{\gamma}}(\mathbf{Z} = \mathbf{z}))) - \mathbb{E}_A(\log A(\mathbf{Z})) =: \text{LB}(A, \boldsymbol{\theta}, \boldsymbol{\gamma}). \quad (8)$$

Here,  $\text{LB}(A, \boldsymbol{\theta}, \boldsymbol{\gamma})$  is known as the evidence lower bound, which is a functional of  $A$  and the model parameters. The expectation with respect to  $A$  over a function  $f$  is  $\mathbb{E}_A(f(\mathbf{Z})) := \sum_{\mathbf{z} \in \mathcal{Z}} A(\mathbf{z})f(\mathbf{z})$ . We optimize  $\text{LB}(A, \boldsymbol{\theta}, \boldsymbol{\gamma})$  with respect to  $A, \boldsymbol{\theta}$ , and  $\boldsymbol{\gamma}$  by alternating block coordinate ascent:

**Step 1:** Update  $A$  to minimize  $\text{LB}(A, \boldsymbol{\theta}, \boldsymbol{\gamma})$  for fixed  $\boldsymbol{\theta}$  and  $\boldsymbol{\gamma}$ ;

**Step 2:** Update  $\boldsymbol{\theta}$  and  $\boldsymbol{\gamma}$  to minimize  $\text{LB}(A, \boldsymbol{\theta}, \boldsymbol{\gamma})$  for fixed  $A$ .

If  $\mathbf{z}$  is observed or can be reasonably derived from covariate information, only Step 2 is needed.

The optimal solution in Step 1 is  $A(\mathbf{Z}) = \mathbb{P}(\mathbf{Z} = \mathbf{z} \mid \mathbf{Y} = \mathbf{y})$ , which does not admit a closed-form expression (Matias and Robin, 2014). For SBMs of moderate size, Gibbs sampling based on the full conditional distributions  $Z_i \mid Z_1, \dots, Z_{i-1}, Z_{i+1}, \dots, Z_N$  can be employed to approximate  $\mathbb{P}(\mathbf{Z} = \mathbf{z} \mid \mathbf{Y} = \mathbf{y})$  (Nowicki and Snijders, 2001). However, for larger networks, as considered in this paper, variational methods offer scalable approximations. For this approach, we restrict  $A$  to a tractable family of distributions parametrized by  $\boldsymbol{\alpha}$  and approximate  $\mathbb{P}(\mathbf{Z} = \mathbf{z} \mid \mathbf{Y} = \mathbf{y}) \approx A_{\boldsymbol{\alpha}}(\mathbf{z})$ , where  $A_{\boldsymbol{\alpha}}(\mathbf{z})$  minimizes the Kullback-Leibler divergence to  $\mathbb{P}(\mathbf{Z} = \mathbf{z} \mid \mathbf{Y} = \mathbf{y})$ . An explicit form of this variational family is provided in Section 4.2.

Still, the dependence of within-block networks (4) inhibits adaption to our setting. In Section 4.1, we adapt the result of Babkin et al. (2020) to signed networks showing that under some conditions the probability (3) can be approximated by the probability of a SSBM, mentioned in Section 2 and Example 1. This result justifies the use scalable optimization methods originally developed for SBMs. More concretely, in Section 4.2 we adapt the approach of Vu et al. (2013) based on MM Steps to the signed setting. After convergence, a point estimate  $\hat{\mathbf{z}}$  is derived Section 4.3 discusses the estimation of  $\boldsymbol{\theta}$  conditional on  $\mathbf{Z} = \hat{\mathbf{z}}$ . Section 4.4 proposes methods to incorporate uncertainty in  $\mathbf{Z}$  neglected by this deterministic assignment.

## 4.1 Signed Stochastic Block Model Approximation

For any model specification, we recover the Signed Stochastic Block Model (SSBM), as shown in Example 1, by setting all parameters in model (4) corresponding to dyad-dependent

statistics (e.g., transitivity) to zero. Note that the sufficient statistics in (5) are already dyad-independent by definition. Therefore, the approximation of the probability of our model by the probability of a SSBM only accrues errors for within-block networks. As  $N$  and  $K$  grow, our joint probability will then be increasingly dominated by terms relating to between-block probabilities, provided single blocks are not too large. This offers a heuristic justification for why the SSBM provides a good approximation in large populations.

Accordingly, we decompose the within-block parameters of the  $k$ th block  $\boldsymbol{\theta}_{k,k} = \text{vec}(\boldsymbol{\theta}_{k,k}^{\setminus}, \boldsymbol{\theta}_{k,k}^{\perp})$ , where  $\boldsymbol{\theta}_{k,k}^{\setminus}$  relates to statistics inducing dependence and  $\boldsymbol{\theta}_{k,k}^{\perp}$  to statistics implying independence. Setting  $\boldsymbol{\theta}_{k,k}^{\setminus} = \mathbf{0}$ , we define  $\boldsymbol{\theta}_{k,k}^{\text{SBM}} = \text{vec}(\mathbf{0}, \boldsymbol{\theta}_{k,k}^{\perp})$  for all  $k \in \{1, \dots, K\}$  and  $\boldsymbol{\theta}^{\text{SBM}} := \text{vec}(\boldsymbol{\theta}_{1,1}^{\text{SBM}}, \dots, \boldsymbol{\theta}_{K,K}^{\text{SBM}}, \boldsymbol{\theta}_{1,2}, \dots, \boldsymbol{\theta}_{K-1,K})$ . With this notation, substituting  $\boldsymbol{\theta}$  by  $\boldsymbol{\theta}^{\text{SBM}}$  in (3), defines the corresponding nested SSBM. Denote by  $m(\mathbf{z})$  the size of the largest block in block allocation  $(\mathbf{z}) \in \mathcal{Z}$  and let  $d : \mathcal{Y} \rightarrow \mathcal{Y}$  be the Hamming distance between two signed networks  $\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{Y}$ :

$$d(\mathbf{y}_1, \mathbf{y}_2) = \sum_{i < j} \mathbb{I}(\mathbf{y}_{1,i,j} \neq \mathbf{y}_{2,i,j}),$$

where  $\mathbb{I}$  denotes the indicator function. We require the following conditions:

**Condition 1: Smoothness of  $\mathbf{s}$ .** A constant  $c > 0$  exists such that for all  $\boldsymbol{\theta} \in \boldsymbol{\Theta}$  and all  $\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{Y}$ ,

$$|\langle \boldsymbol{\theta}, \mathbf{s}(\mathbf{y}_1) - \mathbf{s}(\mathbf{y}_2) \rangle| \leq c d(\mathbf{y}_1, \mathbf{y}_2) m(\mathbf{z}) \log N.$$

**Condition 2: Smoothness of  $\boldsymbol{\theta}$ .** A constant  $c > 0$  exists such that for all  $\boldsymbol{\theta}_{k,l;1}, \boldsymbol{\theta}_{k,l;2} \in \boldsymbol{\Theta}_{k,l}$  and all  $\mathbf{z} \in \mathcal{Z}$ ,

$$|\langle \boldsymbol{\theta}_{k,l;1} - \boldsymbol{\theta}_{k,l;2}, \mathbb{E}(\mathbf{s}_{k,l}(\mathbf{Y})) \rangle| \leq c \|\boldsymbol{\theta}_{k,l;1} - \boldsymbol{\theta}_{k,l;2}\| m(\mathbf{z})^2 \log N.$$



All sufficient statistics considered in this paper satisfy Assumptions 1 and 2. Following Theorem 2 from Babkin et al. (2020), the full conditional probability  $\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y} \mid \mathbf{Z} = \mathbf{z})$  can be then approximated with  $\boldsymbol{\theta}_{SBM}$  as follows

$$\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y} \mid \mathbf{Z} = \mathbf{z}) \approx \mathbb{P}_{\boldsymbol{\theta}^{SBM}}(\mathbf{Y} = \mathbf{y} \mid \mathbf{Z} = \mathbf{z}). \quad (9)$$

## 4.2 Variational Approximation based on MM Updates

Supported by this result, we substitute  $\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} = \mathbf{y} \mid \mathbf{Z} = \mathbf{z})$  in (8) by the probability of a SSBM, which simplifies the optimization in Steps 1 and 2 outlined above. We assume a mean-field variational family for the class of discrete distributions characterized, implying that the joint distribution of  $\mathbf{Z}$  factorizes across nodes:

$$A_{\boldsymbol{\alpha}}(\mathbf{z}) = \prod_{i=1}^N A_{\boldsymbol{\alpha}_i, i}(\mathbf{z}_i),$$

where  $A_{\boldsymbol{\alpha}_i, i}$  denotes the density of a categorical distribution parametrized by  $\boldsymbol{\alpha}_i = (\alpha_{i,k}) \in [0, 1]^K$  with  $\sum_{k=1}^K \alpha_{i,k} = 1$ . The variational parameters that characterize the class of distribution approximating  $\mathbb{P}(\mathbf{Z} = \mathbf{z} \mid \mathbf{Y} = \mathbf{y})$  are  $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N) \in [0, 1]^{K \times N}$ .

The lower bound  $\text{LB}(A, \boldsymbol{\theta}, \boldsymbol{\gamma})$  defined in (8) is then a function of  $\boldsymbol{\alpha}$ :

$$\text{LB}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\gamma}) = \sum_{i < j} \sum_{k=1}^K \sum_{l=1}^K \alpha_{i,k} \alpha_{j,l} \log p_{k,l}(y_{i,j}) + \sum_{i=1}^N \sum_{k=1}^K \alpha_{i,k} (\log \gamma_k - \log \alpha_{i,k}),$$

where  $p_{k,l}(y) = \mathbb{P}_{\boldsymbol{\theta}^{SBM}}(Y_{i,j} = y \mid Z_i = k, Z_j = l)$  denotes the density evaluated at  $y$  of an edge between nodes  $i$  and  $j$ , with actor  $i$  in block  $k$  and actor  $j$  in block  $l$  under the SSBM approximation described in Section 4.1. The dependence of  $p_{k,l}(y)$  on the parameter  $\boldsymbol{\theta}^{SBM}$  is omitted for brevity.

Although one can directly maximize the lower bound  $\text{LB}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\gamma})$  with respect to  $\boldsymbol{\alpha}$  using iterative fixed-point methods (see, e.g., Daudin et al., 2008), a more scalable and robust

approach employs Minorization-Maximization (MM) steps (Vu et al., 2013). Specifically, we introduce the surrogate function  $Q(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\theta}^{(t)}; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\alpha})$ :

$$Q(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\theta}^{(t)}; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\alpha}) = \sum_{i=1}^N \sum_{k=1}^K A_{i,k}(\mathbf{y}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\theta}^{(t)}) \alpha_{i,k}^2 + B_{i,k}(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\alpha}^{(t)}) \alpha_{i,k} \quad (10)$$

which is quadratic in  $\boldsymbol{\alpha}$  and hence easier to maximize. The quadratic term is defined as

$$A_{i,k}(\mathbf{y}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\theta}^{(t)}) := \frac{\Omega_{i,k}^{(t)}(\mathbf{y}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\theta}^{(t)})}{\alpha_{i,k}^{(t)}} - \frac{1}{\alpha_{i,k}^{(t)}},$$

and the linear term as

$$B_{i,k}(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\alpha}^{(t)}) := \log \gamma_k^{(t)} - \log \alpha_{i,k}^{(t)} + 1,$$

where

$$\Omega_{i,k}^{(t)}(\mathbf{y}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\theta}^{(t)}) := \sum_{j \neq i}^N \sum_{l=1}^K \alpha_{j,l}^{(t)} \log p_{k,l}^{(t)}(y_{i,j}).$$

This surrogate function minorizes  $\text{LB}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\gamma})$  in  $\boldsymbol{\alpha}^{(t)}$  and has thus the following two properties:

$$Q(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\theta}^{(t)}, \boldsymbol{\alpha}^{(t)}; \boldsymbol{\alpha}) \leq \text{LB}(\boldsymbol{\alpha}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\theta}^{(t)}) \quad \text{for all } \boldsymbol{\alpha} \in [0, 1]^{K \times N}$$

$$Q(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\theta}^{(t)}, \boldsymbol{\alpha}^{(t)}; \boldsymbol{\alpha}^{(t)}) = \text{LB}(\boldsymbol{\alpha}^{(t)}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\theta}^{(t)}).$$

Consequently, increasing  $Q(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\theta}^{(t)}; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\alpha})$  with respect to  $\boldsymbol{\alpha}$  guarantees an increase in  $\text{LB}(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\gamma})$ . Since  $Q$  is a quadratic function of  $\boldsymbol{\alpha}$ , scalable and robust optimization methods are available to efficiently perform Step 1 (see Stefanov, 2004). A complete derivation of Equation (10) and details on expressing all terms via matrix multiplication for computational efficiency are provided in Supplement B.1 and B.2, respectively. Conditional on  $\boldsymbol{\alpha}^{(t)}$ , the updates of  $\boldsymbol{\theta}^{\text{SBM}}$  and  $\boldsymbol{\gamma}$  are available in closed forms in Step 2, which are provided in the Supplement B.3.

### 4.3 Parameter Step with known Blocks

From the converged lock membership probabilities  $\hat{\boldsymbol{\alpha}}$ , Babkin et al. (2020) assign each node deterministically to the block with the highest posterior probability to obtain the

block membership vector  $\mathbf{z}$

$$\hat{\mathbf{z}} = \left( \arg \max_{k \in \{1, \dots, K\}} \alpha_{i,k} \right)_{i=1}^N.$$

Given this block membership vector, we can estimate the model parameters  $\boldsymbol{\theta}$ . Since (5) expresses  $\boldsymbol{\theta}$  as a function of the population-level parameters  $\boldsymbol{\beta} = (\boldsymbol{\beta}_w, \boldsymbol{\beta}_b)$ , we optimize over these lower-dimensional parameters rather than the full set of block- or block-pair-specific coefficients  $\boldsymbol{\theta}_{k,k}$  and  $\boldsymbol{\theta}_{k,l}$ . Accordingly, we reparametrize (3) using the coefficients  $\boldsymbol{\beta}_{w,\text{vec}} := \text{vec}(\boldsymbol{\beta}_w)$  and  $\boldsymbol{\beta}_{b,\text{vec}} := \text{vec}(\boldsymbol{\beta}_b)$ , together with the block-specific sufficient statistics

$$\mathbf{s}_k : \mathbf{y}_{k,k} \mapsto \mathbf{v}_k \otimes \mathbf{s}(\mathbf{y}_{k,k}), \quad \mathbf{h}_{k,l} : \mathbf{y}_{k,l} \mapsto \mathbf{u}_{k,l} \otimes \mathbf{h}(\mathbf{y}_{k,l}). \quad (11)$$

The term  $\mathbf{a} \otimes \mathbf{b}$  denotes the Kronecker product of vectors  $\mathbf{a} \in \mathbb{R}^m$  and  $\mathbf{b} \in \mathbb{R}^n$ . In the following, we focus on estimating the within-block model parameter  $\boldsymbol{\beta}_{w,\text{vec}}$ . The method applies analogously to  $\boldsymbol{\beta}_{b,\text{vec}}$ . While both MCMC-based maximum likelihood (Hunter and Handcock, 2006) and maximum pseudo-likelihood methods (MPLE, Strauss and Ikeda, 1990) can be used to estimate the parameters in the second step, MPLE is generally better suited for large networks due to its computational efficiency and scalability. If dyadic independence holds, both approaches coincide.

The conditional probability of an edge between nodes  $i$  and  $j$  with  $z_i = z_j = k$  to be  $y \in \mathcal{S}$  given the rest of the network  $\mathbf{Y}_{(-ij)}$  and  $\boldsymbol{\beta}_{w,\text{vec}}$  is

$$\mathbb{P}_{\boldsymbol{\beta}_{w,\text{vec}}}(Y_{i,j} = y \mid \mathbf{Y}_{(-ij)} = \mathbf{y}_{(-ij)}, z_i = z_j = k) = \frac{\exp(\boldsymbol{\beta}_{w,\text{vec}}^\top \mathbf{s}_k(\mathbf{y}_{i,j}^y))}{\sum_{y^*} \exp(\boldsymbol{\beta}_{w,\text{vec}}^\top \mathbf{s}_k(\mathbf{y}_{i,j}^{y^*}))}.$$

Here,  $\mathbf{s}_k$  from (11) is evaluated at the network configuration  $\mathbf{y}_{i,j}^y$ , where the value of edge  $y_{i,j}$  is set to  $y$ , while all other edge values remain fixed as observed.

With  $\Delta_{i,j,k}^{0 \rightarrow +}(\mathbf{y})$  and  $\Delta_{i,j,k}^{0 \rightarrow -}(\mathbf{y})$  being the change in the reparametrized sufficient statistics given in (11) after changing the edge value  $y_{i,j}$  from “0” to “+” or “−”, respectively, we can express the relative log odds of observing  $\mathbf{Y}_{i,j}$  to be “+” abs “−” instead of “0”,

conditional on  $z_i = z_j = k$ :

$$\begin{aligned}\log \left( \frac{\mathbb{P}_{\boldsymbol{\beta}_{w,\text{vec}}}(Y_{i,j} = \text{"+"} \mid \mathbf{Y}_{(-ij)} = \mathbf{y}_{(-ij)}, z_i = z_j = k)}{\mathbb{P}_{\boldsymbol{\beta}_{w,\text{vec}}}(Y_{i,j} = \text{"0"} \mid \mathbf{Y}_{(-ij)} = \mathbf{y}_{(-ij)}, z_i = z_j = k)} \right) &= \boldsymbol{\beta}_{w,\text{vec}}^\top \Delta_{i,j,k}^{0 \rightarrow +}(\mathbf{y}) \\ \log \left( \frac{\mathbb{P}_{\boldsymbol{\beta}_{w,\text{vec}}}(Y_{i,j} = \text{"-"} \mid \mathbf{Y}_{(-ij)} = \mathbf{y}_{(-ij)}, z_i = z_j = k)}{\mathbb{P}_{\boldsymbol{\beta}_{w,\text{vec}}}(Y_{i,j} = \text{"0"} \mid \mathbf{Y}_{(-ij)} = \mathbf{y}_{(-ij)}, \mathbf{z})} \right) &= \boldsymbol{\beta}_{w,\text{vec}}^\top \Delta_{i,j,k}^{0 \rightarrow -}(\mathbf{y}).\end{aligned}$$

The log-likelihood function is then defined as

$$\ell(\boldsymbol{\beta}_{w,\text{vec}}) = \sum_{i < j} \left( \boldsymbol{\beta}_{w,\text{vec}}^\top \Delta_{i,j,k}^{0 \rightarrow y_{i,j}}(\mathbf{y}) - \log \sum_y \exp(\boldsymbol{\beta}_{w,\text{vec}}^\top \Delta_{i,j,k}^{0 \rightarrow y}(\mathbf{y})) \right). \quad (12)$$

Newton Raphson algorithms are commonly used to maximize (12) based on (13) with the following gradient and negative Hessian:

$$\mathbf{u}(\boldsymbol{\beta}_{w,\text{vec}}) = \frac{\partial}{\partial \boldsymbol{\beta}_{w,\text{vec}}} \ell(\boldsymbol{\beta}_{w,\text{vec}}), \quad \mathbf{J}(\boldsymbol{\beta}_{w,\text{vec}}) = -\frac{\partial}{\partial \boldsymbol{\beta}_{w,\text{vec}}} \mathbf{u}(\boldsymbol{\beta}_{w,\text{vec}}). \quad (13)$$

## 4.4 Uncertainty Quantification

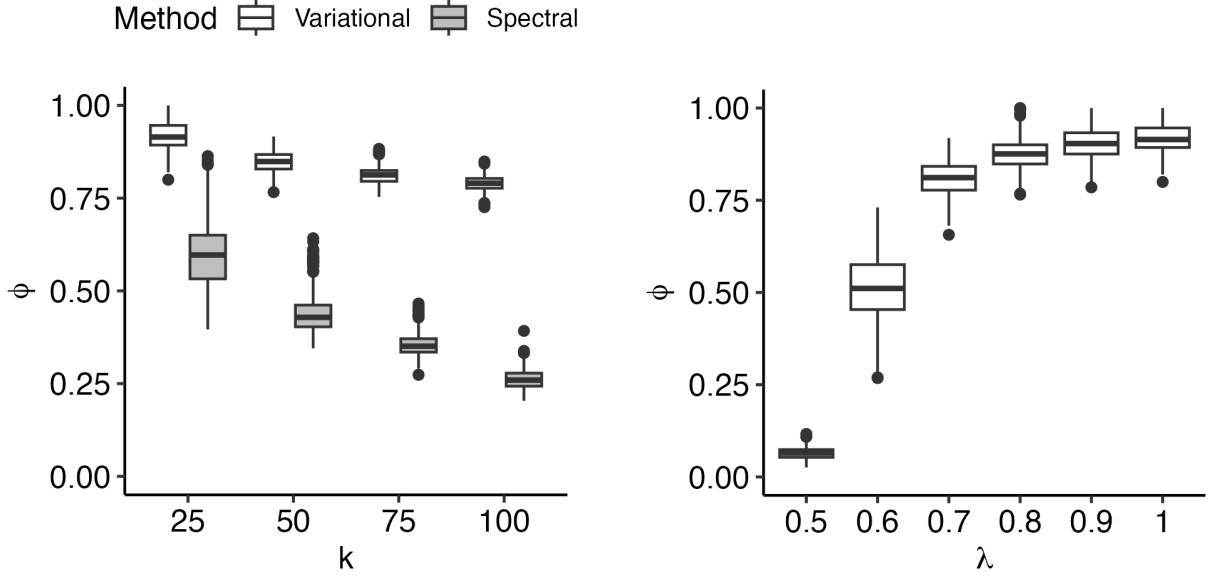
To account for the uncertainty in block allocations, we propose to sample  $T$  times from a multinomial distribution for each node, using their individual block membership probabilities  $\boldsymbol{\alpha}_i$ . For each  $t \in \{1, \dots, T\}$  we sample  $\mathbf{Z}^{(t)}$ , where:

$$\mathbf{Z}_i^{(t)} \sim \text{Multinomial}(1, \boldsymbol{\alpha}_i), \quad \text{for } i = 1, \dots, N$$

Given the previously sampled block memberships,  $\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(T)}$ , the parameters of the model,  $(\boldsymbol{\beta}_{w,\text{vec}}^{(1)}, \boldsymbol{\Sigma}^{(1)}), \dots, (\boldsymbol{\beta}_{w,\text{vec}}^{(T)}, \boldsymbol{\Sigma}^{(T)})$ , are re-estimate for each sampled partition to obtain valid uncertainty quantification. The estimated parameters  $\boldsymbol{\beta}_{w,\text{vec}}^{(t)}$  are pooled as follows

$$\begin{aligned}\mathbb{E}(\boldsymbol{\beta}_{w,\text{vec}}) &= \bar{\boldsymbol{\beta}}_{w,\text{vec}} = \frac{1}{T} \sum_{t=1}^T \hat{\boldsymbol{\beta}}_{w,\text{vec}}^{(t)}, \\ \text{Var}(\boldsymbol{\beta}_{w,\text{vec}}) &= \frac{1}{T} \left( \sum_{t=1}^T \hat{\boldsymbol{\Sigma}}^{(t)} \right) + \frac{1}{T-1} \sum_{t=1}^T \left( \hat{\boldsymbol{\beta}}_{w,\text{vec}}^{(t)} - \bar{\boldsymbol{\beta}}_{w,\text{vec}} \right) \left( \hat{\boldsymbol{\beta}}_{w,\text{vec}}^{(t)} - \bar{\boldsymbol{\beta}}_{w,\text{vec}} \right)^\top.\end{aligned}$$

To assess model fit and enable model selection under this uncertainty quantification, we evaluate the AIC for each sampled partition. By extending the bridge sampler introduced



(a) Block recovery performance in Simulation Study 1, measured by Yule's  $\phi$  coefficient for varying numbers of blocks ( $K = 25, 50, 75, 100$ ) and network sizes ( $N = 1,250$  to  $5,000$ ).

(b) Effect of between-block sparsity on block recovery in Simulation Study 2. Yule's  $\phi$  coefficient is shown for different values of  $\lambda$ , which controls the density of between-block edges.

Figure 1: Block recovery performance across different simulation conditions. Higher  $\phi$  values indicate better agreement with the true block structure.

in Hunter and and Handcock (2006), we compute the AIC for each sample and report the average across all  $T$  partitions. Standard errors  $\hat{\Sigma}^{(t)}$  were computed using the Godambe information matrix. Details are provided in the Supplement A.

## 5 Simulation Study

This section evaluates the performance of the proposed estimation procedure in both steps. For the evaluation of the parameter recovery in Step 2, we use MPLE as they are more scalable than MCMC-based methods. To evaluate the block recovery we use the variational

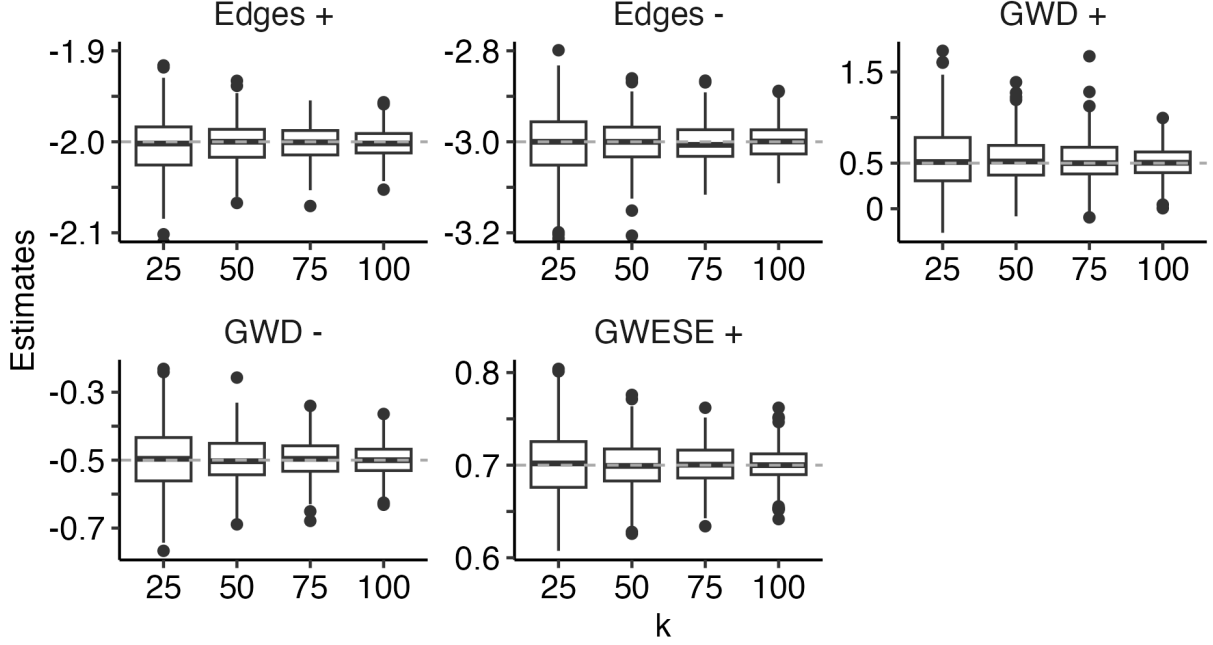


Figure 2: Maximum pseudo-likelihood estimates for within-block parameters in Simulation Study 1. Results are based on known block memberships and used to assess parameter recovery accuracy across increasing network sizes.

approach described in Section 4 and compare it to the spectral clustering described in Lei and Rinaldo (2015), for which we treat the network as a binary one. The block recovery will be assessed using the Yule’s  $\phi$ -coefficient:

$$\Phi(\mathbf{z}^*, \mathbf{z}) = \frac{n_{0,0}n_{1,1} - n_{0,1}n_{1,0}}{\sqrt{(n_{0,0} + n_{0,1})(n_{1,0} + n_{1,1})(n_{0,0} + n_{1,0})(n_{0,1} + n_{1,1})}},$$

where  $n_{0,0}$  represents the number of node pairs assigned to different blocks, while  $n_{1,1}$  counts the pairs assigned to the same block in both. The sum  $n_{0,1} + n_{1,0}$  captures where both assignments differ. The coefficient is 1 if  $\mathbf{z}^*$  and  $\mathbf{z}$  fully agree, and is invariant to block labeling.

**Simulation Study 1: Block & Parameter Recovery.** We simulate networks using the model specification detailed in Example 3 in Section 3, which incorporates geometrically weighted degrees ( $\text{GWD}^{+/-}$ ) and edgewise shared enemies ( $\text{GWESE}^+$ ) statistics. The

networks are drawn using MCMC methods, specifically a Metropolis-Hastings sampler based on the joint distribution, with proposals evaluated using the conditional distribution derived above.

The number of blocks  $K$  each of size 50 on a grid,  $K \in \{25, 50, 75, 100\}$ , resulting in networks of size  $N \in \{1,250, 2,500, 3,750, 5,000\}$ . The within-block coefficients are set to  $\boldsymbol{\theta}_{k,k} = (-2, 0.5, -3, -0.5, 0.7)^\top$  corresponding to the sufficient statistics  $\text{Edges}^+$ ,  $\text{GWD}^+$ ,  $\text{Edges}^-$ ,  $\text{GWD}^-$ , and  $\text{GWESE}^+$ , respectively. The between-block coefficients, corresponding to positive and negative edges, are set to  $\boldsymbol{\theta}_{k,l} = (-1.5, -0.5)^\top \log(N)$ . The log-scaling with  $N$  allows for sparsity in the between-block networks as the total number of nodes increases.

These parameter values are chosen to ensure that the simulated networks meet two criteria. First, the networks exhibit a clear block structure with denser connectivity within blocks than between blocks, which is critical for successful block recovery. Second, the networks resemble real-world signed social networks: the positive  $\text{GWD}^+$  coefficient encourages the formation of nodes with high positive degree, inducing centralization, while the negative  $\text{GWD}^-$  coefficient suppresses high negative degree nodes. The positive  $\text{GWESE}^+$  coefficient reflects structural balance theory, favoring balanced triads where the “*enemy of my enemy is my friend*”.

The generated networks are used to test block recovery by comparing them to binary spectral clustering (see Figure 1a) and to assess parameter recovery using MPLE when true block membership is known (see Figure 2). Our estimation method consistently outperforms spectral clustering in recovering block structure and accurately estimates model coefficients across different numbers of blocks.

**Simulation Study 2: Sparsity.** We conduct a simulation study to evaluate block recovery under varying levels of sparsity. The considered networks consist of  $K = 25$  blocks, each containing 50 nodes, resulting in a total of  $N = 1,250$  nodes. The within-block coefficients remain the same as in the previous scenario. To assess how block recovery is influenced by between-block sparsity, we set the coefficients for positive and negative between-block edges to  $\theta_{k,k} = (-1.5, -0.5)^\top \lambda \log(N)$ , where  $\lambda$  varies from 0.5 to 1. The parameter  $\lambda$  controls the density of the between-block edges: with increasing  $\lambda$ , the sparsity between blocks increases, implying fewer connections between blocks. Figure 1b shows that the block recovery worsens as the number of between-block edges increases (i.e., when  $\lambda$  decreases). This aligns with our expectations as the block structure becomes less pronounced in such regimes.





## 6 Wikipedia Network

We demonstrate the proposed model in an application to a network of Wikipedia editors who contribute to and refine content across a range of topics. Wikipedia is a free, community-edited online encyclopedia that has been the subject of numerous prior studies (Brandes et al., 2009; Iba et al., 2010). A subset of these editors can be considered experts in specific fields, as they focus their contributions on a limited number of Wikipedia pages within their area of expertise. These editors typically exhibit limited awareness of edits or interactions occurring outside their domain. Such behavior gives rise to block structure and localized dependence within the network, where each block corresponds to a distinct area of knowledge.

The edges in the network were constructed in a manner similar to the approach of Lerner and Lomi (2017): a negative edge between users  $i$  and  $j$  is added if either user



Table 1: Network statistics for the Wikipedia editor network, including the number of nodes, positive and negative edges, and frequencies of triads: all-positive (“friend of my friend is my friend”), all-negative, two positive and one negative, and one positive and two negative (“enemy of my enemy is my friend”).

Nodes	Edges +	Edges −				
2,115	875	2,656	78	738	272	934

undoes or deletes the other’s work, and a positive edge if either user redoes the other’s work. The raw dataset includes data from more than 10,000 articles and is described in detail in Lerner and Lomi (2019).

To extract our subsample of experts we filtered the data from the full network of 1,634,189 editors. First, to ensure that we focus on meaningful contributions, we included only users who added at least 100 words per page. We, thereby, distinguish between users who primarily add substantive new content, likely indicating expertise on specific topics, from those who make frequent but less substantive edits, such as corrections or anti-vandalism activities. Second, we excluded users who contributed to more than 10 pages, as this typically indicates bots or users focused on maintenance rather than topic-specific contributions. From the remaining users, we selected 50 pages whose names are provided in Table C. This procedure resulted in an undirected network of  $N = 2,115$  nodes.

## 6.1 Model Specification

We apply the two-stage estimation approach to this network. For this, we assume that each page is an individual block, resulting in 50 blocks in the network. Note, however, that there

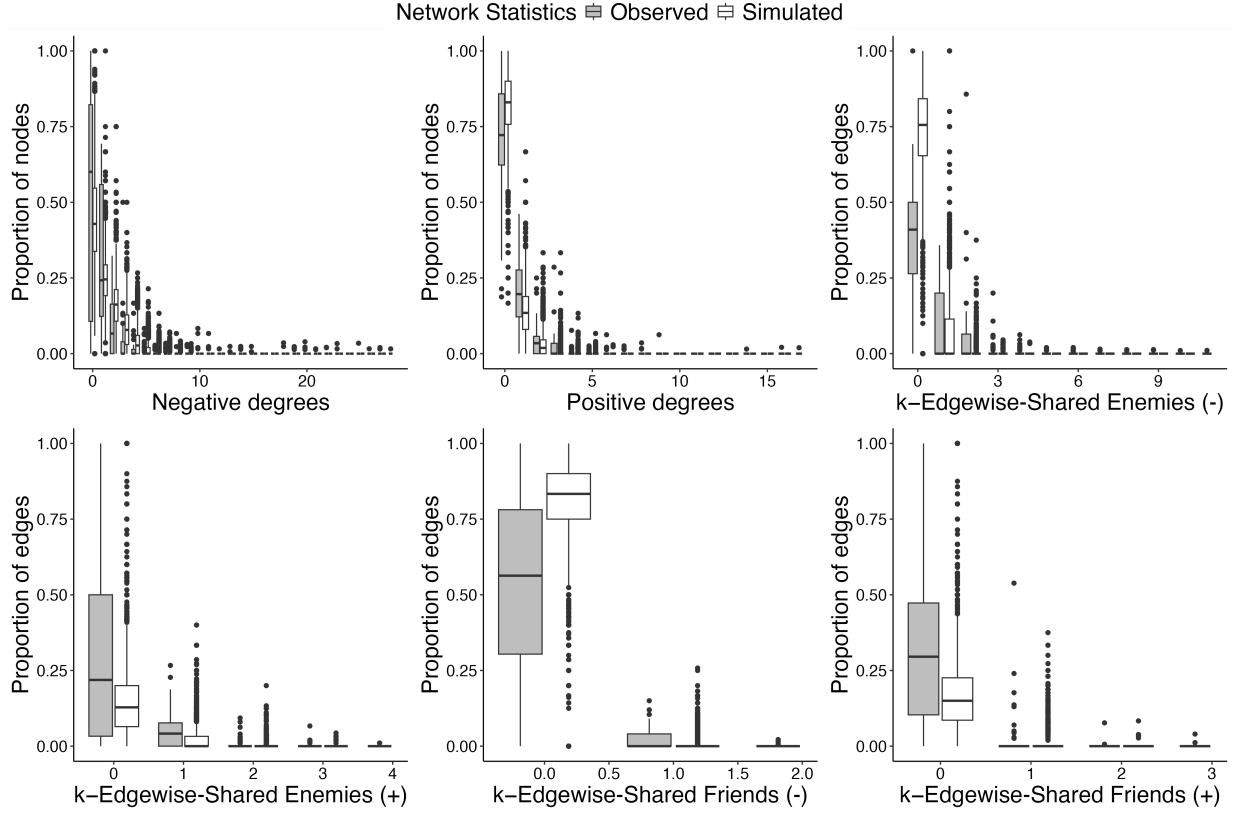


Figure 3: Out-of-sample cross-validation results for the Full Triad model including all triadic terms ( $\text{Edges}^{+/-}$ ,  $\text{GWD}^{+/-}$ , and all GWESP). The distribution of simulated statistics across 100 replications is compared against the observed statistics for each block.

is no ground truth block membership for the nodes, only for the edges, so it is possible that the number of blocks is less than 50 in the case where two pages contain a similar subset of users. Conversely, the number of blocks could be higher than 50 if a single page includes heterogeneous sections that correspond to distinct user groups. That being said, the Yule’s phi coefficient between the estimated block membership and the page an editor contributed to most is 0.69, indicating a moderately strong alignment between the assumed and inferred structures. To account for potential misclassifications in the inferred structure, we apply an uncertainty correction to adjust for estimation error in block assignments.

To decide which model specifications are best suited to describe the observed Wikipedia

network, we adapt the out-of-sample cross-validation for multi-level networks used by Stewart et al. (2019) to our setting. We remove one block and estimate a model for the remaining 49 blocks. The resulting model coefficients are used to individually simulate the previously excluded block 100 times. To best capture signed network structure, we plot the distribution of the observed network statistics against the simulated statistics: degree distributions (positive and negative), edgewise shared friends ( $\text{ESF}^+$  and  $\text{ESF}^-$ ), and edgewise shared enemies ( $\text{ESE}^+$  and  $\text{ESE}^-$ ). In total, we tested four model specifications of increasing complexity, which can be seen in Table 2. The decay term for the geometrically weighted degrees is set to .2, and for the geometrically weighted shared partners (GWESP), it is set to 0, reducing them to a count statistic of the given triad, as introduced in Example 2 in Section 3. The vector of sufficient statistic for the “Full Triad” model is thus given by:

$$\begin{aligned} \mathbf{s}(\mathbf{y}_{k,k}) = & (\text{Edges}^+(\mathbf{y}_{k,k}), \text{Edges}^-(\mathbf{y}_{k,k}), \text{GWD}^+(\mathbf{y}_{k,k}, 0.2), \text{GWD}^-(\mathbf{y}_{k,k}, 0.2), \\ & \text{GWESE}^+(\mathbf{y}_{k,k}, 0), \text{GWESF}^+(\mathbf{y}_{k,k}, 0), \text{GWESE}^-(\mathbf{y}_{k,k}, 0), \\ & \text{GWESF}^-(\mathbf{y}_{k,k}, 0))^{\top}. \end{aligned}$$

The results of the out-of-sample cross-validation, shown in Figure 3 and Supplement C.1, suggest that the Full Triad model, including edges, GWD and all GWESP terms, is better equipped to describe the Wikipedia network.

## 6.2 Results

As described in Section 4, we sample 100 times from the block membership probability to account for the uncertainty within latent block memberships. The results of the second step estimation can be found in Table 2. Standard errors were estimated using the Godambe matrix based on 100 simulations.

The edges terms are modeled as a function of block size  $\log(N_k)$ . The terms  $\text{Edges}^+$  and  $\text{Edges}^-$  represent the intercepts, which alone are not directly interpretable, as they

Table 2: Estimated within-block parameters with uncertainty correction ( $T = 100$ ). Coef. indicates the estimated parameter, and SE is the associated standard error, based on the estimated Godambe information matrix, using 100 simulations per block. Models compared are: I (Independent), I+D (Degree), I+D+PT (Partial Triad), and I+D+FT (Full Triad).  $\Delta\text{AIC}$  values indicate the difference in AIC relative to the independent model.

	Independent		Degree		Partial Triad		Full Triad	
Parameters	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Edges <sup>+</sup>	.543	.268	2.098	.207	1.283	.172	.475	.185
$\times \log(N_k)$	-1.122	.058	-1.160	.042	-1.018	.034	-.932	.036
Edges <sup>-</sup>	.688	.225	1.941	.189	1.700	.168	1.167	.187
$\times \log(N_k)$	-.950	.048	-1.034	.038	-.980	.034	-.904	.033
GWD <sup>+</sup>			-1.106	.065	-1.291	.055	-1.076	.060
GWD <sup>-</sup>			-.945	.051	-1.110	.056	-1.032	.066
GWESE <sup>+</sup>					1.334	.061	1.422	.047
GWESF <sup>+</sup>							.224	.029
GWESE <sup>-</sup>							.126	.028
GWESF <sup>-</sup>							.688	.035
$\Delta\text{AIC}$	0		855		1,296		1,404	

would only apply to hypothetical blocks of size 1. Instead, the combination of intercept and block size effect reflects the expected edge probability within a given block. Across all models, the coefficients for positive edges are lower than for negative edges, indicating that negative edges are more frequent within blocks. The consistently negative coefficients for the block size terms demonstrates that larger blocks tend to be sparser, with edge

probability decreasing as block size increases. This reflects the fact that the number of possible edges grows quadratically, while the number of actual edges grows at a slower pace. Adding GWD terms reduces the edge coefficients, suggesting that some variation in edge probability is explained by degree structure. The GWD coefficients are negative for both edge types in all models, indicating a general tendency against high-degree nodes. The effect is stronger for positive edges, implying that these edges are more evenly distributed and less likely to form hubs. As expected from structural balance theory, the  $\text{GWESE}^+$  terms have positive and statistically significant coefficients in both the Full and Partial Triad model. This means that users are more likely to restore the work of another user if they share a common enemy. While still positive and significant, the other balanced triad represented by  $\text{GWESF}^+$  is closer to zero, similar to the unbalanced triad  $\text{GWESE}^-$ . However, the term  $\text{GWESF}^-$  is significantly larger than  $\text{GWESF}^+$ , which is not consistent with structural balance theory. This discrepancy suggests that structural balance theory may not fully apply to this expert network. For instance, undoing a friend-of-a-friend’s contribution could be more common due to task-related disagreements rather than group-based antagonism. Overall, the more complex models capture important network features, such as triadic relationships, and this is reflected in their lower AIC values, indicating better model fit despite increased complexity.

Our Wikipedia network analysis of thousands of subject matter experts serves an example where global dependence assumption seems unrealistic. Through the use of local dependence, we found some adherence to the behavioral norms predicted by structural balance theory. Specifically, users who have an enemy in common are more likely to restore each other’s contributions. This finding is consistent with Lerner and Lomi (2019).

To validate the results in Table 2, we conducted a conventional ERGM goodness-of-fit

analysis following the method outlined by Hunter et al. (2008). The results are shown in Supplement C.2.

## 7 Conclusion

Our proposed model for signed networks combines the strengths of stochastic block models and exponential random graph models, while being scalable to thousands of nodes. There are several directions for extending the proposed model. First, we assumed that the number of blocks,  $K$ , was known a priori. In practice, this may not hold, and methods to infer  $K$  from data would enhance model flexibility (Saldaña et al., 2017). Second, the assumption of nonoverlapping community structure could be relaxed by allowing nodes to belong to multiple blocks, as in mixed-membership models (see, e.g., Latouche et al., 2011). This would accommodate for more realistic block structures. Third, the observed degree heterogeneity in the Wikipedia network highlights the need for a degree-corrected stochastic block model for signed networks.

## Acknowledgments

The authors would like to thank the HPC Service of FUB-IT, Freie Universität Berlin, for computing time (Bennett et al., 2020).

## References

Babkin, S., J. R. Stewart, X. Long, and M. Schweinberger (2020). Large-scale estimation of random graph models with local dependence. *Computational Statistics & Data Analysis* 152, 107029. doi:10.1016/j.csda.2020.107029.

- Bennett, L., B. Melchers, and B. Proppe (2020). Curta: A general-purpose high-performance computer at ZEDAT, Freie Universität Berlin. doi: 10.17169/refubium-26754.
- Bhamidi, S., G. Bresler, and A. Sly (2011). Mixing time of exponential random graphs. *The Annals of Applied Probability* 21(6), 2146–2170. doi:10.1214/10-AAP740.
- Brandes, U., P. Kenis, J. Lerner, and D. van Raaij (2009). Network analysis of collaboration structure in Wikipedia. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, New York, NY, USA, pp. 731–740. Association for Computing Machinery. doi:10.1145/1526709.1526808.
- Butts, C. T. and Z. W. and Almquist (2015). A flexible parameterization for baseline mean degree in multiple-network ERGMs. *The Journal of Mathematical Sociology* 39(3), 163–167. doi:10.1080/0022250X.2014.967851.
- Dahbura, J. N. M., S. Komatsu, T. Nishida, and A. Mele (2021). A structural model of business card exchange networks. doi:10.48550/arXiv.2105.12704.
- Daudin, J.-J., F. Picard, and S. Robin (2008). A mixture model for random graphs. *Statistics and Computing* 18(2), 173–183. doi:10.1007/s11222-007-9046-7.
- David, B., K. Martin, and M. Irini (2011). *Latent Variable Models and Factor Analysis: A Unified Approach*. Chichester, West Sussex, United Kingdom: John Wiley & Sons, Ltd. doi:10.1002/9781119970583.
- De Nicola, G., C. Fritz, M. Mehrl, and G. Kauermann (2023). Dependence matters: Statistical models to identify the drivers of tie formation in economic networks. *Journal of Economic Behavior & Organization* 215, 351–363. doi:10.1016/j.jebo.2023.09.021.

- Fienberg, S. E. and S. S. Wasserman (1981). Categorical data analysis of single sociometric relations. *Sociological Methodology* 12, 156–192. doi:10.2307/270741.
- Fritz, C., M. Mehrl, P. W. Thurner, and G. Kauermann (2025). Exponential random graph models for dynamic signed networks: An application to international relations. *Political Analysis* 33(3), 211–230. doi:10.1017/pan.2024.21.
- Handcock, M. S. (2003). Statistical models for social networks: Inference and degeneracy. In R. Breiger, K. Carley, and P. Pattison (Eds.), *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, pp. 1–12. Washington, D.C.: National Academies Press. doi:10.17226/10735.
- Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology*. doi:10.1037/h0055425.
- Hunter, D. R., G. , Steven M, and M. S. and Handcock (2008). Goodness of fit of social network models. *Journal of the American Statistical Association* 103(481), 248–258. doi:10.1198/016214507000000446.
- Hunter, D. R. (2007). Curved exponential family models for social networks. *Social Networks* 29(2), 216–230. doi:10.1016/j.socnet.2006.08.005.
- Hunter, D. R. and M. S. and Handcock (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics* 15(3), 565–583. doi:10.1198/106186006X133069.
- Iba, T., K. Nemoto, B. Peters, and P. A. Gloor (2010). Analyzing the creative editing behavior of Wikipedia editors: Through dynamic social network analysis. *Procedia - Social and Behavioral Sciences* 2(4), 6441–6456. doi:10.1016/j.sbspro.2010.04.054.



- Jiang, J. Q. (2015). Stochastic block model and exploratory analysis in signed networks. *Physical Review E* 91(6), 062805. doi:10.1103/PhysRevE.91.062805.
- Krivitsky, P. N., C. , Pietro, and N. and Hens (2023). A tale of two datasets: Representativeness and generalisability of inference for samples of networks. *Journal of the American Statistical Association* 118(544), 2213–2224. doi:10.1080/01621459.2023.2242627.
- Krivitsky, P. N., M. S. Handcock, and M. Morris (2011). Adjusting for network size and composition effects in exponential-family random graph models. *Statistical Methodology* 8, 319–339. doi:10.1016/j.stamet.2011.01.005.
- Latouche, P., E. Birmelé, and C. Ambroise (2011). Overlapping stochastic block models with application to the French political blogosphere. *The Annals of Applied Statistics* 5(1), 309–336. doi:10.1214/10-AOAS382.
- Lazer, D., A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne (2009). Computational social science. *Science* 323(5915), 721–723. doi:10.1126/science.1167742.
- Lei, J. and A. Rinaldo (2015). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics* 43(1), 215–237. doi:10.1214/14-AOS1274.
- Lerner, J. and A. Lomi (2017). The third man: Hierarchy formation in Wikipedia. *Applied Network Science* 2(1), 24. doi:10.1007/s41109-017-0043-2.
- Lerner, J. and A. Lomi (2019). The network structure of successful collaboration in Wikipedia. In *Proceedings of the 52nd Annual Hawaii International Conference*

- on *System Sciences*, Honolulu, HI, pp. 2622–2631. University of Hawai’i at Manoa. doi:10.24251/HICSS.2019.316.
- Leskovec, J., D. Huttenlocher, and J. Kleinberg (2010). Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’10, New York, NY, USA, pp. 1361–1370. Association for Computing Machinery. doi:10.1145/1753326.1753532.
- Li, Y., B. Yang, X. Zhao, Z. Yang, and H. Chen (2023). SSBM: A signed stochastic block model for multiple structure discovery in large-scale exploratory signed networks. *Knowledge-Based Systems* 259, 110068. doi:10.1016/j.knosys.2022.110068.
- Lusher, D., J. Koskinen, and G. Robins (2013). *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge University Press. doi:10.1017/CBO9780511894701.
- Matias, C. and S. Robin (2014). Modeling heterogeneity in random graphs through latent space models: A selective review. *ESAIM: Proceedings and Surveys* 47, 55–74. doi:10.1051/proc/201447004.
- Nowicki, K. and T. A. B. and Snijders (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association* 96(455), 1077–1087. doi:10.1198/016214501753208735.
- Saldaña, D. F., Y. , Yi, and Y. and Feng (2017). How many communities are there? *Journal of Computational and Graphical Statistics* 26(1), 171–181. doi:10.1080/10618600.2015.1096790.
- Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponen-

- tial families. *Journal of the American Statistical Association* 106(496), 1361–1370. doi:10.1198/jasa.2011.tm10747.
- Schweinberger, M. and J. R. Stewart (2020). Concentration and consistency results for canonical and curved exponential-family models of random graphs. *The Annals of Statistics* 48, 374–396. doi:10.1214/19-AOS1810.
- Slaughter, A. J. and L. M. Koehly (2016). Multilevel models for social networks: Hierarchical bayesian approaches to exponential random graph modeling. *Social Networks* 44, 334–345. doi:10.1016/j.socnet.2015.11.002.
- Snijders, T. A. and K. Nowicki (1997). Estimation and prediction for stochastic block-models for graphs with latent block structure. *Journal of Classification* 14(1), 75–100. doi:10.1007/s003579900004.
- Snijders, T. A. B. (2007). Discussion on the paper by Handcock, Raftery and Tantrum. *Journal of the Royal Statistical Society: Statistics in Society Series A* 170, 324. doi:10.1111/j.1467-985X.2007.00471.x.
- Stefanov, S. M. (2004). Convex quadratic minimization subject to a linear constraint and box constraints. *Applied Mathematics Research Express* 2004, 17–42. doi:10.1155/S168712000402009X.
- Stewart, J., M. Schweinberger, M. Bojanowski, and M. 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, D. and M. and Ikeda (1990). Pseudolikelihood estimation for social

networks. *Journal of the American Statistical Association* 85(409), 204–212.  
doi:10.1080/01621459.1990.10475327.

Vu, D. Q., D. R. Hunter, and M. Schweinberger (2013). Model-based clustering of large networks. *The Annals of Applied Statistics* 7(2), 1010–1039. doi:10.1214/12-AOAS617.

Wasserman, S. and K. Faust (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511815478.

# Supplementary Materials:

## Scalable Signed Exponential Random Graph Models under Local Dependence

A	Godambe Information . . . . .	1
B	Computation . . . . .	2
B.1	Surrogate Function . . . . .	2
B.2	Sparse Matrix Multiplication . . . . .	3
B.3	Update Rules . . . . .	5
C	Wikipedia Network . . . . .	6
C.1	Out-of-Sample Cross Validation . . . . .	7
C.2	Goodness-of-Fit . . . . .	13

### A Godambe Information

When using MPLE, the standard errors  $\hat{\Sigma}^{(t)}$  obtained from a logistic regression are, as Strauss and Ikeda (1990) and ? pointed out, inappropriate because dyadic independence is assumed. ? suggested a more appropriate method for estimating MPLE standard errors by calculating the Godambe matrix ?. The Godambe matrix is

$$\mathbf{G}(\boldsymbol{\theta}^{(t)}) = \mathbf{J}(\boldsymbol{\theta}^{(t)})^{-1} \mathbf{V}(\boldsymbol{\theta}^{(t)}) \mathbf{J}(\boldsymbol{\theta}^{(t)})^{-1},$$

where  $\mathbf{J}(\boldsymbol{\theta}^{(t)})$  is defined in (13) and  $\mathbf{V}(\boldsymbol{\theta}^{(t)}) := \text{Var}(\mathbf{u}(\boldsymbol{\theta}^{(t)}))$  is called the variability matrix. However,  $\mathbf{V}(\boldsymbol{\theta}^{(t)})$  cannot be directly computed and must be approximated by simulating  $R$  networks using MCMC procedures and calculating the vector of first derivatives of the pseudo-likelihood function  $\mathbf{u}_1(\boldsymbol{\theta}), \dots, \mathbf{u}_R(\boldsymbol{\theta})$  for each of the simulated network. The

variability matrix is then approximated by

$$\hat{\mathbf{V}}(\boldsymbol{\theta}^{(t)}) = \frac{1}{R-1} \sum_{r=1}^R (\mathbf{u}_r(\boldsymbol{\theta}^{(t)}) - \bar{\mathbf{u}}(\boldsymbol{\theta}^{(t)})) (\mathbf{u}_r(\boldsymbol{\theta}^{(t)}) - \bar{\mathbf{u}}(\boldsymbol{\theta}^{(t)}))^T,$$

where  $\bar{\mathbf{u}}(\boldsymbol{\theta}^{(t)}) := 1/R \sum_{r=1}^R \mathbf{u}_r(\boldsymbol{\theta}^{(t)})$  denotes the sample mean of the score vectors.

## B Computation

### B.1 Surrogate Function

In order to detail how the updates of  $\boldsymbol{\alpha}$  are carried out using sparse matrix operations, we introduce a surrogate function  $Q(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\theta}^{(t)}; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\alpha})$  that provides a tractable lower bound that can be maximized iteratively and rearrange it as follows:

$$\begin{aligned} Q(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\theta}^{(t)}; \boldsymbol{\alpha}^{(t)}, \boldsymbol{\alpha}) &= \sum_{i < j} \sum_{k=1}^K \sum_{l=1}^K \left( \alpha_{i,k}^2 \frac{\alpha_{j,l}^{(t)}}{2\alpha_{i,k}^{(t)}} + \alpha_{j,l}^2 \frac{\alpha_{i,k}^{(t)}}{2\alpha_{j,l}^{(t)}} \right) \log p_{k,l}(y_{i,j}) \\ &+ \sum_{i=1}^N \sum_{k=1}^K \alpha_{i,k} \left( \log \gamma_k^{(t)} - \log \alpha_{i,k}^{(t)} - \frac{\alpha_{i,k}}{\alpha_{i,k}^{(t)}} + 1 \right) \\ &= \sum_{i=1}^N \sum_{k=1}^K \frac{\Omega_{i,k}^{(t)}(\mathbf{y}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\theta}^{(t)})}{\alpha_{i,k}^{(t)}} \alpha_{i,k}^2 + \sum_{i=1}^N \sum_{k=1}^K \alpha_{i,k} \left( \log \gamma_k^{(t)} - \log \alpha_{i,k}^{(t)} - \frac{\alpha_{i,k}}{\alpha_{i,k}^{(t)}} + 1 \right) \\ &= \sum_{i=1}^N \sum_{k=1}^K \left( \frac{\Omega_{i,k}^{(t)}(\mathbf{y}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\theta}^{(t)})}{\alpha_{i,k}^{(t)}} - \frac{1}{\alpha_{i,k}^{(t)}} \right) \alpha_{i,k}^2 + \left( \log \gamma_k^{(t)} - \log \alpha_{i,k}^{(t)} + 1 \right) \alpha_{i,k} \\ &= \sum_{i=1}^N \sum_{k=1}^K A_{i,k}(\mathbf{y}, \boldsymbol{\alpha}^{(t)}) \alpha_{i,k}^2 + B_{i,k}(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\alpha}^{(t)}) \alpha_{i,k} \end{aligned}$$

with

$$\Omega_{i,k}^{(t)}(\mathbf{y}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\theta}^{(t)}) := \sum_{j \neq i} \sum_{l=1}^K \alpha_{j,l}^{(t)} \log p_{k,l}(y_{i,j}),$$

the quadratic term

$$A_{i,k}(\mathbf{y}, \boldsymbol{\alpha}^{(t)}) := \frac{\Omega_{i,k}^{(t)}(\mathbf{y}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\theta}^{(t)})}{\alpha_{i,k}^{(t)}} - \frac{1}{\alpha_{i,k}^{(t)}}$$

and the linear term of the quadratic problem

$$B_{i,k}(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\alpha}^{(t)}) := \log \gamma_k^{(t)} - \log \alpha_{i,k}^{(t)} + 1.$$

## B.2 Sparse Matrix Multiplication

To update our estimate of  $\boldsymbol{\alpha}$ , we need to evaluate  $A_{i,k}(\mathbf{y}, \boldsymbol{\alpha}^{(t)})$  and  $B_{i,k}(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\alpha}^{(t)})$ , with the former being the problematic part, giving us a complexity of  $\mathcal{O}(N^2 K^2)$ . Therefore, we decompose  $\Omega_{i,k}^{(t)}(\mathbf{y}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\theta}^{(t)})$  into two parts: one when the network is completely empty, i.e.,  $y_{i,j} = 0$  for all  $i, j = 1, \dots, N$ , and the other capturing all dyads where  $y_{i,j} \neq 0$ , which allows us to exploit the sparse nature of the network:

$$\begin{aligned}
\Omega_{i,k}^{(t)}(\mathbf{y}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\theta}^{(t)}) &:= \sum_{j \neq i}^N \sum_{l=1}^K \alpha_{j,l}^{(t)} \log p_{k,l}^{(t)}(y_{i,j}) \\
&= \sum_{j \neq i}^N \sum_{l=1}^K \alpha_{j,l}^{(t)} \log p_{k,l}^{(t)}("0") + \sum_{j \neq i}^N \mathbb{I}(y_{i,j} = "+") \sum_{l=1}^K \alpha_{j,l}^{(t)} \log \frac{p_{k,l}^{(t)}("+")}{p_{k,l}^{(t)}("0")} \\
&\quad + \sum_{j \neq i}^N \mathbb{I}(y_{i,j} = "-") \sum_{l=1}^K \alpha_{j,l}^{(t)} \log \frac{p_{k,l}^{(t)}("-")}{p_{k,l}^{(t)}("0")} \\
&= \Omega_{i,k}^{(t)}(\mathbf{0}, \boldsymbol{\alpha}^{(t)}) + \Lambda_{i,k}^{(t)}(\mathbf{y}, \boldsymbol{\alpha}^{(t)})
\end{aligned}$$

with

$$\begin{aligned}
\Lambda_{i,k}^{(t)}(\mathbf{y}, \boldsymbol{\alpha}^{(t)}) &= \sum_{j \neq i}^N \mathbb{I}(y_{i,j} = "+") \sum_{l=1}^K \alpha_{j,l}^{(t)} \log \frac{p_{k,l}^{(t)}("+")}{p_{k,l}^{(t)}("0")} \\
&\quad + \sum_{j \neq i}^N \mathbb{I}(y_{i,j} = "-") \sum_{l=1}^K \alpha_{j,l}^{(t)} \log \frac{p_{k,l}^{(t)}("-")}{p_{k,l}^{(t)}("0")}
\end{aligned}$$

The first part, in which we assume the network is completely empty, can be computed using matrix multiplication as follows:

$$\begin{aligned}
\Omega_{i,k}^{(t)}(\mathbf{0}, \boldsymbol{\alpha}^{(t)}) &= \sum_{j \neq i}^N \sum_{l=1}^K \alpha_{j,l}^{(t)} \log p_{k,l}^{(t)}("0") \\
&= \sum_{l=1}^K \left( \underbrace{\sum_{j=1}^N \alpha_{j,l}^{(t)}}_{:=\tau_l^{(t)}} - \alpha_{i,l}^{(t)} \right) \log p_{k,l}^{(t)}("0") \\
&= \sum_{l=1}^K \left( \tau_l^{(t)} - \alpha_{i,l}^{(t)} \right) \log p_{k,l}^{(t)}("0")
\end{aligned}$$

Define

$$\mathbf{A}_0^{(t)} := \begin{bmatrix} \tau_1^{(t)} - \alpha_{1,1}^{(t)} & \tau_2^{(t)} - \alpha_{1,2}^{(t)} & \dots & \tau_K^{(t)} - \alpha_{1,K}^{(t)} \\ \tau_1^{(t)} - \alpha_{2,1}^{(t)} & \tau_2^{(t)} - \alpha_{2,2}^{(t)} & \dots & \tau_K^{(t)} - \alpha_{2,K}^{(t)} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_1^{(t)} - \alpha_{N,1}^{(t)} & \tau_2^{(t)} - \alpha_{N,2}^{(t)} & \dots & \tau_K^{(t)} - \alpha_{N,K}^{(t)} \end{bmatrix}$$

and

$$\mathbf{P}_0^{(t)} := \begin{bmatrix} \log p_{1,1}^{(t)}("0") & \log p_{1,2}^{(t)}("0") & \dots & \log p_{1,K}^{(t)}("0") \\ \log p_{2,1}^{(t)}("0") & \log p_{2,2}^{(t)}("0") & \dots & \log p_{2,K}^{(t)}("0") \\ \vdots & \vdots & \ddots & \vdots \\ \log p_{K,1}^{(t)}("0") & \log p_{K,2}^{(t)}("0") & \dots & \log p_{K,K}^{(t)}("0") \end{bmatrix}.$$

Then  $\Omega_{i,k}^{(t)}(\mathbf{0}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\theta}^{(t)})$  is given by the  $(i, k)$  entry of  $\mathbf{A}_0^{(t)} \mathbf{P}_0^{(t)}$ .

Next, we correct the error arising from the assumption of an entirely empty network by calculating the term where this assumption does not apply:

$$\begin{aligned} \Lambda_{i,k}^{(t)}(\mathbf{y}, \boldsymbol{\alpha}^{(t)}) &= \sum_{j \neq i}^N \mathbb{I}(y_{i,j} = "+") \sum_{l=1}^K \alpha_{j,l}^{(t)} \log \frac{p_{k,l}^{(t)}("+")}{p_{k,l}^{(t)}("0")} \\ &+ \sum_{j \neq i}^N \mathbb{I}(y_{i,j} = "-") \sum_{l=1}^K \alpha_{j,l}^{(t)} \log \frac{p_{k,l}^{(t)}("-")}{p_{k,l}^{(t)}("0")} \end{aligned}$$

Define

$$\mathbf{P}_+^{(t)} := \begin{bmatrix} \log \frac{p_{11}^{(t)}("+")}{p_{11}^{(t)}("0")} & \log \frac{p_{12}^{(t)}("+")}{p_{12}^{(t)}("0")} & \dots & \log \frac{p_{1K}^{(t)}("+")}{p_{1K}^{(t)}("0")} \\ \log \frac{p_{21}^{(t)}("+")}{p_{21}^{(t)}("0")} & \log \frac{p_{22}^{(t)}("+")}{p_{22}^{(t)}("0")} & \dots & \log \frac{p_{2K}^{(t)}("+")}{p_{2K}^{(t)}("0")} \\ \vdots & \vdots & \ddots & \vdots \\ \log \frac{p_{K1}^{(t)}("+")}{p_{K1}^{(t)}("0")} & \log \frac{p_{K2}^{(t)}("+")}{p_{K2}^{(t)}("0")} & \dots & \log \frac{p_{KK}^{(t)}("+")}{p_{KK}^{(t)}("0")} \end{bmatrix}$$



and

$$\mathbf{P}_{-}^{(t)} \underset{(K \times K)}{:=} \begin{bmatrix} \log \frac{p_{11}^{(t)}(\text{"-"})}{p_{11}^{(t)}(\text{"0"})} & \log \frac{p_{12}^{(t)}(\text{"-"})}{p_{12}^{(t)}(\text{"0"})} & \dots & \log \frac{p_{1K}^{(t)}(\text{"-"})}{p_{1K}^{(t)}(\text{"0"})} \\ \log \frac{p_{21}^{(t)}(\text{"-"})}{p_{21}^{(t)}(\text{"0"})} & \log \frac{p_{22}^{(t)}(\text{"-"})}{p_{22}^{(t)}(\text{"0"})} & \dots & \log \frac{p_{2K}^{(t)}(\text{"-"})}{p_{2K}^{(t)}(\text{"0"})} \\ \vdots & \vdots & \ddots & \vdots \\ \log \frac{p_{K1}^{(t)}(\text{"-"})}{p_{K1}^{(t)}(\text{"0"})} & \log \frac{p_{K2}^{(t)}(\text{"-"})}{p_{K2}^{(t)}(\text{"0"})} & \dots & \log \frac{p_{KK}^{(t)}(\text{"-"})}{p_{KK}^{(t)}(\text{"0"})} \end{bmatrix}.$$

Then  $\Lambda_{i,k}^{(t)}(\mathbf{y}, \boldsymbol{\alpha}^{(t)})$  is given by the  $(i, k)$  entry of  $\mathbf{y}_+ \boldsymbol{\alpha}^{(t)} \mathbf{P}_+^{(t)} + \mathbf{y}_- \boldsymbol{\alpha}^{(t)} \mathbf{P}_-^{(t)}$ .

The extension of this methodology is straightforward and is demonstrated for the binary case in ?.

### B.3 Update Rules

In this section we provide details on the update rules of the estimated block membership probability  $\boldsymbol{\alpha} \in [0, 1]^{N \times K}$  for each node  $n$  for each block  $k$ , the prior block probability  $\gamma_k \in [0, 1]$  for each block  $k$  and the edge probability  $\pi_{k,l} \in [0, 1]$  between each pair of blocks  $(k, l)$ . These parameters are required for the first step of our two step estimation approach.

The updated rules of  $\boldsymbol{\alpha}$ ,  $\gamma_k$  and  $\pi_{k,l}(y_{i,j})$  follow

$$\begin{aligned} \boldsymbol{\alpha}^{(t+1)} &:= \arg \max_{\boldsymbol{\alpha}} Q(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\theta}^{(t)}, \boldsymbol{\alpha}^{(t)}; \boldsymbol{\alpha}) \\ \gamma_k^{(t+1)} &:= \frac{1}{N} \sum_{i=1}^N \alpha_{i,k}^{(t+1)}, \text{ for } k = 1, \dots, K \\ \pi_{k,l}^{(t+1)}(y) &:= \frac{\sum_{i < j} \alpha_{i,k}^{(t+1)} \alpha_{j,l}^{(t+1)} \mathbb{I}(y_{i,j} = y)}{\sum_{i < j} \alpha_{i,k}^{(t+1)} \alpha_{j,l}^{(t+1)}} \end{aligned}$$

for  $k, l = 1, \dots, K$  and  $y \in \mathcal{S} := \{\text{"-"}, \text{"0"}, \text{"+"}\}$ .

Table 3: List of selected Wikipedia pages used to construct the editor network.

Wikipedia pages	
Second Taranaki War	Operation Nifty Package
Te Kooti's War	Box Hill Hawks Football Club
2011 Turkish sports corruption scandal	Comparison of MUTCD-influenced traffic signs
Turkish Cup	Calogero Vizzini
Navenby	Partial pressure
Sunderland Echo	Edmund Lyons, 1st Baron Lyons
2009-10 Ukrainian First League	John Jervis, 1st Earl of St Vincent
2010 S.League	Falcon's Fury
King's Cup	Hogwarts Express (Universal Orlando Resort)
TT Pro League	Francization of Brussels
Akhtar Hameed Khan	Leonel Brizola
Ishaq Dar	Glenda Farrell
Alias (season 5)	List of Maverick episodes
List of The Listener episodes	Herne Hill railway station
Argentina women's national field hockey team	LSWR N15 class
Australia national baseball team	Hillsborough Area Regional Transit
Lena Park	List of state highways in Arkansas
Battle of Flamborough Head	The Verge
Uzalo	Iveta Mukuchyan
Margaret (singer)	Japanese aircraft carrier Hiryū
Principalía	Kulothunga Chola III
Northern Province, Sri Lanka	Lee Purcell
New Guinea singing dog	Page Two (EP)
Talbot Tagora	Tatiana Troyanos
Little Thetford	Temple of Eshmun

## C Wikipedia Network

This appendix provides supplementary material for the Wikipedia network application described in Section 6. It includes the list of the 50 selected Wikipedia pages, out-of-sample cross-validation as well as in-sample goodness of fit plots and a visualization of the full Wikipedia network. These materials support the evaluation of model fit and illustrate the network structure underlying the analysis.

## C.1 Out-of-Sample Cross Validation

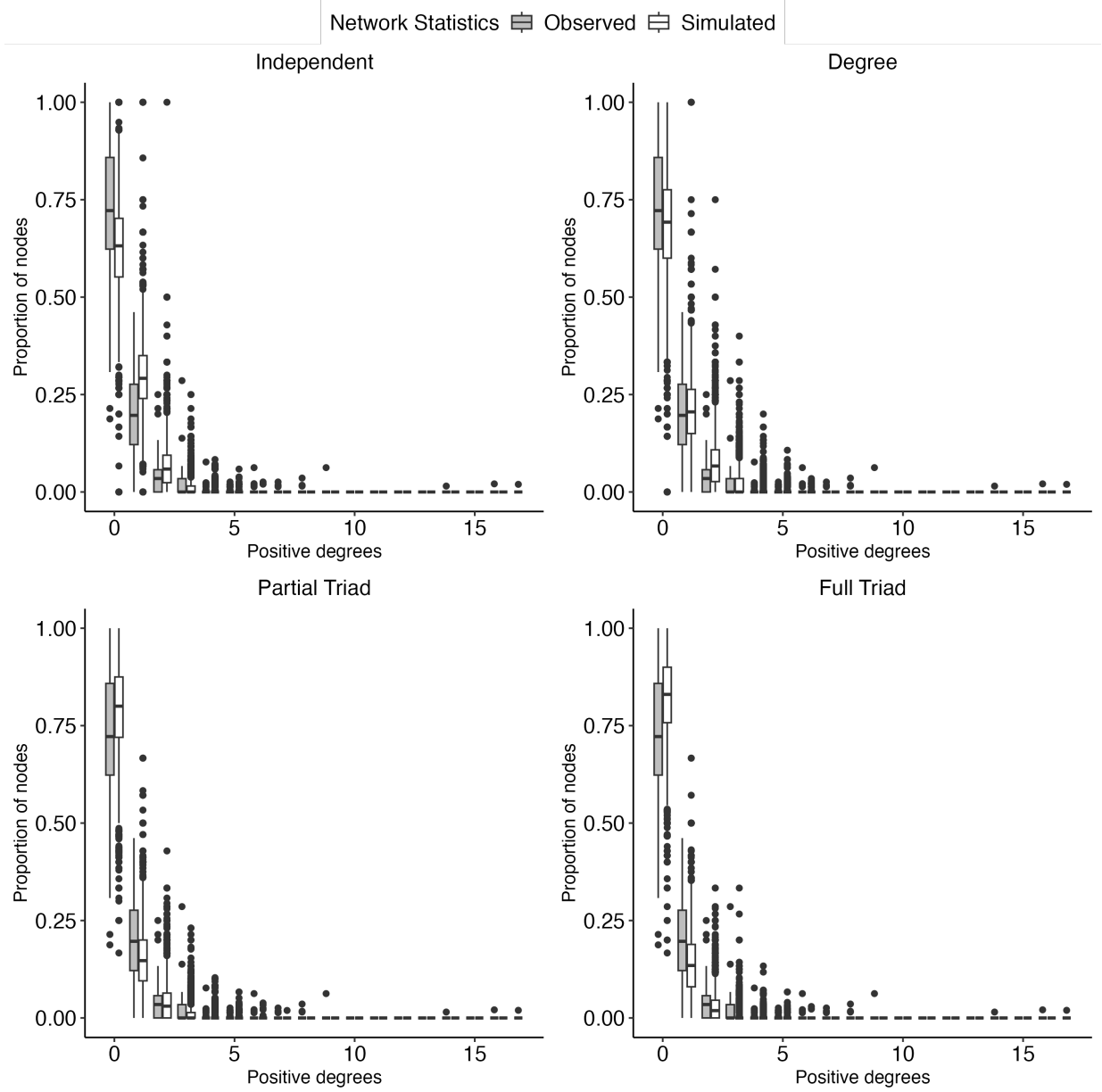


Figure 4: Comparison of out-of-sample cross-validation results for positive degree distribution. The distribution of simulated statistics across 100 replications is compared against the observed statistics for each block. Models compared are: I (Independent), I+D (Degree), I+D+PT (Partial Triad), and I+D+FT (Full Triad).

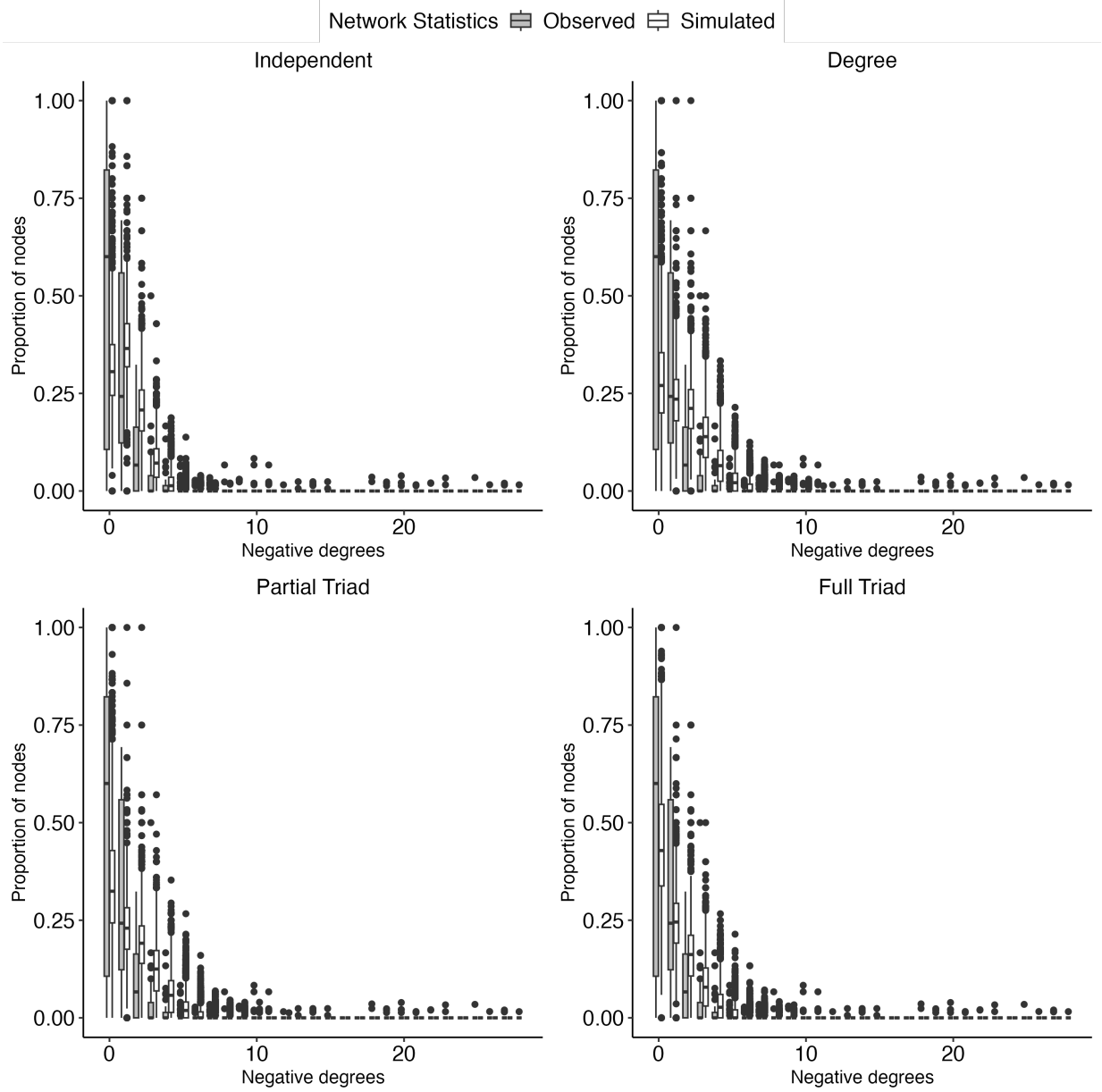


Figure 5: Comparison of out-of-sample cross-validation results for negative degree distribution. The distribution of simulated statistics across 100 replications is compared against the observed statistics for each block. Models compared are: I (Independent), I+D (Degree), I+D+PT (Partial Triad), and I+D+FT (Full Triad).

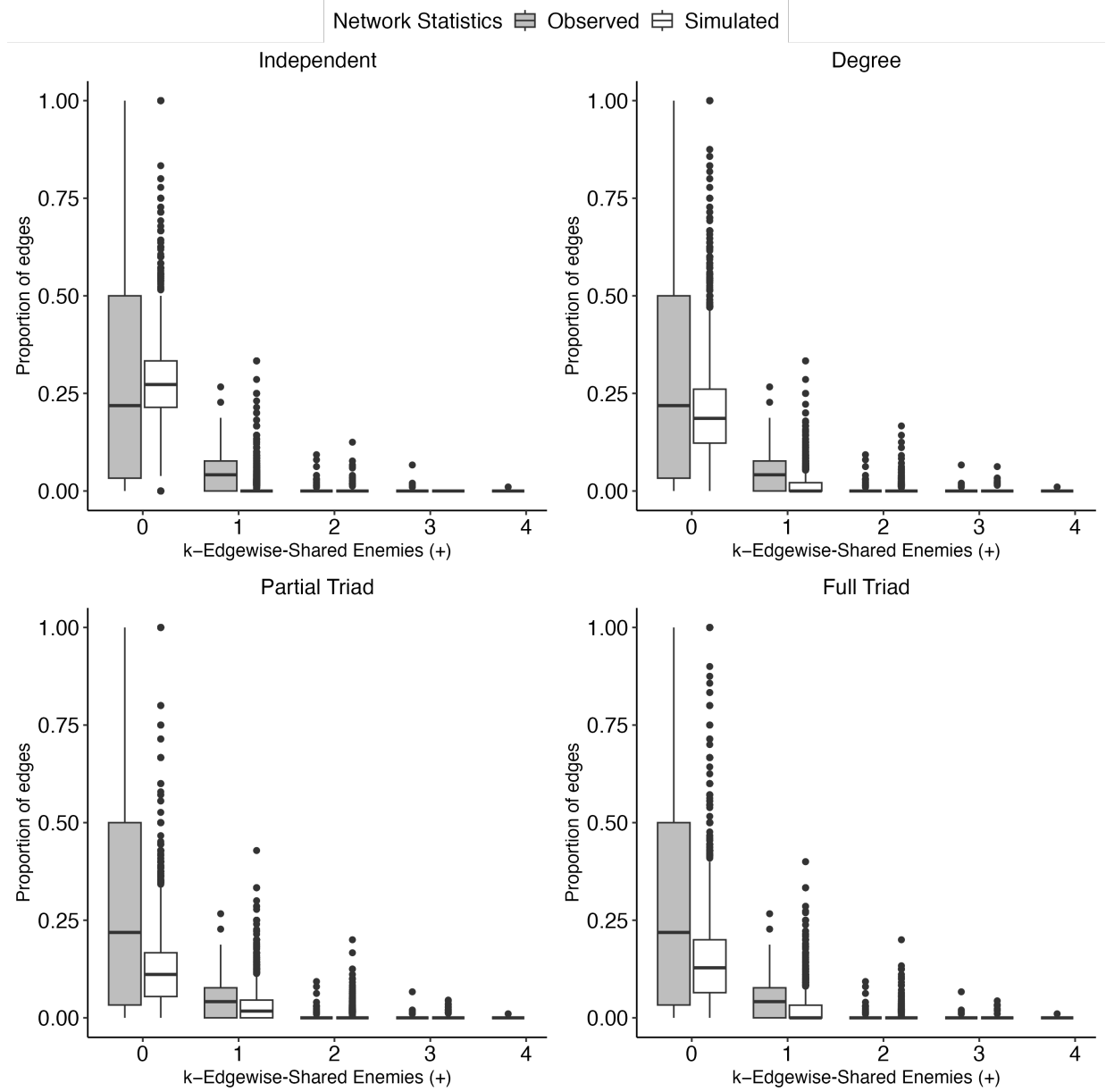


Figure 6: Comparison of out-of-sample cross-validation results for positive edgewise shared enemies (ESE +). The distribution of simulated statistics across 100 replications is compared against the observed statistics for each block. Models compared are: I (Independent), I+D (Degree), I+D+PT (Partial Triad), and I+D+FT (Full Triad).

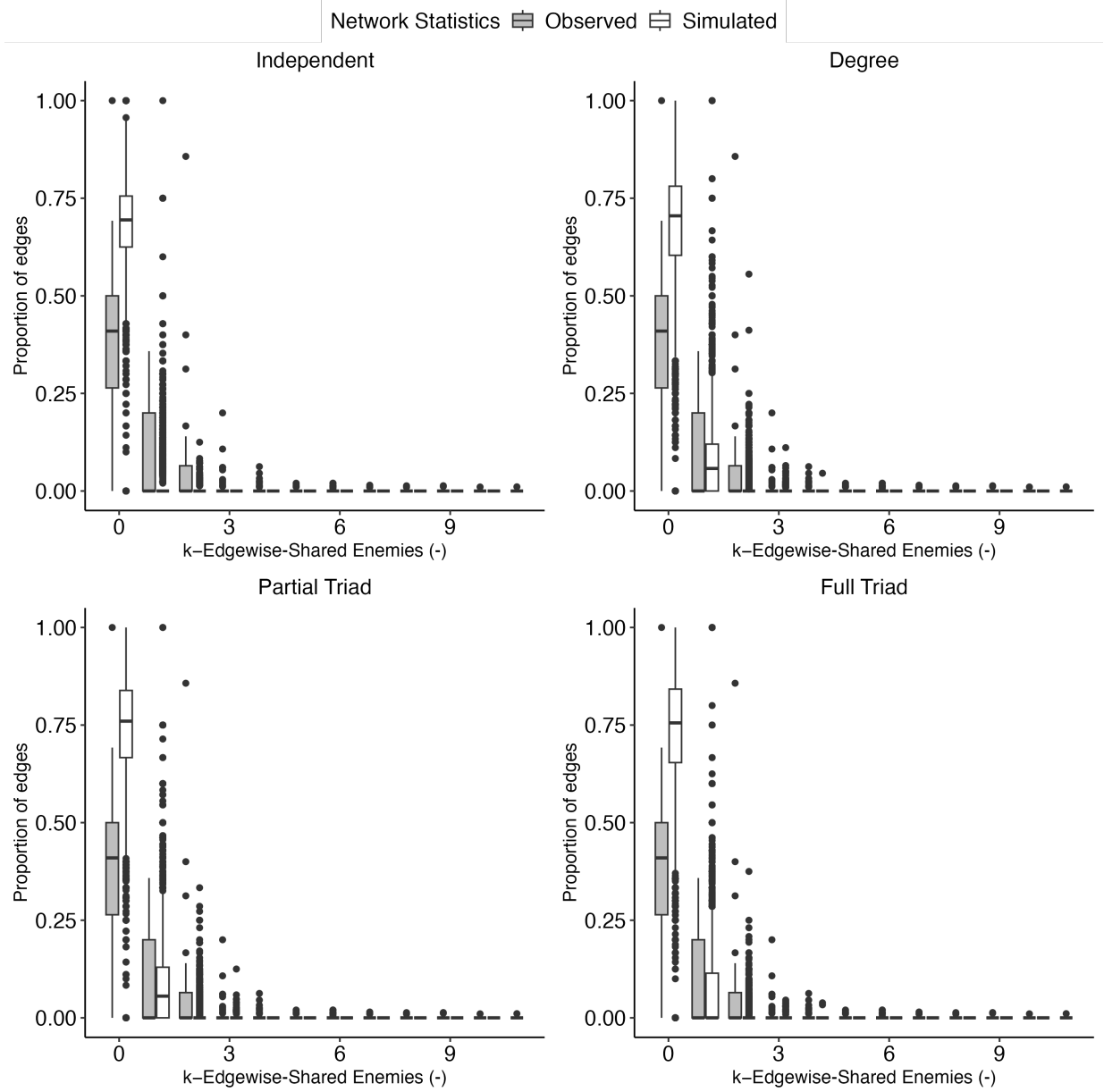


Figure 7: Comparison of out-of-sample cross-validation results for negative edgewise shared enemies (ESE  $-$ ). The distribution of simulated statistics across 100 replications is compared against the observed statistics for each block. Models compared are: I (Independent), I+D (Degree), I+D+PT (Partial Triad), and I+D+FT (Full Triad).

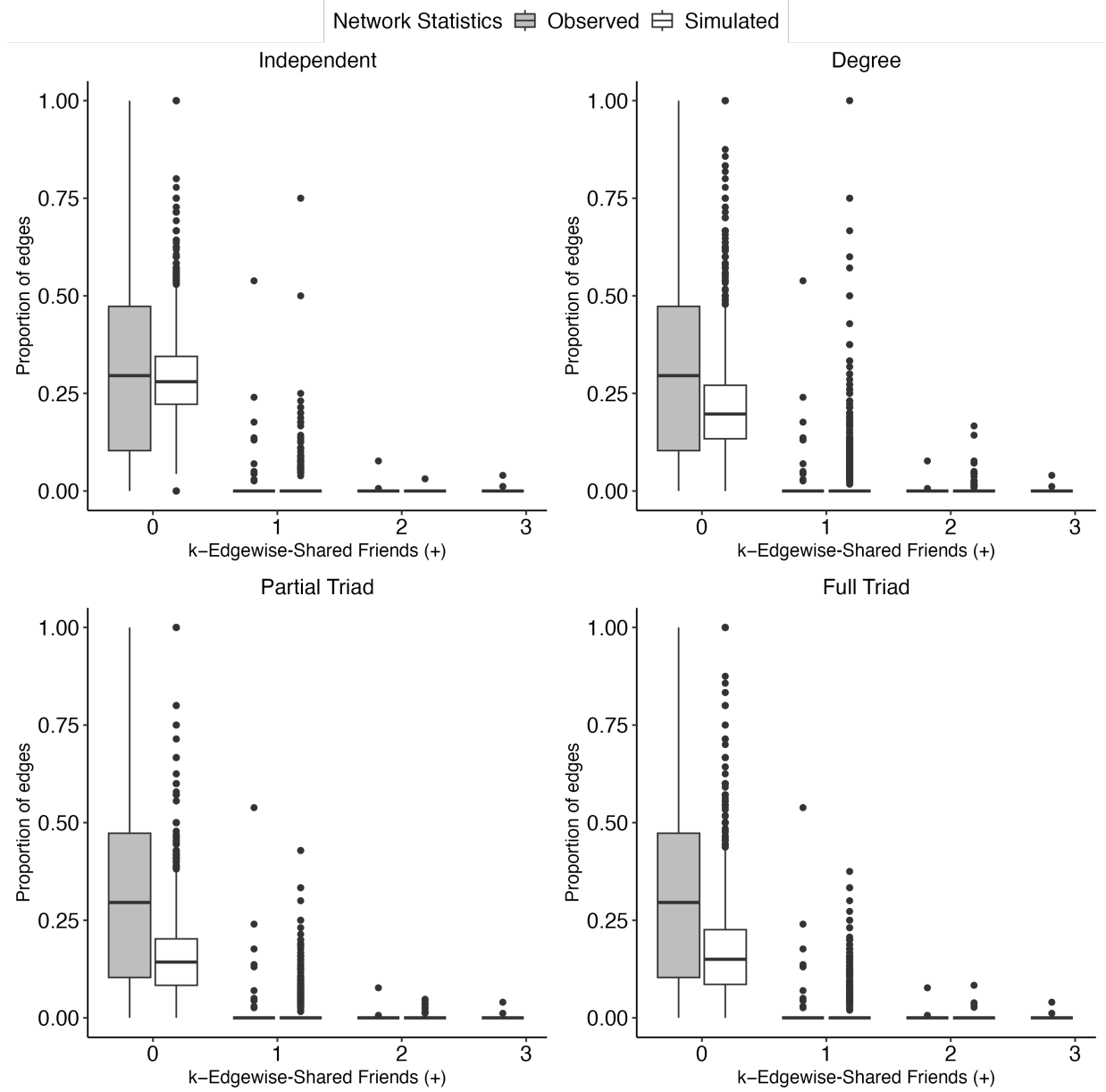


Figure 8: Comparison of out-of-sample cross-validation results for positive edgewise shared friends (ESF +). The distribution of simulated statistics across 100 replications is compared against the observed statistics for each block. Models compared are: I (Independent), I+D (Degree), I+D+PT (Partial Triad), and I+D+FT (Full Triad).

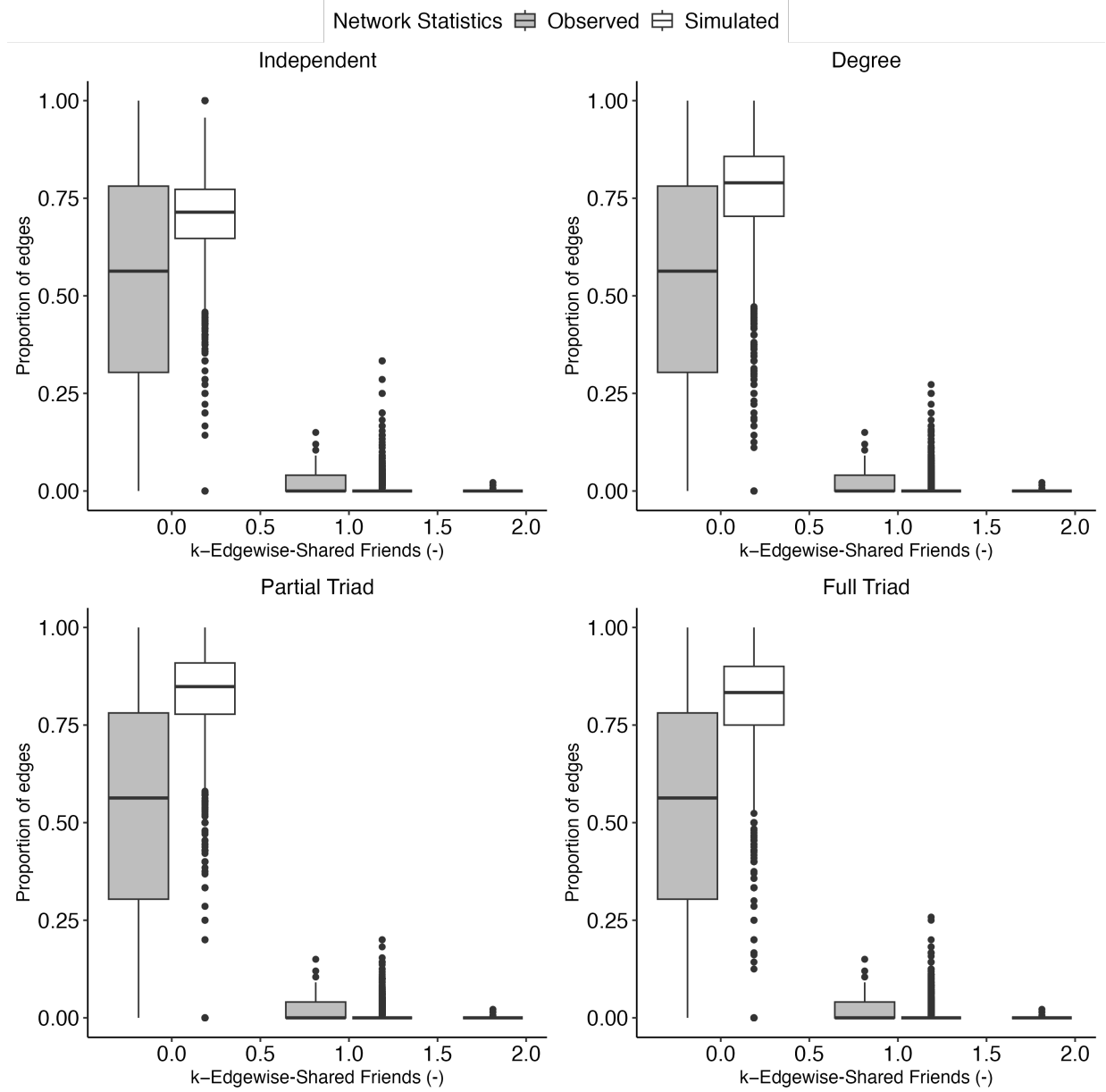


Figure 9: Comparison of out-of-sample cross-validation results for negative edgewise shared friends (ESF -). The distribution of simulated statistics across 100 replications is compared against the observed statistics for each block. Models compared are: I (Independent), I+D (Degree), I+D+PT (Partial Triad), and I+D+FT (Full Triad).



## C.2 Goodness-of-Fit

To validate the results in Table 2, we conduct a conventional ERGM goodness-of-fit analysis following the method outlined by Hunter et al. (2008). Using the estimated coefficients, we simulate 500 networks and compare the simulated network statistics with the observed statistics. For the network statistics describing the signed network, we use the same metrics as in Section 6.1. The red line marks the observed network’s statistic and should ideally be near the median of the simulated values, shown by the center of the boxplots.

The results show that while the sparse, low-degree nature of the Wikipedia network is captured by the simulations, the statistics distribution of the observed network is not replicated with high accuracy. Overall, the in-sample goodness-of-fit analysis aligns with the out-of-sample analysis, indicating that the more complex models, Partial Triad and Full Triad, best describe the data.

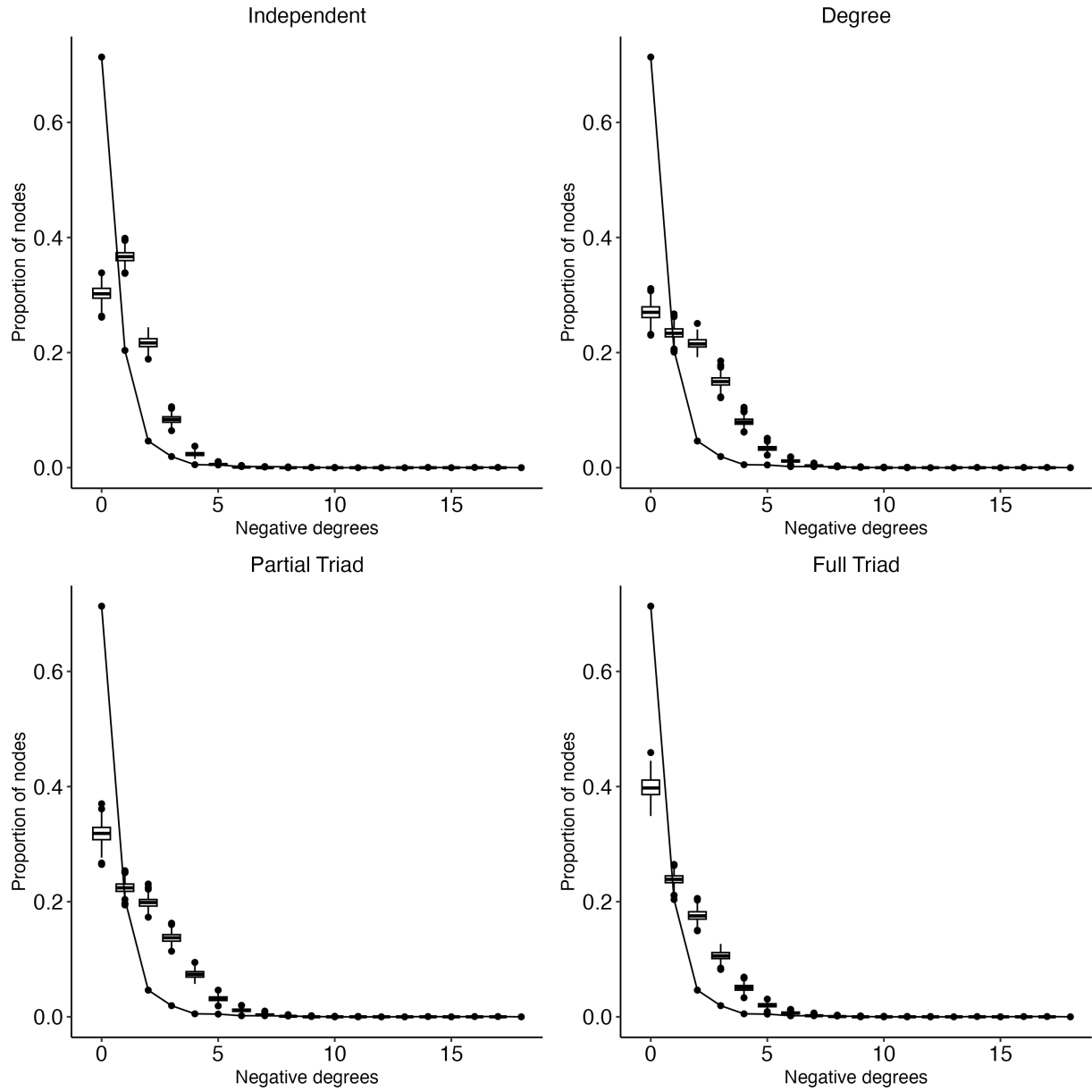


Figure 10: Comparison of goodness-of-fit results for positive degree distribution. Observed network statistics (line) are compared to the distribution of statistics from 500 simulated networks. Models compared are: I (Independent), I+D (Degree), I+D+PT (Partial Triad), and I+D+FT (Full Triad).

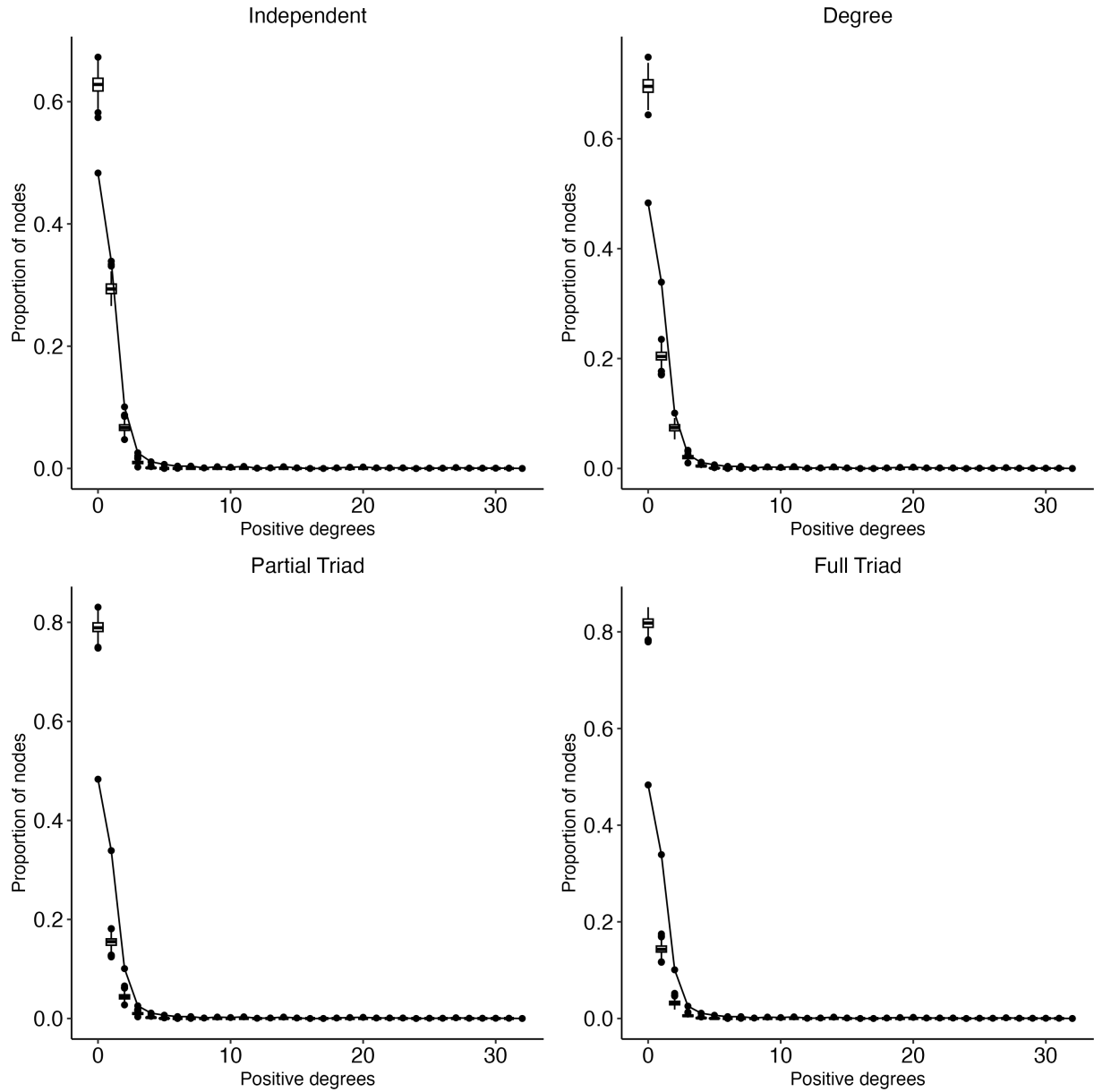


Figure 11: Comparison of goodness-of-fit results for negative degree distribution. Observed network statistics are compared to the distribution of statistics from 500 simulated networks. Models compared are: I (Independent), I+D (Degree), I+D+PT (Partial Triad), and I+D+FT (Full Triad).

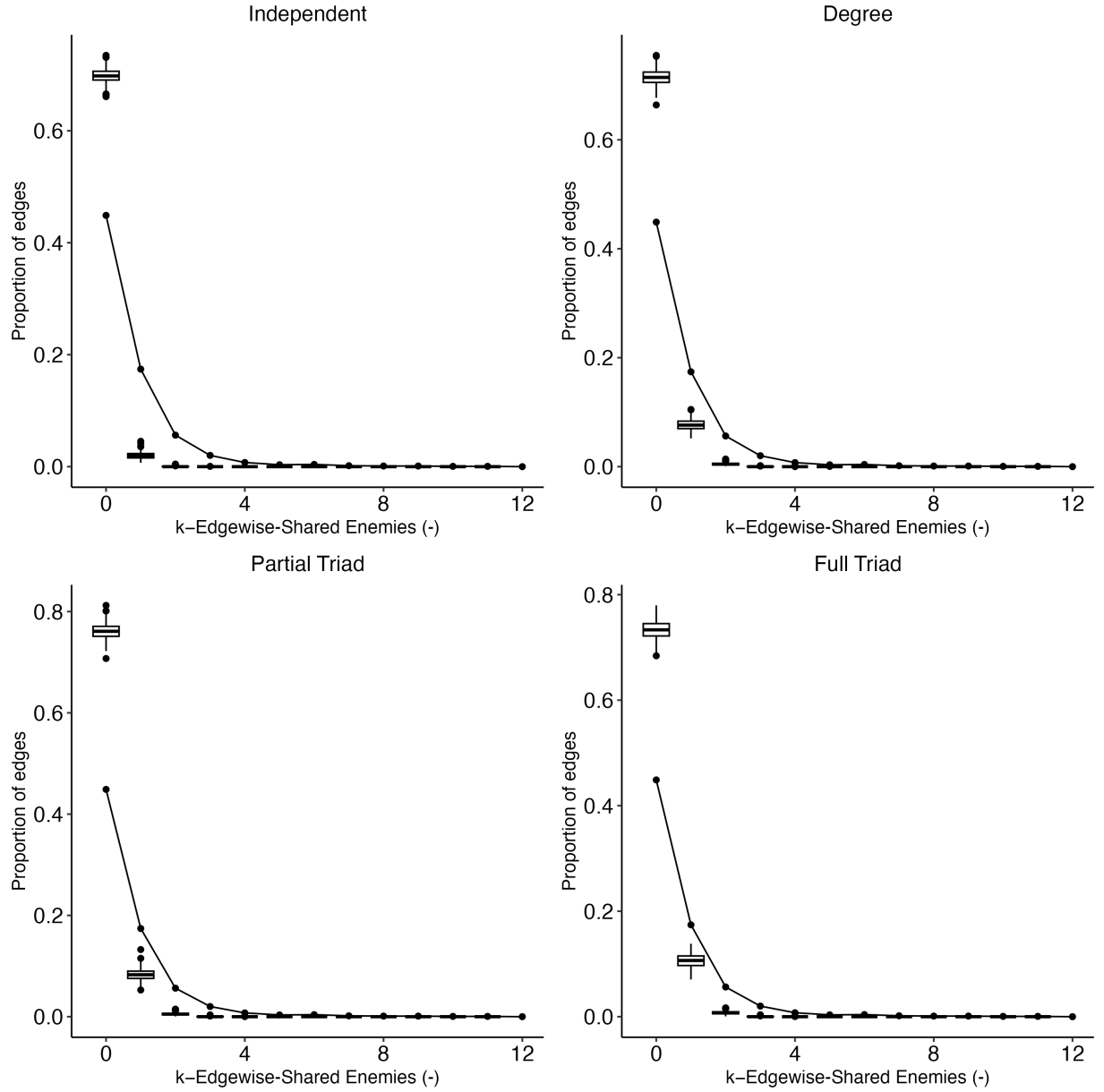


Figure 12: Comparison of goodness-of-fit results for positive edgewise shared enemies (ESE +). Observed network statistics are compared to the distribution of statistics from 500 simulated networks. Models compared are: I (Independent), I+D (Degree), I+D+PT (Partial Triad), and I+D+FT (Full Triad).

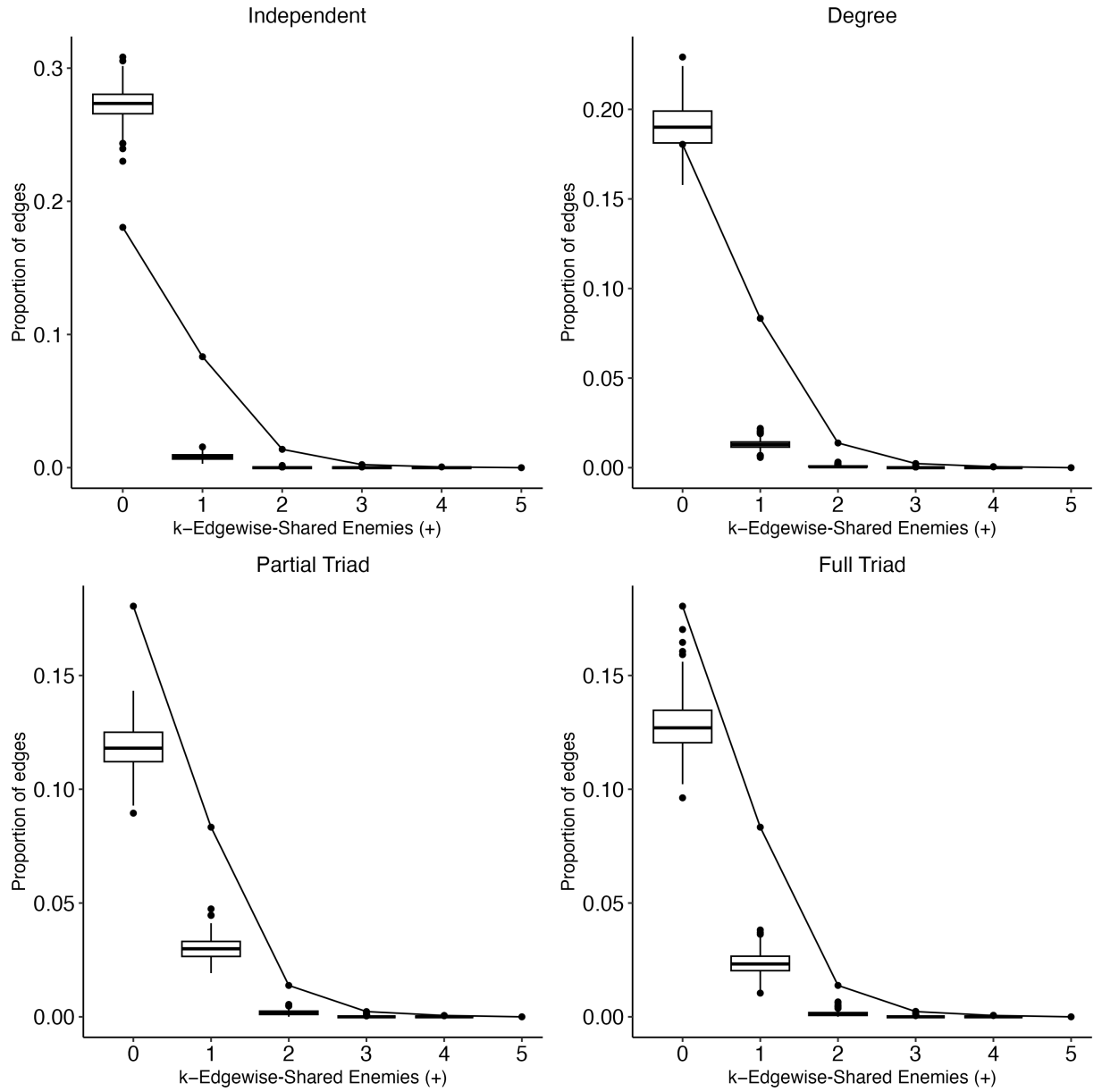


Figure 13: Comparison of goodness-of-fit results for negative edgewise shared enemies (ESE  $-$ ). Observed network statistics are compared to the distribution of statistics from 500 simulated networks. Models compared are: I (Independent), I+D (Degree), I+D+PT (Partial Triad), and I+D+FT (Full Triad).

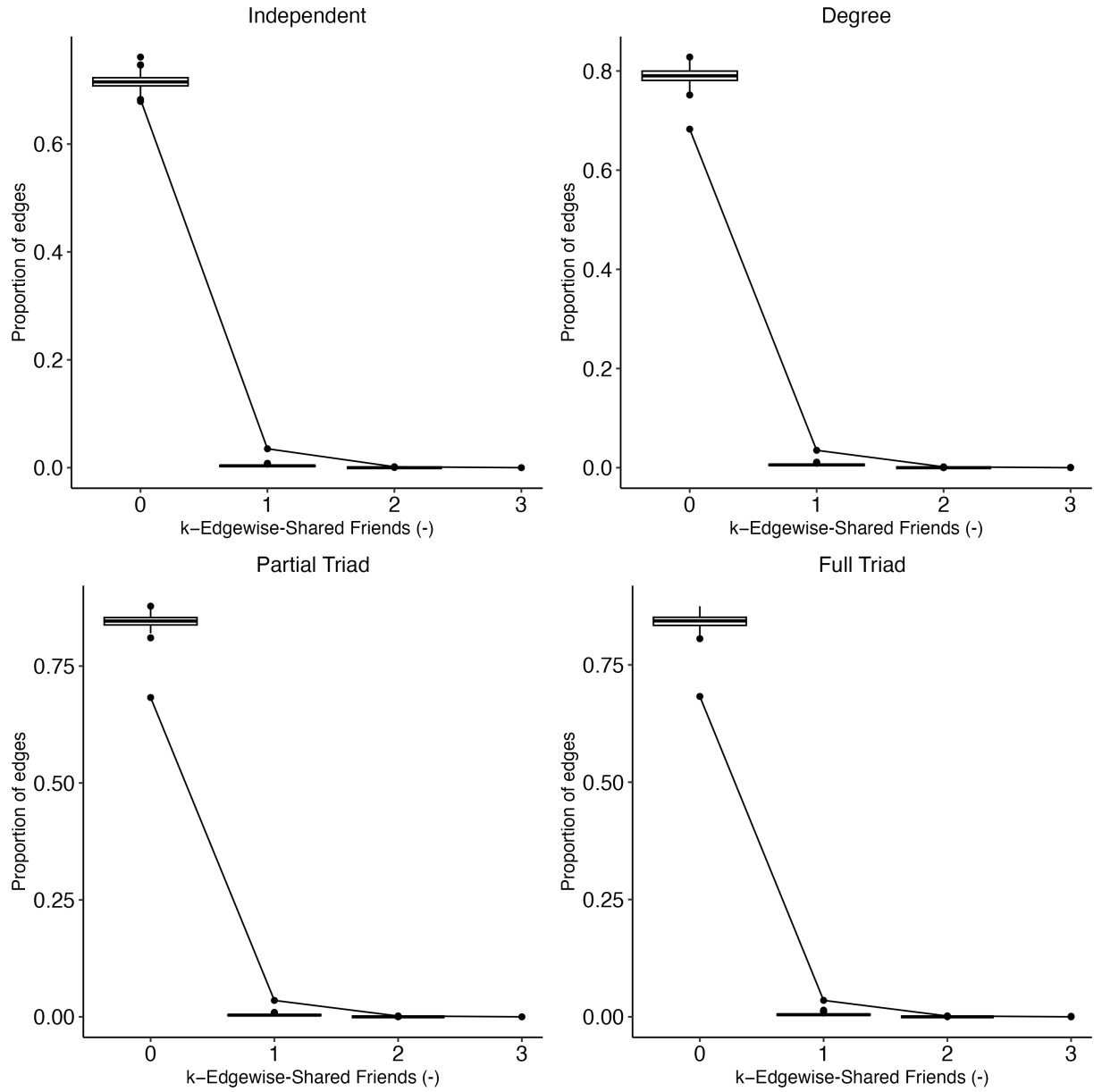


Figure 14: Comparison of goodness-of-fit results for positive edgewise shared friends (ESF +). Observed network statistics are compared to the distribution of statistics from 500 simulated networks. Models compared are: I (Independent), I+D (Degree), I+D+PT (Partial Triad), and I+D+FT (Full Triad).

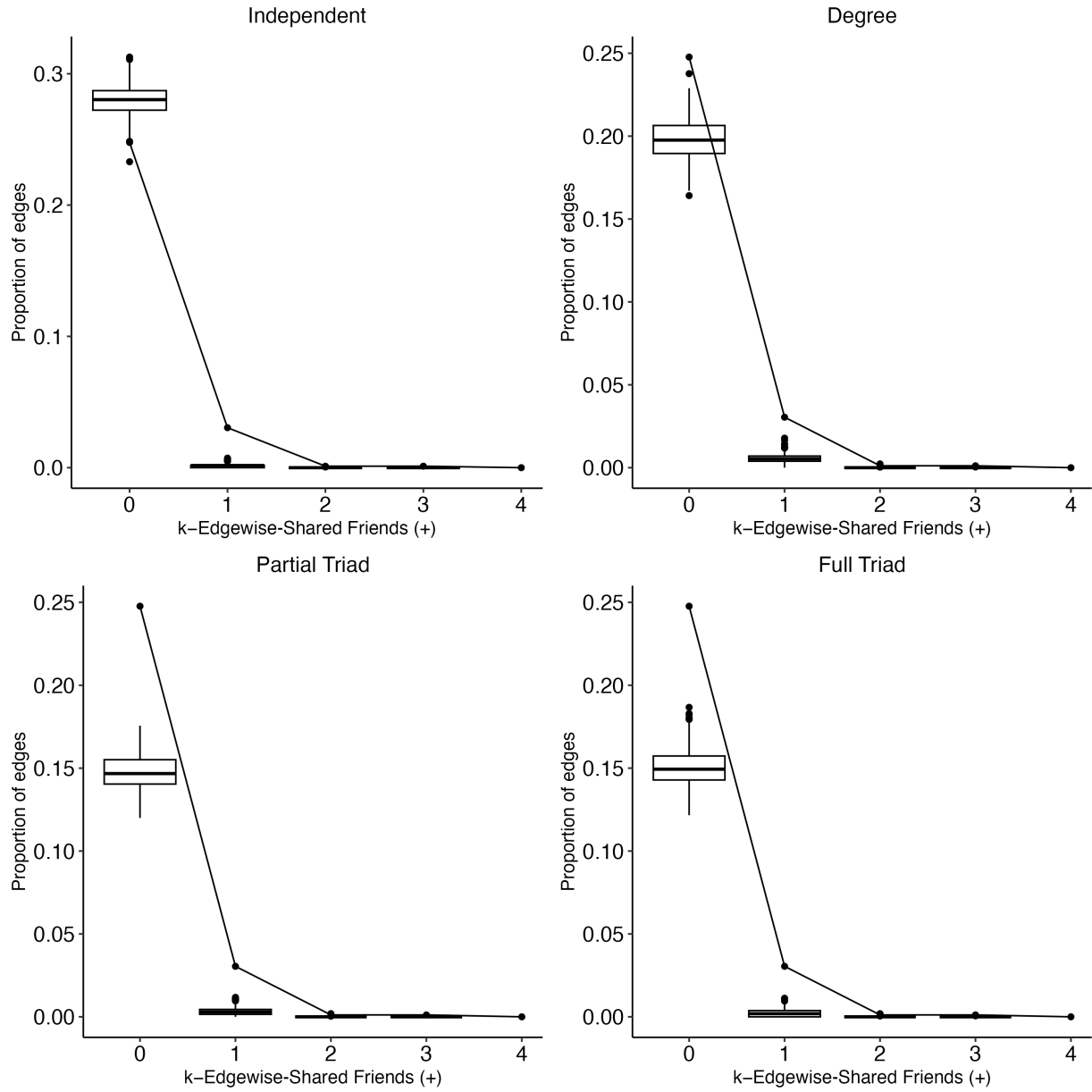


Figure 15: Comparison of goodness-of-fit results for negative edgewise shared friends (ESF  $-$ ). Observed network statistics are compared to the distribution of statistics from 500 simulated networks. Models compared are: I (Independent), I+D (Degree), I+D+PT (Partial Triad), and I+D+FT (Full Triad).

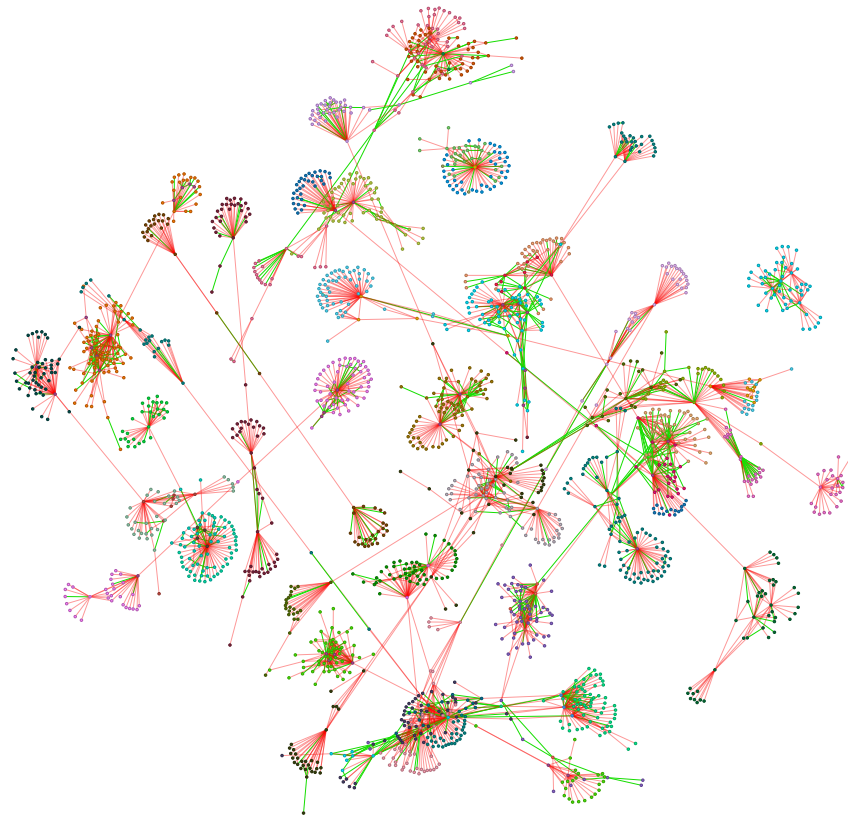


Figure 16: Visualization of the Wikipedia editor network based on selected pages.