

Direct blockmodeling of valued and binary networks: a dichotomization-free approach

Carl Nordlund

The Institute for Analytical Sociology, Linköping University, S-601 74 Norrköping, Sweden

ARTICLE INFO

Keywords:

Blockmodeling
Valued networks
Goodness-of-fit functions
Weighted correlation coefficient
Dichotomization

ABSTRACT

A long-standing open problem with direct blockmodeling is that it is explicitly intended for binary, not valued, networks. The underlying dilemma is how empirical valued blocks can be compared with ideal binary blocks, an intrinsic problem in the direct approach where partitions are solely determined through such comparisons. Addressing this dilemma, valued networks have either been dichotomized into binary versions, or novel types of ideal valued blocks have been introduced. Both these workarounds are problematic in terms of interpretability, unwanted data reduction, and the often arbitrary setting of model parameters.

This paper proposes a direct blockmodeling approach that effectively bypasses the dilemma with blockmodeling of valued networks. By introducing an adaptive weighted correlation-based criteria function, the proposed approach is directly applicable to both binary and valued networks, without any form of dichotomization or transformation of the valued (or binary) data at any point in the analysis, while still using the conventional set of ideal binary blocks from structural, regular and generalized blockmodeling.

The proposed approach seemingly solves two other open problems with direct blockmodeling. First, its standardized goodness-of-fit measure allows for direct comparisons across solutions, within and between networks of different sizes, value types, and notions of equivalence. Secondly, through an inherent bias of point-biserial correlations, the approach puts a premium on solutions that are closer to the mid-point density of blockmodels. This, it is argued, translates into solutions that are more intuitive and easier to interpret.

The approach is demonstrated by structural, regular and generalized blockmodeling applications of six classical binary and valued networks. Finding feasible and intuitive optimal solutions in both the binary and valued examples, the approach is proposed not only as a practical, dichotomization-free heuristic for blockmodeling of valued networks but also, through its additional benefits, as an alternative to the conventional direct approach to blockmodeling.

1. Introduction

Blockmodeling¹ is concerned with the identification and classification of actors that are equivalent in some meaningful sense, and the analysis of the patterns of ties within and between such sets of equivalent actors. Given a network represented as an adjacency matrix, a blockmodel is created by sorting the rows and columns according to these sets of equivalent actors (also known as ‘positions’ or ‘clusters’), where the delineated sub-matrix ‘blocks’² within and between such positions are compared to different sets of ideal blocks. For an adequately fitted blockmodel, i.e. where the empirical blocks can be reasonably approximated in terms of such ideal blocks, the blockmodel can

be reduced to a so-called blockimage, a representation of the original network that capture its underlying ‘functional anatomy’ (Nordlund, 2016, p. 161). Such blockimages can subsequently be compared across different networks, possibly of vastly different sizes, as well as interpreted and compared with the set of ideal blockimage templates that capture common archetypical structures, such as core-periphery structures, hierarchies, transitive structures, and cohesive subgroups (see Doreian et al., 2005, p. 236ff; Wasserman & Faust, 1994, p. 419,423; White et al., 1976, pp. 742–744).

The notion of equivalence is foundational in blockmodeling, where different ideas and definitions of what is meant by actor equivalence determine the outcome and interpretation of a blockmodel analysis.

E-mail addresses: carl.nordlund@liu.se, carl.nordlund71@gmail.com.

¹ For a more detailed introduction to blockmodeling and role-analysis, see Wasserman and Faust (1994) and Doreian et al. (2005). In addition to the deterministic kind of blockmodeling that this article is concerned with, there is also stochastic blockmodeling – see Wasserman and Faust (1994, Chapter 16), a chapter that also contains a section of goodness-of-fit measures of relevance for this current article.

² The term ‘block’ is here used to denote the specific submatrix in an adjacency matrix that is specified by one or two sets of actors.

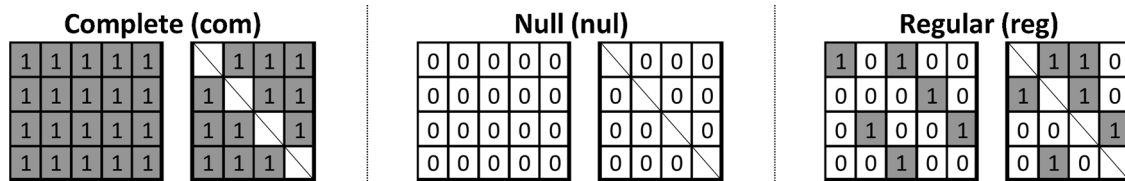


Fig. 1. Examples of complete, null and regular blocks located off or on the diagonal of a blockmodel.

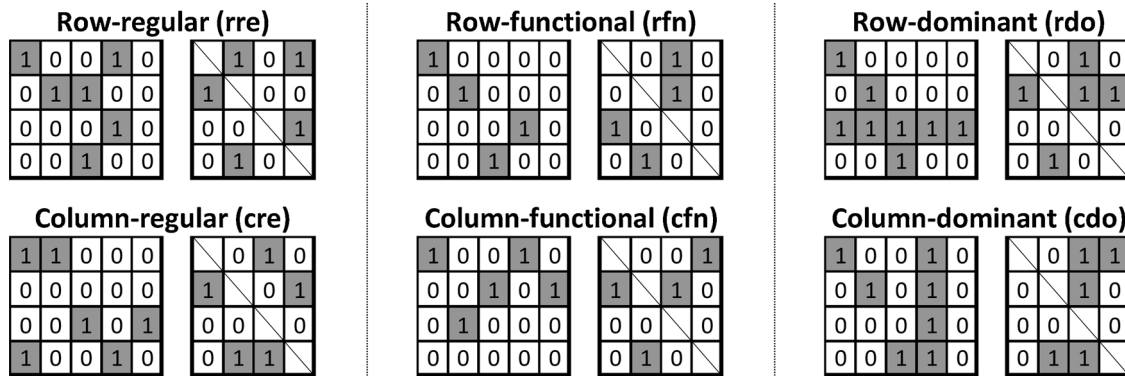


Fig. 2. Examples of ideal block types associated with generalized blockmodeling (see Nordlund 2016:162; Doreian et al., 2005:212; Žiberna 2007a:108).

The initial notion of equivalence was that of *structural* equivalence: formally introduced by Lorrain and White (1971), two actors are deemed structurally equivalent if they have identical ties to their common alters. Under this notion of equivalence, the empirical blocks in a structural equivalence blockmodel should ideally be either filled with ties – a so-called complete block – or void of any ties whatsoever – which constitute a null block. Loosening the criteria for structural equivalence, White and Reitz (1983, 1985; building on Sailer, 1978) proposed the notion of *regular* equivalence. Two actors are deemed regularly equivalent if they have ties to alters that in turn are also regularly equivalent. In regular equivalence blockmodeling, this kind of equivalence is captured by the regular block, an ideal block type where there is at least one tie in each row and column of the block. Supplementing the regular block, classical regular blockmodeling also includes the null block. Examples of complete, null and regular blocks are given in Fig. 1, for both off-diagonal blocks (i.e. with ties between two sets of actors) and diagonal blocks (i.e. for ties among a single set of actors).

Generalized blockmodeling is a broad analytical framework that allows for a wider variety of equivalences. Incorporating the three ideal blocks used to capture structural and regular equivalence, generalized blockmodeling also introduces additional types of ideal blocks (see Doreian et al., 2005, p. 212). Separating the criteria of regular blocks into rows and columns, respectively, the row-regular ideal block has at least one tie in each row of the block, and a column-regular has at least one tie in each block column. For row-functional blocks, there is exactly one tie in each row of the block (with corresponding column-wise criteria for column-functional blocks). A third novel type are the dominant-directional blocks, where the ideal row-dominant (column-dominant) block implies that at least one row (column) in the block is filled with ties. Examples of the directional varieties of the regular, functional and dominant block types are given in Fig. 2, for off- and on-diagonal blockmodel locations.

The procedures to identify and classify actors into equivalence³ sets

can be divided into indirect and direct approaches. In the indirect approaches, pair-wise proxy measures for a given type of equivalence are first calculated⁴ for each pair of actors in the network (see, e.g., Doreian et al., 2005, p. 177ff). Through a cluster analysis of these indirect measures of equivalence, identified classes of near-equivalent actors are subsequently used to sort the original adjacency matrix into a blockmodel. The emerging blocks are finally interpreted by comparing them with the set of ideal blocks of relevance for the specific notion of equivalence that the indirect measure is argued to capture. Contrasting this, the direct approach instead tries to fit a given set of ideal blocks directly with the original adjacency matrix. As an optimization problem where the number of positions is given a priori, a suitable relocation algorithm is used to find the particular actor partitions that minimizes the number of inconsistencies between the observed empirical blocks and a given set of ideal blocks that represents a specific type of equivalence. Although computationally heavy for larger networks and with the risk of search algorithms possibly failing to find optimal solutions, direct approaches nevertheless have several advantages over the indirect approaches (see Doreian et al., 2005, p. 184; Doreian, 2006, p. 124). The most profound and, in this case, crucial advantage of direct approaches is the lack of indirect measures or algorithms that can capture the more complex and varied ideal block types as found in generalized blockmodeling. In what follows, the focus is thus exclusively on direct blockmodeling approaches.

1.1. The dilemma with blockmodeling of valued networks

An open and seemingly sticky problem in blockmodeling is that it is explicitly geared at binary, rather than valued, networks (Doreian, 2006; Nordlund, 2007, 2016; Žiberna, 2007b). The underlying dilemma is that the standard set of ideal blocks (Fig. 1 and Fig. 2) are all specified

³ Or, more commonly, near-equivalent. It is rare that actors in real-world empirical networks are perfectly equivalent, meaning that classes/subsets of actors typically contain some variance with respect to how structural, regular or otherwise equivalent they are.

⁴ The classical indirect measure of structural equivalence of two actors involves the comparison of their corresponding row- and column vectors in the adjacency matrix, often using either Pearson correlations, Euclidean distances or Hamming distances to compare these vectors. For regular equivalence, White and Reitz (1983, 1985; also see varieties by Žiberna, 2007b) introduced the REGE algorithm, a point-scoring algorithm that, it is argued, produces adequate pair-wise measures of regular equivalence of actors.

in terms of binary tie patterns, making it a non-trivial exercise to match and count inconsistencies between such ideal binary ties and empirical valued ties. This is of particular concern in the direct approaches where the identification and clustering of equivalent actors stem directly from such comparisons, but it is equally problematic when interpreting blocks for partitions derived from indirect methods or for given, pre-specified hypothetical partitions.

Addressing this dilemma, a number of workarounds have been proposed. The classical yet obviously crude approach is to simply transform the network into a binary version through dichotomization. Instead of picking a single network-wide dichotomization threshold on statistical, theoretical or arbitrary grounds, an alternative is to instead apply a series of threshold values at periodic intervals, repeating and comparing the blockmodeling analyses for a larger ensemble of dichotomized binary networks (e.g. Doreian, 1969).

The opposite type of workaround is to instead introduce novel types of ideal blocks that cater for valued ties, such as the sum- and max-regular blocks proposed by Žiberna (2007a). In these novel ideal blocks, designed explicitly for valued networks, a threshold parameter is nevertheless needed to distinguish between the fulfillment or not of the specific block criteria. In a similar vein, Nordlund (2007) suggests a novel heuristic for the identification of regular blocks in valued networks, a heuristic that equally requires the choosing of a suitable threshold to distinguish between prominent and non-prominent ties.

1.2. A novel direct approach to blockmodeling of binary and valued networks

This paper proposes a direct blockmodeling approach that effectively bypasses the dilemma with blockmodeling of valued networks. Inspired by the core-periphery heuristic of Borgatti and Everett (2000), the herein proposed approach replaces the classical inconsistency-counting criteria functions in direct blockmodeling with a generalized version of the weighted correlation coefficient approach suggested in Nordlund (2018). Applicable in both exploratory and confirmatory studies of both structural, regular and generalized blockmodeling, possibly with prespecified partitions, blockimages, or both, it is argued that the proposed approach not only represents a practical dichotomization-free solution to the valued blockmodeling problem but also provides several additional advantages compared to existing direct blockmodeling approaches, for binary and valued networks alike.

First, the proposed approach is equally applicable to binary as well as valued networks, without any type of dichotomization, transformation or reduction of the data resolution of the valued network data at any stage in the analysis. Despite this, secondly, the proposed approach still uses the conventional set of binary ideal blocks of structural, regular and generalized blockmodeling, without modifying or introducing any specialized valued versions of ideal blocks (cf. Žiberna, 2007a). In line with the generalized blockmodeling framework (Doreian et al., 2005, p. 245ff), the proposed approach also allows for the introduction of additional ideal blocks, specified in terms of binary tie patterns, including ‘overlaid’ blocks like the combined dependency-dominance blocks as suggested in Nordlund (2018).

With the weighted correlation coefficient as the criteria function, the relative contribution of each empirical block to the final goodness-of-fit measure can be kept proportional to the relative block sizes of a blockmodel, irrespective of the type of ideal block that is tested for. For regular blocks, this bypasses the need to have different ways to measure block inconsistencies depending on whether we are conducting a classical regular equivalence blockmodeling or working with the broader range of blocks in generalized blockmodeling (see Doreian et al., 2005, pp. 188, 224). By following this principle of size-proportional influence to the overall goodness-of-fit measure, the need to balance the penalties for different types of ideal blocks is greatly diminished.

Whereas “the magnitudes of the [conventional inconsistency-counting] criterion functions cannot be compared across different types

of blockmodels” (Doreian, 2006, p. 127), it is argued here that a standardized correlation-based measure makes such comparisons possible. It thus also provides a solution to this hitherto unsolved problem of comparing goodness-of-fit measures across different blockmodel solutions, with possibly vastly different network sizes and types of equivalences, for both binary and valued networks.

Two additional advantages stem from using a correlation-based goodness-of-fit function for direct blockmodeling. The first is the inherent bias of point-biserial correlations (i.e. a correlation when a naturally dichotomous variable is correlated with a continuous variable) that favors solutions where a split of the data along the dichotomous variable results in equally large subsets. Although this is occasionally perceived as a problem in various applications, it is here argued⁵ to be an advantage in this context, as it puts a premium on ideal blockmodels that are closer to the mid-point (0.5) density than further away from it. Stemming from this, the second advantage with using a correlation-based heuristic for direct blockmodeling is that it is less likely to find multiple, equally optimal solutions than what is the case for the conventional inconsistency-minimizing approach. Although this makes it imperative to also explore near-optimal solution, possibly classifying near-optimal solutions as equally optimal as the max-correlation solution, the correlation-based measure can act as a guide in choosing which out of many near-optimal solutions to use as a representative blockmodel.

The final crucial advantage with the approach proposed here is that for the binary networks so far tested, it seems to capture either identical or very similar solutions as those found when using the existing conventional approach for direct binary blockmodeling. As there are other advantages with this proposed approach, even when analyzing binary empirical networks, the weighted correlation coefficient heuristic is not only proposed as a premier solution to direct blockmodeling of valued networks, it is also proposed as an alternative to the conventional direct approach to binary network blockmodeling.

The remainder of this article is divided into three parts. Starting with the Borgatti-Everett core-periphery heuristic and its power-relational extensions proposed by Nordlund (2018), the next section will discuss the foundations for the proposed heuristic, providing pseudocode for how various ideal binary blocks can be captured. This is followed by an example section. Beginning with a small toy network, two classical binary networks will be analyzed, followed by three valued networks: the Hlebec notesharing data (Hlebec, 1996; Žiberna, 2007a), the UCINET river primates network (Borgatti & Everett, 2000, p. 380), and the Freeman’s EIES friendship networks (Freeman & Freeman, 1980). The paper is concluded with a short summary of the proposed heuristics, observations from the provided examples, and potential shortcomings.

2. Specification of proposed heuristic

The direct approach for binary and valued generalized blockmodeling proposed here takes inspiration from the core-periphery measure and heuristic proposed by Borgatti and Everett (2000) and the extensions to this approach as proposed by Nordlund (2018). Starting with these correlation-based approaches for identifying core-periphery structures, this section describes how these can be generalized and extended. This is followed by specifications on how the ideal blocks in structural, regular and generalized⁶ equivalence can be identified, measured and combined into the aggregate goodness-of-fit measure proposed in this paper.

⁵ This will be discussed and detailed in the next section of the paper.

⁶ Of the six types of ideal blocks typically specified for generalized blockmodeling – the row- and column-varieties of, respectively, regular, functional and dominant ideal blocks (see Fig. 2 above) – this paper will specify and provide pseudocode for the regular and functional types, leaving the row- and column-dominant ideal blocks for would-be future implementations.

		C			P		
		a	b	c	d	e	f
C	a	1	1	0	1	1	
	b	0	1	0	0	0	
	c	1	1	0	0	1	
P	d	0	1	1	0	0	
	e	1	0	0	0	1	
	f	0	0	1	0	1	

Fig. 3. 6-actor directional binary network example for core-periphery correlation testing.

2.1. The Borgatti-Everett measure of core-periphery structures

To capture the classical categorical core-periphery notion of dense cores and internally disconnected peripheries for a hypothetical 2-positional blockmodel, Borgatti and Everett (2000) suggest using the (population-wise) correlation coefficient resulting from correlating all intra-core cells with unity and corresponding intra-peripheral cells with zero. This measure thus reaches unity when the set of core actors form a clique⁷ and when there are no ties between the remaining set of peripheral actors. Borgatti and Everett demonstrate how this correlation can be used as a goodness-of-fit measure in a relocation algorithm for finding optimal core-periphery partitions. Exemplifying the measure using the hypothetically partitioned 6-actor directional binary network in Fig. 3 below, the calculation of the core-periphery correlation for this particular blockmodel is given in Table 1.

By fitting an empirical network to the prespecified complete and null blocks, the Borgatti-Everett heuristic for identifying core-periphery structures is a direct approach for structural equivalence blockmodeling. What differs from conventional direct blockmodeling is its criteria function: the classical procedure of counting and minimizing the total number of inconsistencies (i.e. the Hamming distance between observed and ideal value-pairs) is in the Borgatti-Everett approach replaced by a correlation⁸ measure. As exemplified in Table 1, the complete and null block tests are each represented by subsets of value-pairs capturing the empirical and ideal values for each cell, which when combined form the whole set of observed and ideal ties for the overall measure of fit.

While the ideal ties for a core-periphery structure indeed are binary (Y in Table 1), the observed values (X in Table 1) could be either binary or valued (Borgatti & Everett, 2000, p. 379). This heuristic for measuring and fitting core-periphery structures is thus applicable to both binary and valued networks, without any form of dichotomization at any point in the analysis.

As the values in Y are always binary, this correlation is also known as the point-biserial correlation, expressed in the following formula:

$$r_{pb} = \frac{M_1 - M_0}{s_n} \sqrt{\frac{n_1 n_0}{n^2}} \quad (1)$$

where M_1 and M_0 are the mean values for the subsets of X values with corresponding Y values being, respectively, 1 and 0, s_n being the

⁷ I.e. where all possible ties exist and where they have equal tie values (i.e. unity in binary networks).

⁸ A correlation coefficient of two binary variables is also called the phi (ϕ) coefficient or, as in Hubert and Baker (1978), the Γ index. As the correlation coefficient is 'designed' to measure the linear association between two continuous variables, its applicability for dichotomous data has been criticized (see also Carrington et al., 1979, p. 223ff). Borgatti and Everett however point out that they are merely using the Pearson correlation coefficient as a measure of association between two variables, without the associated inferential test (2000, p. 379; cf. Arable et al., 1978). This is exactly the setting in which this current paper and its proposed heuristic operate.

population-wide standard deviation of all X values, n_1 and n_0 being the number of Y values that are, respectively, 1 and 0, and n being the total length of the list L .

The point-biserial representation (Eq. 1) is particularly instructive for understanding an important aspect of its usage as a blockmodeling criteria function, such as the Borgatti-Everett core-periphery measure. Dividing up the values in the X vector into two sets, one for respective Y value, the left-side factor of the point-biserial formula is concerned with the differences in mean values between these sets. The right-hand factor in the formula, however, is only concerned with the relative sizes of these two sets, a factor that reaches a maximum of 0.5 when n_1 equals n_0 . As a measure of goodness-of-fit in the Borgatti-Everett core-periphery heuristic and, as elaborated below, its generalization as a criteria function for direct structural equivalence, this preference for an equal number of 1:s and 0:s in the Y vector translates into a bias towards optimal blockmodel solutions whose ideal counterparts are closer to a mid-way density. Whereas this particular 50/50 phenomena of point-biserial correlations has been deemed problematic and worthy of 'corrections' in certain context (see, e.g., Kemery et al., 1988; cf. Williams, 1990), it is here argued to instead be beneficial in the blockmodeling context by its inherent 'bias', or rather preference, for more interpretable solutions.

The first step towards the approach suggested in this paper is a generalization of the Borgatti-Everett heuristic for identifying core-periphery structures. By allowing for arbitrary configurations of complete and null blocks, for two or more sets of equivalent actors, we obtain a direct approach for structural equivalence blockmodeling that, similar to the original core-periphery heuristic, is equally applicable to binary and valued data.

Whereas the standard (point-biserial) correlation coefficient works well for capturing the complete and null ideal block types, it is less useful for capturing the more complex tie patterns of the ideal blocks used in regular and generalized blockmodeling. Rather, building on and generalizing how the complex ideal block types of dependency and dominance were operationalized in Nordlund (2018), the weighted correlation coefficient is instead used in the specification which we now turn to.

2.2. A weighted correlation-based direct blockmodeling approach⁹

Given the adjacency matrix A of a (possibly directional) network, where $a_{row,col}$ is the tie value from actor row to actor col (with self-ties undefined), the set of actors are partitioned into k non-overlapping subsets $P = \{P_1, \dots, P_k\}$. The blockmodel is given by $B(A, P)$, where $B_{r,c}$ corresponds to the particular empirical block of A that is delineated by the actors in P_r and P_c . For diagonal blocks, r equals c . We also have a specified ideal blockimage I , where $I_{r,c}$ represents the ideal block to be compared with the relational patterns of block $B_{r,c}$. For structural equivalence blockmodeling, I contains either complete or null blocks.

To calculate the weighted correlation measure for a specific partition, we compile a list L of ordered value-triplets ($x; y; w$), values that correspond to, respectively, an observed value (x), an ideal value (y), and a weight (w) for each ($x; y$) value-pair. Given the blockmodel B and blockimage I , the subset of value-triplets for the specific empirical and ideal block at row r and column c of the blockmodel is denoted by $f_{BLK}(A, P_r, P_c)$ where BLK is the specific ideal block – $I_{r,c}$ – that is being tested. Concatenating the value-triplets for all blocks into the list L , the goodness-of-fit for this blockmodel and partition is the weighted correlation coefficient r_w where each value-pair in the combined X and Y vectors uses the corresponding weights in the combined W vector:

⁹ A demonstrational Windows client that implements the proposed heuristic is available at <http://carlnordlund.net/>.

Table 1
Calculating the Borgatti-Everett core-periphery measure for the blockmodel in Fig. 3.

Block	C→C (Complete)						P→P (Null)					
Cell	a→b	a→c	b→a	b→c	c→a	c→b	d→e	d→f	e→d	e→f	f→d	f→e
Observed (X)	1	1	0	1	1	1	0	0	0	0	0	1
Ideal (Y)	1	1	1	1	1	1	0	0	0	0	0	0
	Number of inconsistencies:						corr(X, Y):					
	2						2/3					

Table 2

Pseudocode to generate correlation value-triplets for complete and null blocks.^a

$f_{\text{COMPLETE}}(A, P_i, P_j)$:	$f_{\text{NULL}}(A, P_i, P_j)$:
$L_{i,j} = \{ \}$ for each r in P_i : for each c in P_j : if $r < > c$: $L_{i,j} \leftarrow (a_{r,c}; 1; 1)$ return $L_{i,j}$	$L_{i,j} = \{ \}$ for each r in P_i : for each c in P_j : if $r < > c$: $L_{i,j} \leftarrow (a_{r,c}; 0; 1)$ return $L_{i,j}$

^a It is here assumed that the network contains no self-ties. If self-ties exist and are allowed, the if-statements have to be removed.

$$r_w = \frac{\sum w_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum w_i (x_i - \bar{x})^2} \sqrt{\sum w_i (y_i - \bar{y})^2}} \quad (2)$$

where x_i and y_i are, respectively, the observed and ideal values for value-triplet i , w_i is the weight for the $(x_i; y_i)$ pair, and where the weighted means are $\bar{x} = \sum w_i x_i / \sum w_i$ and $\bar{y} = \sum w_i y_i / \sum w_i$.

Given the empirical block $B_{i,j}$ as specified by the adjacency matrix A and the two subsets of actors P_i and P_j , the pseudocode for generating the $(x; y; w)$ value-triplets to test this empirical block with the complete and null blocks, respectively, is given in Table 2. As the criteria for these blocks imply that each cell is checked exactly once, the weighting for each value-pair is equal to unity for these blocks.

The regular block criteria stipulate that there must be *at least* one tie in each row and column, respectively, of an empirical block. It is thus only necessary to check the largest tie values¹⁰ in each row and column, respectively. For the correlation-based heuristic proposed herein, this would imply that the number of observed-vs-ideal value-pairs for checking a regular block would be far fewer than the number of value-pairs needed to check for a complete or a null block. For instance, for an off-diagonal block with 4 rows and 6 columns, only 10 observations would be needed to check for a regular block, whereas a null or a complete block would need to observe all 24 cells in this particular block. Thus, to ensure that each block contributes in a size-proportional manner to the overall goodness-of-fit measure for the blockmodel as a whole, the weights for the checked empirical-ideal value-pairs are adjusted accordingly. In the example with a 4-by-6 block, each of the 10 value-pairs would then carry a weight of 2.4, adjusting so that the relative contribution for this block to the aggregate correlation would be proportional to its size. Pseudocode to generate value-triplets for regular blocks in this way is given in Table 3 below.

Separating the criteria for regular blocks into rows and columns, respectively, subsequently also adjusting the weights for each checked value-pair, pseudocode for row-regular and column-regular blocks are found in Table 4 below. For row-regular blocks, the maximum value in each block will be correlated with unity, weighting this by the number of columns in the block (adjusted for potential self-ties). For column-regular blocks, a corresponding weighting is done with respect to the number of block rows.

For the row-functional ideal block, there is *exactly one* tie on each row in the block. In the conventional generalized blockmodeling

Table 3

Pseudocode to generate correlation value-triplets for regular blocks (Iverson square brackets, i.e. true→1, false→0; $|P_i|$: cardinal number of actors in position P_i ; $\max_x a_{x,y}$: the maximum value in column y of the A matrix)

$f_{\text{REGULAR}}(A, P_i, P_j)$:
$L_{i,j} = \{ \}$ $w = (P_i \cdot (P_j - [i = j])) / (P_i + P_j)$ for each r in P_i : $L_{i,j} \leftarrow (\max_c a_{r,c}; 1; w)$ for each c in P_j : $L_{i,j} \leftarrow (\max_r a_{r,c}; 1; w)$ return $L_{i,j}$

Table 4

Pseudocode to generate correlation value-triplets for row-regular (left) and column-regular (right) blocks ($\max_c a_{r,c}$: the maximum value in row vector r in A ; other notation as in Table 3).

$f_{\text{ROW-REGULAR}}(A, P_i, P_j)$:	$f_{\text{COLUMN-REGULAR}}(A, P_i, P_j)$:
$L_{i,j} = \{ \}$ $w = P_j - [i = j]$ for each r in P_i : $L_{i,j} \leftarrow (\max_c a_{r,c}; 1; w)$ return $L_{i,j}$	$L_{i,j} = \{ \}$ $w = P_i - [i = j]$ for each c in P_j : $L_{i,j} \leftarrow (\max_r a_{r,c}; 1; w)$ return $L_{i,j}$

framework for binary networks (see Doreian et al., 2005, p. 224), the incurred penalty for a block row depends on whether it is empty or contains more than one tie. For an empty block row, the incurred penalty is equal to the length of this row, i.e. the number of columns that the block has. Conversely, a single penalty is incurred for the number of ties in a block row exceeding one. Corresponding penalties are incurred for column-functional blocks.

Translating the criteria for row- and column-functional blocks into value-triplets for the correlation-based approach is less straight-forward than what was the case for the previously specified ideal blocks. In Nordlund (2018, p. 354ff), the suggested approach for capturing row-functionality is to correlate the largest value in a block row with unity and all remaining row values with zero. For block rows completely devoid of ties, only one of these zero values would then be correlated with unity where remaining zeros were correlated with zero. Inspired by the full-row penalty for empty block rows in the conventional generalized approach, we instead opt for the more “penalizing” alternative suggested in Nordlund (2018, p. 355, note 11) where a similar full-row weighted penalty is incurred for would-be empty block rows. Pseudocode for capturing row- and column-functional blocks in this way is given in Table 5 below.

3. Example section

This section will demonstrate the proposed direct blockmodeling approach on three binary and three valued networks. Starting with a small 10-actor toy network previously used to demonstrate core-periphery structures (Borgatti & Everett, 2000; Nordlund, 2018), this is followed by analyses of two classical binary networks previously used to demonstrate generalized blockmodeling: friendship among the Transatlantic industries Little League baseball team (Fine, 1987;

¹⁰ In binary networks, this would then ideally be unity.

Table 5

Pseudocode to generate correlation value-triplets for row-functional (left) and column-functional (right) blocks (notation as in Table 3 and Table 4).

$f_{\text{ROW-FUNCTIONAL}}(A, P_i, P_j)$:	$f_{\text{COLUMN-FUNCTIONAL}}(A, P_i, P_j)$:
$L_{i,j} = \{ \}$ $w = P_j - [i = j]$ for each r in P_i : $\max_r = \max_c a_{r,c}$ if $\max_r = 0$: $L_{i,j} \leftarrow (0; 1; w)$ else: index = -1 for each c in P_j : if $a_{r,c} = \max_r$: index_c = c for each c in P_j : if $c = \text{index}_c$: $L_{i,j} \leftarrow (a_{r,c}; 1; 1)$ else: $L_{i,j} \leftarrow (a_{r,c}; 0; 1)$ return $L_{i,j}$	$L_{i,j} = \{ \}$ $w = P_i - [i = j]$ for each c in P_j : $\max_c = \max_r a_{r,c}$ if $\max_c = 0$: $L_{i,j} \leftarrow (0; 1; w)$ else: index = -1 for each r in P_i : if $a_{r,c} = \max_c$: index_r = r for each r in P_i : if $r = \text{index}_r$: $L_{i,j} \leftarrow (a_{r,c}; 1; 1)$ else: $L_{i,j} \leftarrow (a_{r,c}; 0; 1)$ return $L_{i,j}$

Doreian et al., 2005, p. 48), and the Kansas Search and Rescue data (Drabek et al., 1981; Doreian et al., 2005, p. 203). For these binary networks, the results from the proposed correlation-based approach will be compared with corresponding results from using the conventional direct approach for structural, regular and generalized block-modeling.

Following the binary examples are three valued examples: the Hlebec notesharing network (Hlebec, 1996; Žibera, 2007a), the co-presence of Florida primates (Borgatti & Everett, 2000, p. 380), and the Freeman EIES friendship data (Freeman & Freeman, 1980). These examples will demonstrate both how the proposed goodness-of-fit measure can be used for finding optimal partitions, blockimages, or both, as well as for hypothesis testing with given partitions.

3.1. Binary example 1: The BEfig1 core-periphery network

The first example is the 'BEfig1' network in Fig. 4 below (Borgatti & Everett, 2000, p. 377; Nordlund, 2018, p. 358), a trivial binary, symmetric 10-actor network introduced by Borgatti and Everett to exemplify how a typical core-periphery structure could look like. With the default Borgatti-Everett core-periphery specification, i.e. prespecifying the blockimage with a complete block for intra-core ties, a null block for intra-peripheral ties, and a 'do-not-care' block for core-periphery

	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	0	0	0	0	0
2	1	1	1	1	0	1	1	1	0	0
3	1	1	1	1	0	0	0	1	1	0
4	1	1	1	1	1	0	0	0	0	1
5	1	0	0	1	1	0	0	0	0	0
6	0	1	0	0	0	1	0	0	0	0
7	0	1	0	0	0	0	1	0	0	0
8	0	1	1	0	0	0	0	1	0	0
9	0	0	1	0	0	0	0	0	1	0
10	0	0	0	1	0	0	0	0	0	1

Fig. 4. The 'BEfig1' core-periphery example, blocked according to the intuitive core (see Borgatti & Everett, 2000, p. 377; Nordlund, 2018, p. 358).

Table 6

Optimal solutions for 2-positional structural equivalence blockmodeling of the 'BEfig1' network.

Correlation-based optimal solution [com, nul], k = 2				
# solutions	Blockimage ^a	corr (penalties)	P1	P2
1	[com,nul],[nul,nul]	0.5837 (16)	1,2,3,4	rest
1	[com,com],[com,nul]	0.5324 (24)	2,3,4	rest
Inconsistency-based optimal solutions [com, nul], k = 2				
# solutions	Blockimage	Penalties (corr)	P1	P2
3	[com,nul],[nul,nul]	16 (0.5837)	1,2,3,4	rest
	“-	16 (0.5645)	1,2,3,4,5	rest
	“-	16 (0.5645)	1,2,3,4,8	rest
3	[com,com],[com,nul]	22 (0.4664)	2,3	rest
	“-	22 (0.4664)	2,4	rest
	[-,com],[com,nul]	22 (0.3840)	2	rest

^a In the notation used here, $[a_{11}, a_{12}], [a_{21}, a_{22}]$ indicates a 2-by-2 blockimage where block position a_{rc} is the ideal block between actor set P_r and P_c . Dashes (–) are do-not-care or not-applicable, e.g. for singleton positions.

ties, both the correlation-based and inconsistency-based approaches find a single ideal solution ($corr = 1.000$; $penalty = 0$) when actors 1 to 4 constitute the core, i.e. the outlined blockmodel in Fig. 4 below.

Allowing for any non-trivial 2-positional configuration of complete and null blocks, i.e. a direct exploratory structural equivalence block-model analysis of the BEfig1 network, both the penalty-based and the correlation-based heuristics confirm that a default core-periphery structure with null blocks in the off-diagonal blocks constitutes the optimal solution ($corr = 0.5837$; $pen = 16$) – see Table 6. However, whereas both heuristics place actors 1 to 4 in that solution, the inconsistency-based heuristics finds two additional, equally optimal, solutions that also place actor 5 and 8, respectively, in the core. Even though these solutions indeed contain 16 inconsistencies, the correlation-based approach deems these additional 5-actor-core solutions as slightly less optimal ($corr = 0.5645$) than the more intuitive 4-actor core (i.e. $corr = 0.5837$).

The second-most optimal 2-positional blockimage is the default core-periphery structure with complete blocks in the off-diagonal blocks – see Table 6. However, while the correlation-based approach finds a single optimal solution when the core consists of actors 2, 3 and 4 ($corr = 0.5324$; $pen = 24$), the conventional direct approach returns three equally optimal partitions ($pen = 22$), none of which are the same as the 2,3,4-core partition suggested by the correlation-based approach. Whereas two of these optimal solutions have a corresponding correlation of 0.4664, the solution with the singleton actor in the core only has a corresponding correlation-based goodness-of-fit measure of 0.3840.

Repeating the free-search optimization with three types of ideal blocks – complete, regular, and null blocks (in this order of priority) – both heuristics find identical ideal solutions as given in Table 7. In addition to the intuitive clique core with actors 1 to 4, there is also an alternative ideal solution where actors 2 to 5 form a core in the form of a regular block.

Table 7

Optimal solutions for 2-positional generalized equivalence blockmodeling (complete, regular and null ideal blocks; this order of priority) of the 'BEfig1' network.

Optimal solutions (both heuristics) [com, reg, nul], k = 2				
# solutions	Blockimage	corr / Penalty	P1	P2
2	[com,reg],[reg,nul]	1.0000 / 0	1,2,3,4	rest
	[reg,reg],[reg,nul]	1.0000 / 0	2,3,4,5	rest

Table 8

The Transatlantic Industries team friendship network (Fine, 1987, p. 144; Doreian et al., 2005, pp. 8, 48).

	Ron_1	Tom_2	Frank_3	Boyd_4	Tim_5	John_6	Jeff_7	Jay_8	Sandy_9	Jerry_10	Darrin_11	Ben_12	Arnie_13
Ron_1	1	0	1	1	1	0	0	0	0	0	0	0	0
Tom_2	1	1	1	0	0	0	0	0	0	0	1	0	0
Frank_3	1	0	1	1	0	0	0	0	0	0	1	0	0
Boyd_4	1	1	1	1	0	0	0	0	0	0	0	0	0
Tim_5	1	0	1	1	1	0	0	0	0	0	0	0	0
John_6	0	0	0	0	1	1	0	0	0	0	0	1	1
Jeff_7	0	1	0	0	0	0	1	1	0	0	0	0	0
Jay_8	0	1	0	0	0	0	1	1	1	0	0	0	0
Sandy_9	0	1	0	0	0	0	1	1	1	0	0	0	0
Jerry_10	1	0	0	0	0	1	0	0	0	1	0	0	0
Darrin_11	1	1	0	0	0	0	0	0	0	1	1	0	0
Ben_12	1	0	0	0	0	1	0	0	0	1	1	1	0
Arnie_13	0	0	0	0	1	1	0	0	0	0	0	0	1

3.2. Binary example 2: Little League baseball teams (Transatlantic Industries)

From data collected by Fine (1987) on friendship nominations within Little League baseball teams, Doreian and colleagues look closer at one of these teams – the Transatlantic Industries team – to demonstrate aspects of direct structural equivalence blockmodeling and the enhanced features of the generalized blockmodeling framework they propose. With each of the 13 team members nominating three¹¹ friends, the adjacency matrix for this binary, directed Transatlantic Industries network is given in Table 8.

Replicating the direct structural equivalence blockmodeling of Doreian et al. (2005, p. 196ff), i.e. allowing for complete and null blocks in arbitrary configurations, their Hamming-optimized solutions for partitions with 2 to 5 positions are given in the left side of Table 9 below. For each solution, corresponding blockmodels can be found in Fig. 5. Repeating the blockmodel analysis when using the correlation-based goodness-of-fit measure, we arrive at the optimal solutions as given in the right-hand side of Table 9 below.

Three of the Hamming-optimized solutions in Fig. 5 contain blocks that have perfect 0.5 densities. For the 4-positional solution (blockmodel D), the $B_{2,1}$ block could be either a complete or a null block, yielding the same number of inconsistencies in both cases. Similarly, one block each in two of the 5-positional solutions ($B_{5,5}$ in blockmodel E; $B_{4,2}$ in blockmodel F) have mid-point densities where either complete or null blocks would yield the same number of inconsistencies. However, the corresponding correlation-based measures differ in all three cases depending on whether we model these blocks as complete or null.

The correlation-based approach produces single optimal blockmodel and blockimage solutions for each specified number of positions. For the 2-positional model, both heuristics arrive at the same optimal solution. When fitting three positions, the correlation-based approach finds a single optimal solution (blockmodel H) that has four more inconsistencies than the two optimal solutions found when using the conventional approach (blockmodel B and C). Of particular interest in the blockmodel H solution is the $B_{3,3}$ block, i.e. the intra-positional ties between John (6), Jerry (10), Ben (12), and Arnie (13). With 6 out of a total of 12 potential ties among these four boys, it would seem reasonable that this could either be categorized as a complete or a null

block. However, in the correlation-based approach, this is not the case. When modeling $B_{3,3}$ as a complete block, the goodness-of-fit measure is about 10 percent higher (0.5752) than what would be the case when modeled as a null block (0.5173). These are effects from the inherent 50/50 bias of the point-biserial correlation previously discussed.

The particularities with the correlation-based heuristic is further emphasized when examining the correlation-based optimal solution for the 4-positional blockmodel (blockmodel I in Fig. 5). Examining block $B_{1,2}$, we note that two individuals nominate Tom as a friend, the latter constituting a singleton position (P_2) in this solution. Despite this block only having 2 out of a total of 5 possible ties, the correlation-based heuristic nevertheless categorizes this as a complete block. Pre-specifying with the partition and blockimage for this optimal solution, setting this block as a null block instead ($I_{1,2} = \text{nul}$), the correlation for this solution would be the slightly lower (0.6309).¹²

Expanding the analysis towards the generalized framework, allowing for complete, regular and null blocks (in this order), optimal solutions for the two approaches are given in Table 10 above. As neither goodness-of-fit measure improve when running 3-positional optimizations, only optimal 2-positional solutions are presented in this table. Where the conventional direct approach finds 8 equally optimal solutions, the correlation-based approach identifies a single optimal solution, corresponding to the first of the 8 solutions in Table 10.

3.3. Binary example 3: Kansas Search and Rescue communication

The final binary example consists of the communication network between organizations and units involved in a search and rescue operation following a boat accident on Lake Pomona in Kansas, USA. Collected and analyzed by Drabek et al. (1981), the original data was directional and valued, capturing frequencies of communication between 20 organizations and offices involved in the emergency operation. Prior to their blockmodel analysis, Drabek and colleagues binarized this data by coding all reported communication ties as binary ties, this being the data that was subsequently analyzed by Doreian et al. (2005), as well as being the data analyzed here.

In the original study by Drabek and colleagues (1981), the CONCOR algorithm (see, e.g., Wasserman & Faust, 1994, p. 376) was used to partition the network into five positions, resulting in the blockmodel and blockimage as shown in Fig. 6 below.

As noted by Doreian et al. (2005, p. 203), the CONCOR-derived blockmodel contains 79 inconsistencies, which Doreian and colleagues deemed as less than optimal. When instead conducting a direct structural equivalence blockmodeling with five positions, Doreian et al. found four equally optimal solutions, each with 57 inconsistencies (Doreian et al., 2005, pp. 205, 208). These four optimal partitions and, especially, their corresponding blockimages are indeed quite similar to each other – see Table 11 below – and substantially different from the CONCOR-derived blockmodel in Drabek et al. (1981).

The correlation-based goodness-of-fit measure for the CONCOR-derived solution in Fig. 6 above is somewhat low (0.5608). For the four optimal solutions in Table 11, the corresponding correlations are all higher and almost identical, with the first solution having a slightly

¹¹ Note that Jerry (10) and Arnie (13) only nominated two friends each (Fine, 1987, p. 144).

¹² Calculating the correlations for these two solutions using the notation for the point-biserial formula (see Eq. 1) is as follows. Both solutions have 156 datapoints (n) with the standard deviation s of the X values being ~ 0.4254 . For the optimal solution (i.e. with $B_{1,2} = \text{com}$), $r_{pb} = ((0.6275 - 0.0476)/s) \cdot \sqrt{51 \cdot 105/n^2} = 1.3632 \cdot 0.4691 = 0.6395$, and when $B_{1,2} = \text{nul}$, $r_{pb} = ((0.6522 - 0.0636)/s) \cdot \sqrt{46 \cdot 110/n^2} = 1.3836 \cdot 0.4560 = 0.6309$. Thus, even though the null block increases the difference between the mean X values of M_1 and M_2 , this gain to the point-biserial correlation is more than offset as the ideal blockmodel (i.e. the values in Y) moves away from a midpoint density of 0.5.

Table 9

Optimal solutions for conventional and correlational direct structural equivalence blockmodeling of Transatlantic Industries baseball team, 2 to 5 positions (consult Fig. 5 below for respective blockmodel solution).

k	Hamming-based optimal solutions			Correlation-based optimal solutions		
	Hamming	(corr)	Blockmodel	corr	(Hamming)	Blockmodel
2	29	(0.4046)	A	0.4046	(29)	A
3	23	(0.5559)	B	0.5752	(27)	H
	23	(0.5534)	C			
4	20	(0.6191)	D	0.6395	(24)	I
5	17	(0.7080)	E	0.7080	(17)	E
	17	(0.6871)	F			
	17	(0.6812)	G			

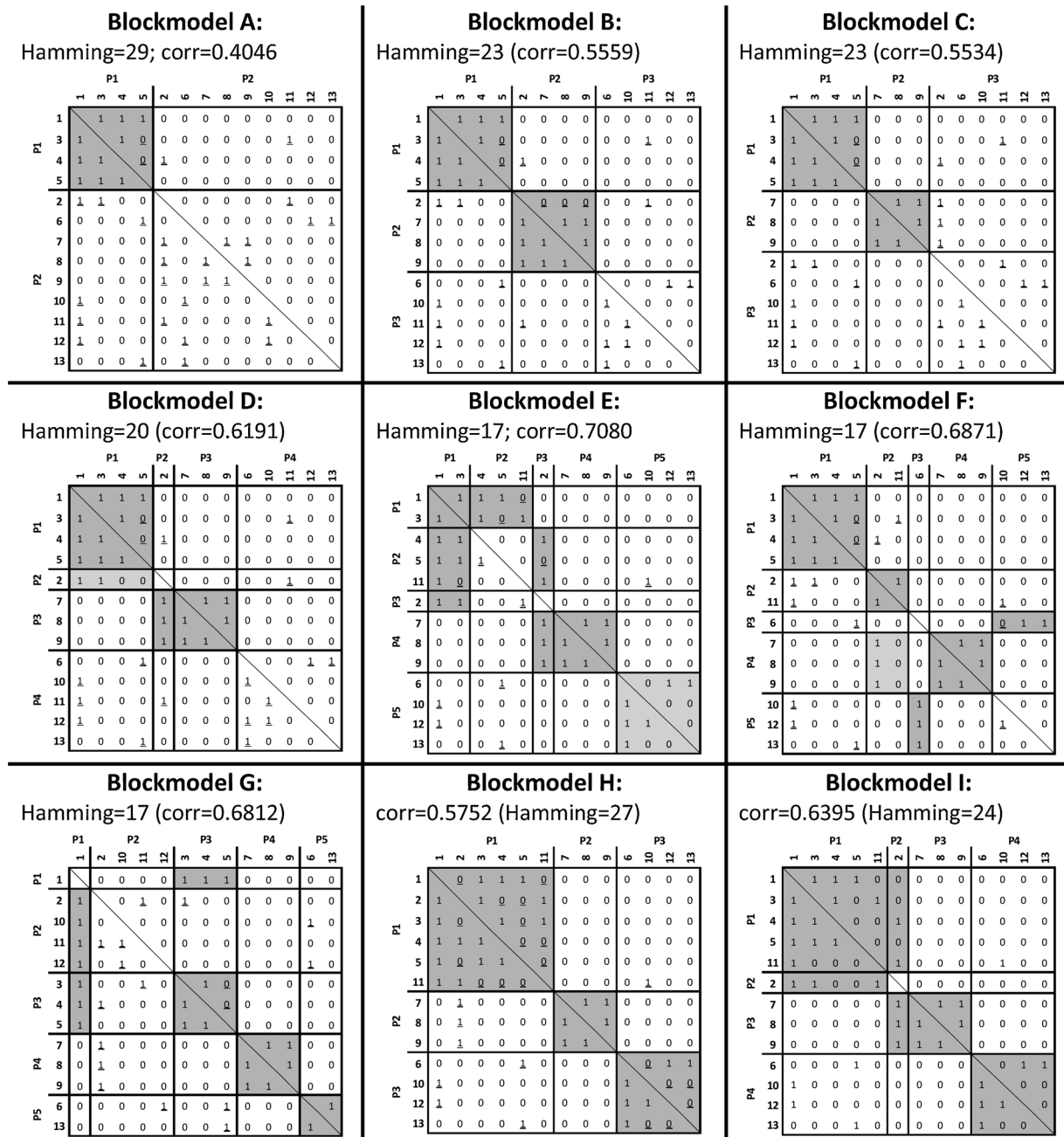


Fig. 5. Optimal blockmodel solutions for the Transatlantic Industries team (dark-shaded blocks: complete blocks; light-shaded: complete or null blocks; non-shaded: null blocks; underlined cell values: block inconsistencies).

Table 10

Optimal 2-positional solutions for conventional and correlational generalized equivalence blockmodeling of Transatlantic Industries baseball team (complete, regular and null blocks in this order of priority)

k	Penalty	Correlation	Blockmodel	P1	P2
2	3	0.9598	[com,nul],[nul,reg]	7,8,9	rest
	“-	0.9250	[-,nul],[nul,reg]	13	rest
	“-	0.9120	[nul,reg],[reg,reg]	1,2,7,12,13	rest
	“-	“-	“-	1,2,8,12,13	rest
	“-	“-	“-	1,2,9,12,13	rest
	“-	“-	“-	1,3,7,12,13	rest
	“-	“-	“-	1,3,8,12,13	rest
	“-	“-	“-	1,3,9,12,13	rest

optimal¹³ correlations as optimal, the correlation-based approach indeed seems very apt at finding the same solutions as those found when using the conventional Hamming-based criteria function.

Beyond the structural equivalence blockmodels of these previous studies, this third binary example is rounded off by fitting the Kansas Search and Rescue communication network to the ideal regular and null blocks. When fitting to a 2-positional regular blockmodel, both the conventional and weighted correlation-based direct heuristics find the same single optimal solution ($pen = 2$; $corr = 0.9793$) as outlined in Fig. 7 below (solid lines).

Extending to 3-positional solutions, fitting the network to all unique, non-trivial regular blockimage varieties, the correlation-based heuristic finds the single optimal blockmodel solution as outlined in

		P1		P2					P3			P4				P5						
		A	E	C	F	G	I	K	D	L	N	B	H	J	S	M	O	P	Q	R	T	
P1	Sheriff (A)		1	1	1	1	1	1	0	0	0	1	0	0	0	0	1	1	1	1	0	1
	HighwayP (E)	1		1	1	1	0	1	1	1	1	0	0	0	0	1	0	1	1	0	1	
P2	Coroner (C)	1	1		1	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	
	ParksRes (F)	1	1	1		1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	
	GameFish (G)	1	1	1	1		1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	
	ArmyCorps (I)	1	1	1	1	1		1	0	0	0	0	0	0	0	0	0	0	1	0	0	
	CrableAmb (K)	1	1	1	1	1	1		0	1	0	0	0	0	0	0	0	0	0	0	0	
	Attorney (D)	1	1	1	1	1	1	0		0	1	1	0	0	0	0	0	0	0	0	0	
P3	FrankCoAmb (L)	1	1	0	0	1	0	0	0		0	0	0	0	0	0	0	0	1	0	0	
	Shawney (N)	1	1	1	1	1	1	0	0	1		0	0	0	0	1	0	0	1	1	1	
	CivilDef (B)	1	1	1	1	1	1	0	0	0	1		0	0	0	0	0	1	1	0	0	
P4	KansasDOT (H)	1	1	0	0	0	0	0	0	0	0	1		1	0	0	0	0	0	0	0	
	ArmyRes (J)	1	1	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	
	CarbDF (S)	1	1	0	1	1	1	0	0	0	0	1	0	0		0	0	0	1	0	0	
	LeeRescue (M)	1	1	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	
P5	BurlPolice (O)	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1		1	1	1	1	
	LyndPolice (P)	1	1	0	1	1	0	1	0	0	0	1	0	0	0	0	0		0	0	0	
	RedCross (Q)	1	1	1	1	1	1	0	0	0	1	1	0	1	0	0	0	0		0	0	
	TopicaFD (R)	1	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0		0	
	TopekaRBW (T)	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0		

Fig. 6. The 5-positional CONCOR-derived blockmodel and blockimage as found by Drabek et al. (1981). Notation as in Fig. 5.

Table 11

Optimal solutions for Hamming-based direct structural equivalence (*: For solution 4, i.e. when the Shawnee County Underwater Rescue Team (N) joins the Burlingame Police (O) in position 5, the ideal block $B_{5,2}$ turns into a null block; see also Doreian et al., 2005, p. 208).

	P1	P2	P3	P4	P5	Common members	Solution-specific members			
							1	2	3	4*
P1	com	com	com	nul	nul	A,E				
P2	com	nul	com	nul	nul	B,D,K,P	+N,Q	+N,S	+N,Q,S	+S
P3	com	nul	com	nul	nul	C,F,G,I				+Q
P4	com	nul	nul	nul	nul	H,J,L,M,R,T	+S	+Q		
P5	com	com*	com	com	nul	O				+N

higher correlation ($corr_1 = 0.6856$) than the remaining three ($corr_3 = 0.6831$; $corr_4 = 0.6813$; $corr_2 = 0.6811$).

Prespecifying with the two almost identical blockimages in Table 11, i.e. with $B_{5,2}$ either being a complete or a null block, the correlation-based direct heuristic finds two of the same solutions as found when using the conventional Hamming-based approach. For the blockimage when $B_{5,2}$ is a complete block, the correlation-based approach finds the same optimal solution as solution 1 in Table 11, and solution 4 is found when $B_{5,2}$ is a null block. Thus, if also categorizing solutions with near-

¹³ Looking at the corresponding correlation-based measures for the four Hamming-optimized 57-penalty solutions in Table 11, it can be noted that the solution with the lowest correlation ($corr_2 = 0.6811$) is only 0.6 percent lower than the solution with the highest correlation ($corr_1 = 0.6856$). When finding the optimal partition when using correlations, it thus seems advisable to also explore solutions with correlation scores close to the optimal solution, either to ensure that the chosen solution sticks out from the next-best solution, or that a set of almost-equally-optimal solutions form a distinguishable cluster vis-à-vis subsequent solutions.

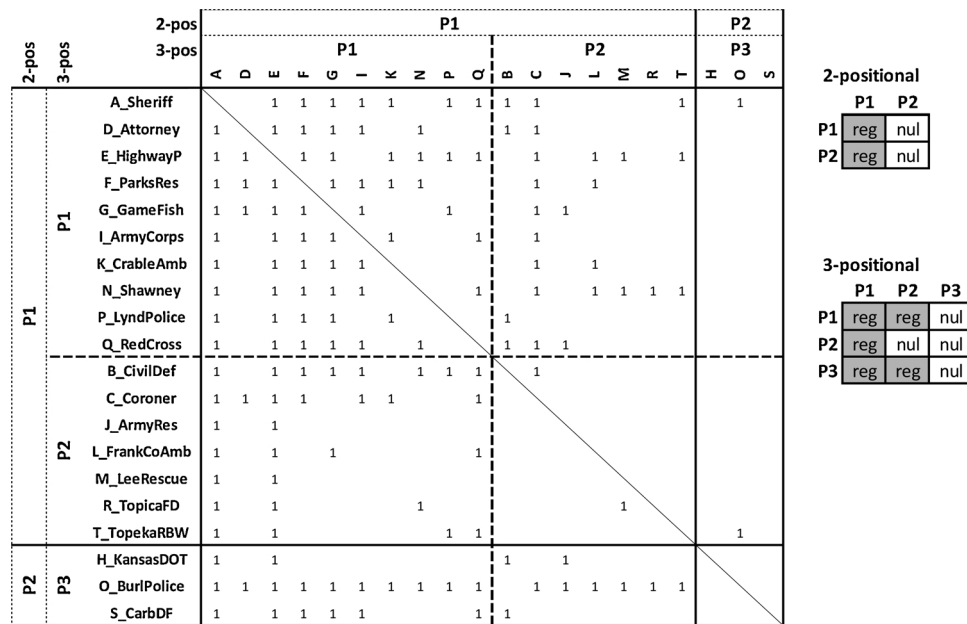


Fig. 7. Optimal regular 2- and 3-positional blockmodels for the Kansas SAR communication network (2-positional partition as solid lines; 3-positional partition as solid and dashed lines).

Fig. 7 (solid and dashed lines). At a goodness-of-fit correlation at 0.9727, the 3-positional solution is objectively less ideal than the optimal 2-positional solution. However, the 3-positional solution does find a single solution that splits the previous 17-actor regular block into a tighter regular core, consisting of 10 organizations whose communications among each other form a 3-core, and a set of 7 actors that virtually only have ties to the regular core.

Instead, when optimizing using the conventional inconsistency-based measure, the same 3-positional solution found by the weighted correlation-approach is found (Fig. 7) – as well as 2946 other solutions, all similarly having four inconsistencies each. Of these alternative solutions, it turns out that all but one are solutions for a trivial¹⁴ extension of the 2-positional regular blockimage solution, resulting in a plethora of possible actor permutations that all fulfil the criteria for such adjacent regular blocks. Filtering out the solutions for this trivial blockimage, a single alternative 4-penalty optimized solution¹⁵ remains: with a marginally lower correlation (0.9633) than the optimal solution (0.9727), both solutions should reasonably be viewed as equally viable, where a would-be choice between them should rest on substantive, interpretational grounds rather than a minuscule difference in a goodness-of-fit measure.

3.4. Valued example 1: Hlebec notesharing network

Building on survey data collected by Hlebec (1996), this network captures the self-reported sharing of notes among 13 Slovenian social-informatics students. With tie values ranging from 1 to 19 and tie directionality capturing the relation from borrower to lender, this network has been used to demonstrate the valued blockmodeling approach

¹⁴ The blockimage for these solutions is similar to the 3-positional solution in Fig. 6 above, though with $B_{2,2}$ being a regular block. This implies that position 1 and 2 are virtually identical in this blockimage – the two positions in this blockimage are structurally equivalent! – leading to a large set of equivalent solutions. Notably, the weighted correlation measures are all identical for each of these permutations.

¹⁵ $I = [\text{reg}, \text{reg}, \text{nul}], [\text{reg}, \text{reg}, \text{nul}], [\text{reg}, \text{nul}, \text{nul}]$; $P = (\{A, B, F, G, I, J, N, Q\}, \{C, D, E, K, L, M, O, P, R, T\}, \{H, S\})$.

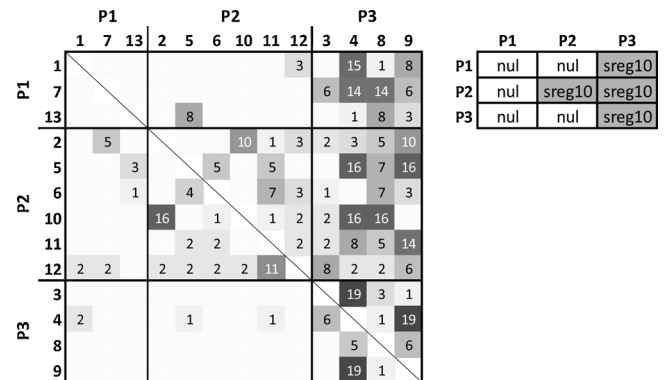


Fig. 8. Optimal blockmodel and blockimage using Žibera's valued network approach (sreg10:sum-regular blocks with $m = 10$).

of Žibera (2007a) as well as the deviational blockmodeling approach of Nordlund (2016). As the latter approach involves a transformation of the valued data, the comparisons in this section is primarily with respect to the findings of Žibera.

In his analysis, Žibera (2007a) searched for optimal 3-positional blockmodels containing either conventional null blocks or sum-regular blocks, the latter being one of the novel valued block types suggested herein. The criteria for a sum-regular block is that the sum of each individual row and column is at least m , the latter being a network-wide parameter that he chooses to set to 10 in this particular analysis. This results in the single solution as given in Fig. 8.

Switching to the correlation-based heuristic, the optimal regular blockmodels and their corresponding blockimages for 2, 3 and 4 positions are given in Fig. 9. The blockimage for the correlation-based 3-positional solution is identical to that found by Žibera (see Fig. 8), but their respective partitions are somewhat different, with only 8 out of 13 actors placed in the corresponding positions for respective solution. Despite this, their corresponding correlation-based measures are nevertheless quite similar: for the solution found by Žibera using his suggested approach (Fig. 8), the corresponding correlation coefficient is 0.8189, i.e. only 7 percent lower than that found when optimizing on this correlation measure (blockmodel B in Fig. 9 below).

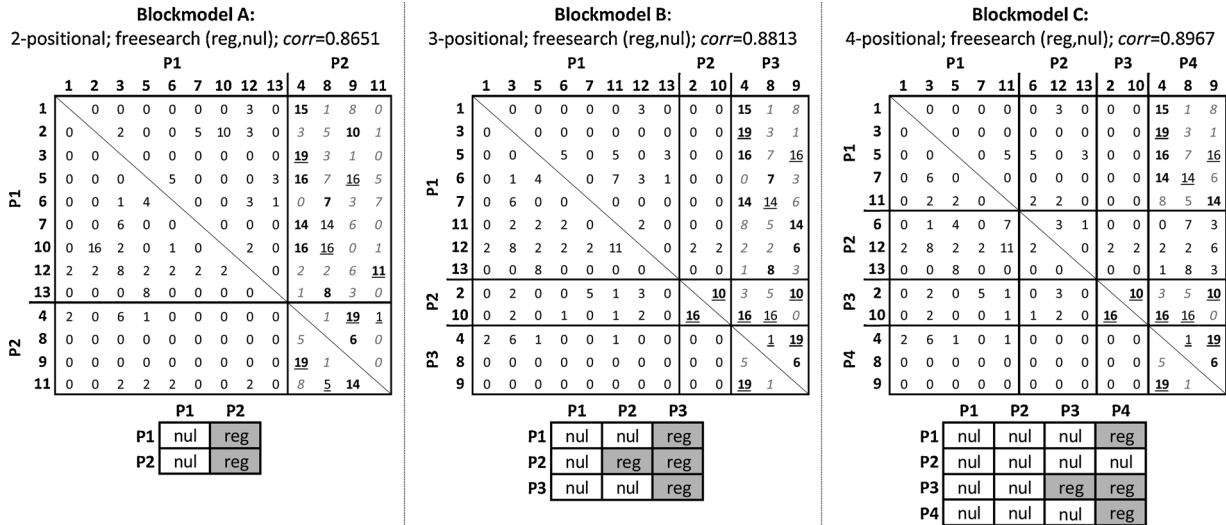


Fig. 9. Correlation-optimal regular blockmodels ($k = 2,3,4$) of Hlebec notesharing network (cell notation for checking block criteria: bold: row-regular criteria; underline: column-regular criteria; default: null block criteria; italicized/grey: excluded).

Extending to four positions, once again testing for ideal regular and null blocks, we arrive at an optimal solution with a fractionally higher correlation (0.8967) – see blockmodel C in Fig. 9 below. This 4-positional optimal solution is a single-split descendant from the 3-positional optimal solution, with students 6, 12 and 13 in position P_1 of the 3-positional solution (blockmodel B) forming their own position in the 4-positional solution. A particular feature of these three students (i.e. P_2 in blockmodel C) is that they are relatively disconnected from the overall note-borrowing circus, both as lenders and borrowers, a social role not readily observed in the optimal blockimages for the optimal 2- and 3-positional solutions. Thus, even though the optimal correlation for the 4-positional solution is only marginally higher than that of the optimal 3-positional solution, the former solution might still make more sense in this particular context, i.e. that there indeed might be students that are less keen to both borrow and lend out notes to their peers.

3.5. Valued example 2: primates at a Florida river

Based on observations of a troop of 20 monkeys at a river in Ocala, Florida, this symmetric valued network captures the total number of joint presences at the river during a 3-month period. With 5 male and 15 female monkeys, Borgatti and Everett (2000) used this projected 2-mode network to demonstrate how their core-periphery measure can be used with given partitions. Testing the hypothesis that core-vs-periphery membership followed a male-female division, with the off-diagonal female-male interactions modeled as complete blocks, Borgatti and Everett found a correlation of 0.2058, whose insignificance was demonstrated with a QAP-test (2000, p. 381).

Using their prespecified 2-positional core-periphery model (i.e. with the off-diagonal blocks modeled as complete blocks) in search of a correlation-wise optimal partition, the most optimal solution (corr = 0.5463) places one male and two female monkeys in the core (blockmodel A in Fig. 10 below). When allowing for any non-trivial 2-positional structural blockmodel, the optimal solution (corr = 0.5821) is a core-periphery structure with null blocks in the off-diagonal blocks, with four males and eight females constituting the core – see blockmodel B in Fig. 10 below.

Extending the freesearching to three positions, a single optimal solution is found (blockmodel C in Fig. 10) at a correlation of 0.6637. Both the partition and blockimage for this solution is a ‘hierarchical child’ of the optimal 2-positional solution. Both solutions have identical ‘peripheral’ positions, both with respect to its actors and associated null

blocks, and the 3-positional blockimage splits the 2-positional core to a cohesive and a non-cohesive part. Continuing to a 4-positional configuration, this pattern is repeated. At a slightly higher correlation of 0.7058, the optimal 4-positional solution (blockmodel D in Fig. 10) has the same peripheral positions as in the 2- and 3-positional optimal solutions. The 4-positional blockimage extends in a similar way on the optimal 2- and 3-positional solutions. The ‘peripheral’ position in the 4-positional solution remains the same as in the optimal 2- and 3-positional solutions, where the remaining positions are only hierarchically linked to the grander core position of the 2-positional solution.

Switching to regular equivalence, i.e. allowing for regular and null blocks, the testing of all unique, non-trivial regular blockimages finds a single optimal partition at a correlation of 0.8894. This optimal solution is identical to its structural-equivalence counterpart (i.e. blockmodel B in Fig. 10 above) where a regular block replaces the complete block. Increasing the number of positions, the optimal 3-positional regular blockmodel has a blockimage that is also similar to its structural-equivalent counterpart (i.e. blockmodel C in Fig. 10) where regular blocks replace the complete ones). As before, the disconnected ‘peripheral’ position in the 3-positional regular solution has the same actors as all other structural and regular optimal solutions. The optimal 3-positional regular solution does however have a slightly lower goodness-of-fit (0.8777) than its 2-positional counterpart, which together with position P_2 being a singleton position (holding F12) indicates that the 2-positional representation is the better choice for representing regular equivalence in this network.

Can the functional anatomy of this network be adequately expressed in terms of the more complex ideal blocks used in generalized blockmodeling? Prespecifying a core-periphery blockimage where periphery-to-core¹⁶ relations are modeled as row-regular, the optimal solution with a correlation of 0.7556 has one male and four females in the core – see blockmodel A in Fig. 11 below. Changing the ideal off-diagonal blocks to regular blocks, the same optimal partition is found, though this time at a slightly higher correlation (0.7836).

When switching the periphery-to-core block into a column-regular block, i.e. with core-to-periphery ties in the form of a row-regular block, the optimal solution – see blockmodel B in Fig. 11 below – has a correlation of 0.8903. With peripheral dependency no longer being a

¹⁶ And due to the symmetry of this network, core-to-periphery relations are here modeled as column-regular.

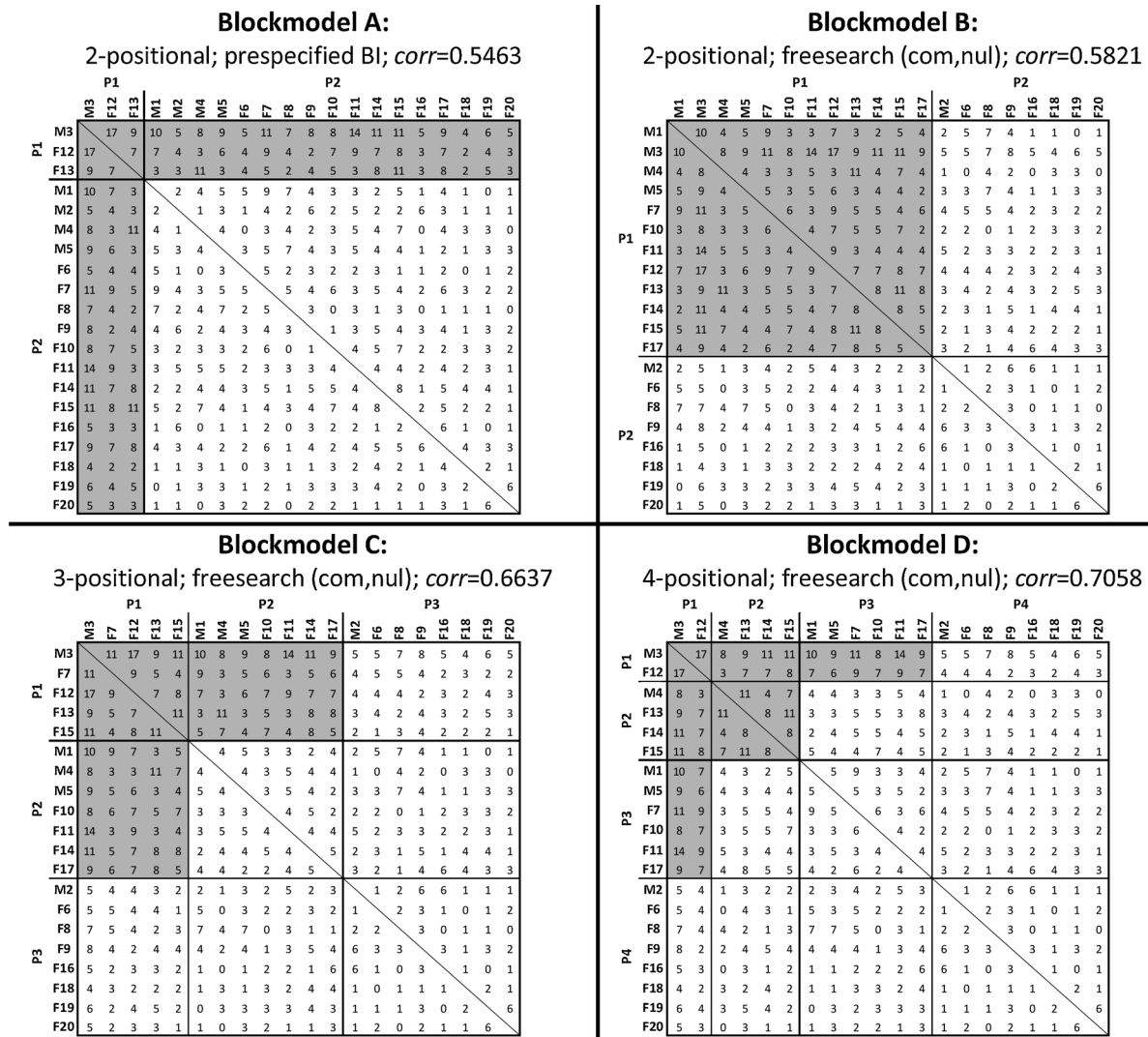


Fig. 10. Correlation-optimal structural blockmodels of Florida river primates network (shaded blocks: complete blocks; non-shaded blocks: null blocks).

criteria for core-periphery relations (see Nordlund, 2018, p. 354), the composition of the core is slightly different from the previous model (blockmodel A), where two female monkeys (F7 and F13) become peripheral as one additional male monkey (M4) joins the core. Of particular interest is the strong tie between M4 and F13, this being a dominance tie in both solutions in Fig. 11, though where the involved monkeys swap position in the evident core-periphery structure.

3.6. Valued example 3: The Freeman EIES friendship network

The third and final valued example in this article revisits the classical Freeman EIES acquaintance data (Freeman & Freeman, 1980). Capturing the self-reported existence and strength of friendships among 32 scholars interested in social network analysis both prior to (T1) and seven months after (T2) being introduced to a computer-based message and conferencing system, the two networks are directional and valued, with values ranging from 0 (unknown) to 4 (close personal friend). Data was also collected on the academic backgrounds of each scholar, these being either sociology (S), anthropology (A), mathematics/statistics (M), or other (O). During these seven months, 118 novel friendship ties were formed (whereas nine existing friendship ties seem to have ended), increasing the topological density from 66 to 77 percent and the average friendship intensity from 1.37 to 1.68.

Looking at the initially reported friendship ties, testing for all non-trivial structural equivalence blockimages for 2 to 5 positions, the optimal solutions for respective number of positions are given in Fig. 12 below. As the number of positions increase, the optimal goodness-of-fit correlation increases as well as the detail-level of the corresponding blockimages. Whereas the 2- and 3-positional optimal solutions have mixed disciplinary memberships, the 4- and 5-positional optimal solutions reflect something akin to scholarly homophily. This is particularly visible for the group of four anthropologists – 14_A, 21_A, 43_A, and 45_A – that together with a math-statistician (40_M) and an unspecified scholar (44_O) form positions in both the 4- and 5-positional optimal solutions. Seemingly knowing each other well enough to form a complete block, they also share a tendency to *not* knowing most of the sociologists. This anthropology-dominated subset is however well connected to the primary scholar¹⁷ in the network – 1_S – who together with another sociologist and two other anthropologists form a well-nominated core group of scholars.

Instead of searching for optimal solutions, the correlation-based approach can also be used for hypothesis testing. Approaching the

¹⁷ As this specific actor surely corresponds to the late professor Linton Freeman himself, the particular index number of this actor is indeed befitting.

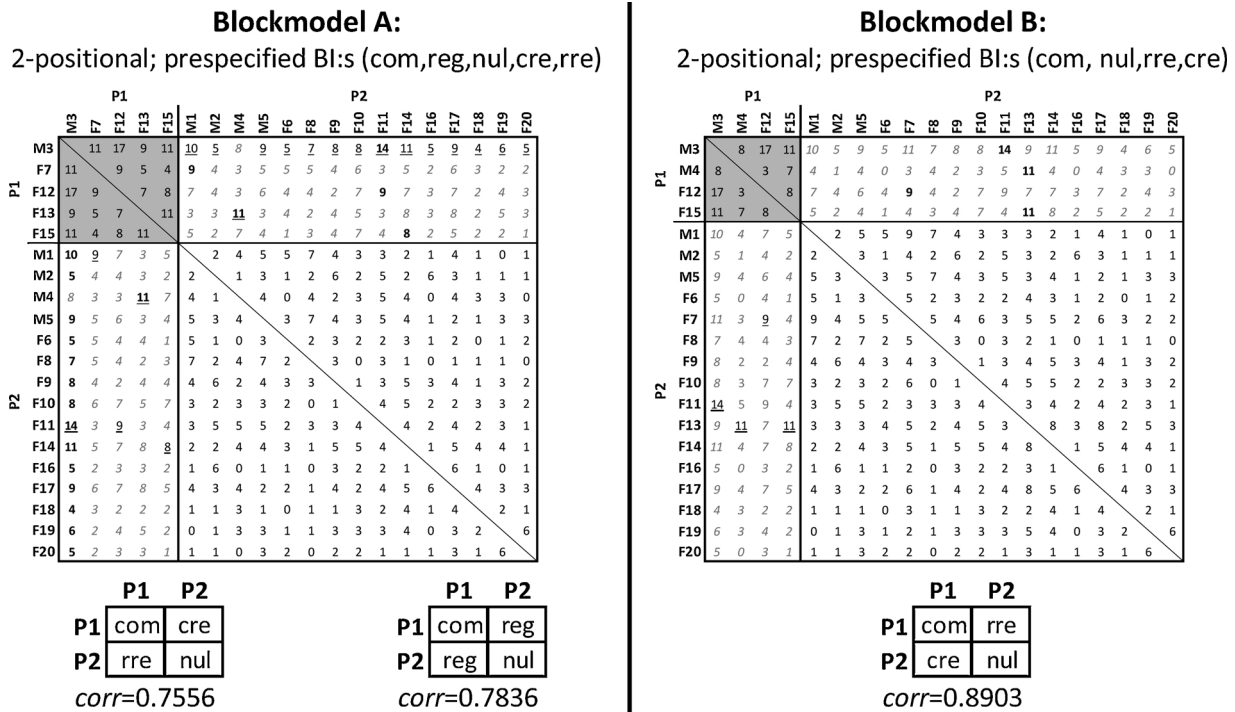


Fig. 11. Optimal 2-positional solutions, prespecified generalized equivalence ideal blocks (bold: row-regular checks; underline: column-regular checks; background-shaded: complete-block checks; default: null-block checks; gray/italics: ignored).

friendship network in this way, we can test and measure the degree of scholastic ‘isolation’ in the friendship networks. Prespecifying¹⁸ a 4-positional cohesive subgroup blockimage and partitioning the actors into discipline-specific positions, the blockmodel below (Fig. 13) corresponds to both time T1 and T2. Although the correlation is low at T1 (0.2522), it is even lower at T2 (0.1770), indicating a decrease in discipline-wise isolation^{19, 20} over time.

From the low goodness-of-fit score and from inspecting the blockmodel in Fig. 13, it is evident that the conventional cohesive subgroups template is unsuitable for testing an hypothesis of discipline-specific role equivalence. For instance, the residual ‘Others’ category indeed is less internally cohesive than what is the case within, respectively, the anthropologists and mathematician-statisticians, whereas friendship among the sociologists is perhaps best captured by a regular block. Additionally, many of the off-diagonal blocks in Fig. 13 above seem quite far away from the ideal null blocks of the classical cohesive-subgroup structural template.

Based on an ocular inspection of Fig. 13, parts of the 4-positional blockimage is modified, allowing for alternative ideal blocks where such seem feasible – see Fig. 14A. However, whereas the conventional inconsistency-minimizing direct heuristic allows for testing individual blocks independently of each other, i.e. by finding the particular ideal block that results in the least number of inconsistencies with respect to the corresponding empirical block, the correlation-based heuristics make such block-wise tests interdependent. That is, if the correlation-based heuristic finds that a regular block is a better representation of an empirical block than a null block, such a change has repercussions on

the mean and standard deviations of the Y values, possibly altering how other empirical blocks are evaluated. This implies that the correlation-based heuristic is inherently global, implying that all potential blockimage configurations, such as the 384 alternative configurations in Fig. 14A, must be evaluated separately.²¹

From the friendship data at time 1, testing the discipline-wise 4-positional partition (see Fig. 13) on the 384 alternative blockimages, the most optimal blockimage, at a correlation of 0.7232, is the one given in Fig. 14B. The full blockmodel for this hypothetical partition is given in Fig. 15 below. For the friendship network at time 2, an almost identical blockimage is found at a slightly lower correlation (0.7076), where the friendship nominations from the anthropologists to the sociologists is better depicted as row-regular at time point 2.

4. Summary and conclusion

This article has proposed a correlation-based approach to direct blockmodeling for binary and valued networks. Building on the core-periphery metric and heuristic of Borgatti and Everett (2000) and its proposed extensions in Nordlund (2018), the suggested approach replaces the conventional inconsistency-based goodness-of-fit measure in direct blockmodeling with a weighted correlation coefficient. This allows for direct blockmodeling of valued as well as binary networks, using the standard set of ideal binary blocks from structural, regular and generalized blockmodeling, without any kind of dichotomization or data reduction at any point in the analysis.

Effectively bypassing the long-standing dilemma with blockmodeling of valued networks, the proposed heuristic has additional advantages. First, contrary to the conventional inconsistency-based

¹⁸ For prespecified blockmodeling, see Doreian et al. (2005, p. 233ff; also Borgatti & Everett, 2000, p. 379ff)

¹⁹ QAP-tests for both these hypothetical partitions point to significance: $p(\text{corr}_{T1} \geq 0.2522) = 0.0002$; $p(\text{corr}_{T2} \geq 0.1770) = 0.0030$.

²⁰ This analysis was also done with regular, rather than complete, blocks in the diagonal of the ideal blockimage, capturing a much smaller goodness-of-fit decrease from 0.7129 to 0.6920.

²¹ In addition to finding optimal partitions for given blockimage configurations, the accompanying software also allows for the creation and comparative testing of such full sets of blockimage ensembles, either for given hypothetical partitions or while also finding the optimal partitions for each permuted blockimage.

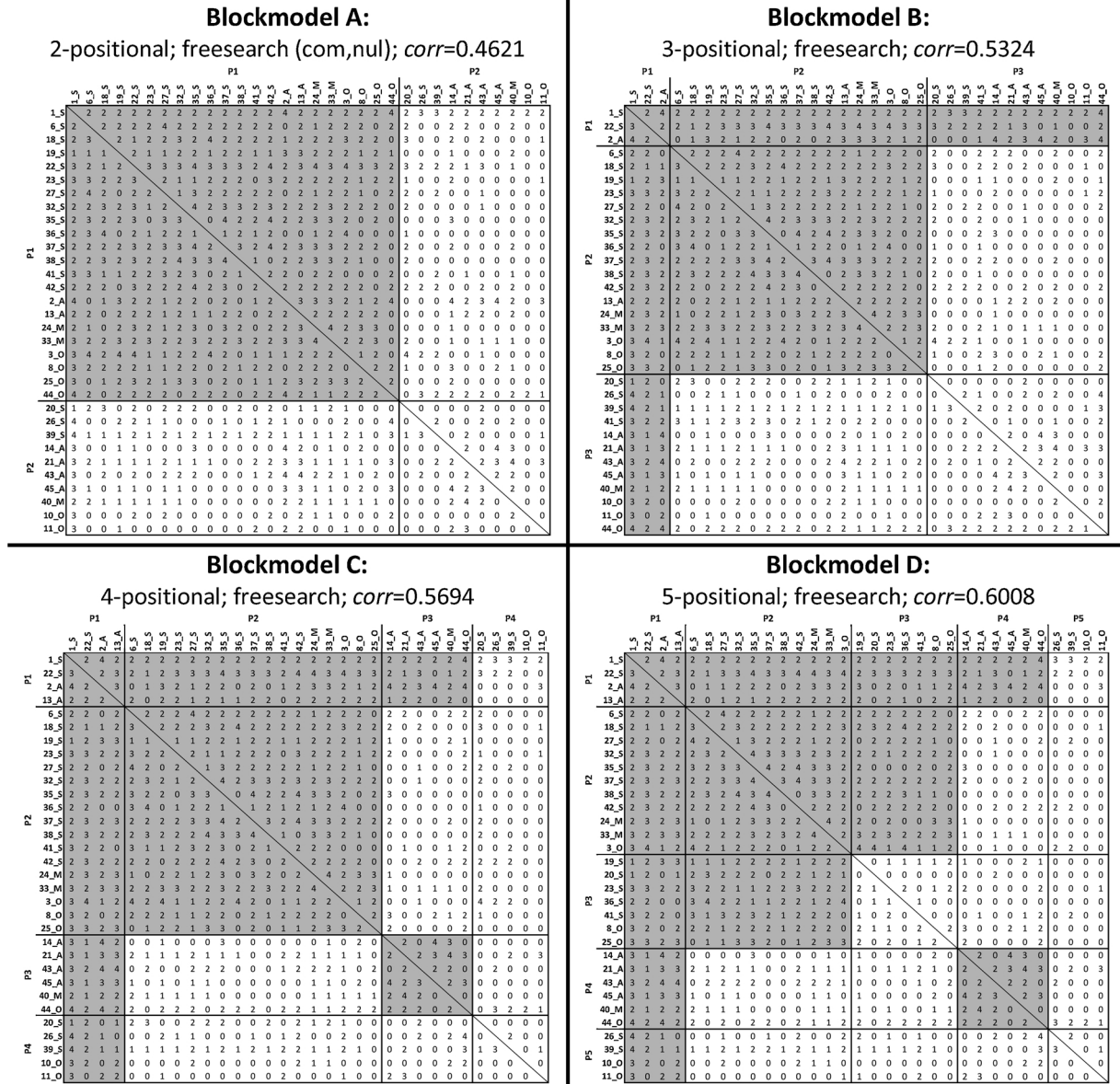


Fig. 12. Correlation-optimal structural blockmodels of EIES Friendships time 1 (shaded blocks: complete blocks; non-shaded blocks: null blocks; S = sociology, A = anthropology, M = mathematics/statistics, O = other).

heuristic, a standardized goodness-of-fit measure is provided, allowing for the comparison of solutions between different network sizes, value types, and equivalence notions. Additionally, benefiting from a specific particularity of point-biserial correlations, the proposed approach is inherently biased towards optimal blockmodel solutions that, it is argued, are easier to interpret. An implication of this bias is that the heuristic tends to distinguish between solutions that in the conventional approach would be deemed equally optimal, preferring solutions whose ideal blockmodels have mid-point densities of 0.5.

The proposed heuristic was demonstrated on six networks. For the three binary examples, the proposed heuristic either finds identical or very similar solutions as those found using the conventional inconsistency-minimizing direct approach. Occasionally finding optimal solutions that are non-optimal in the conventional sense, these solutions are nevertheless argued to be more intuitive and representative than the often multiple optimal solutions found when using the conventional direct approach. The approach specified in this paper is therefore not only proposed for valued networks, but also as a viable alternative, and possibly a replacement, of the conventional direct blockmodeling

approach based on minimizing inconsistencies.

As demonstrated in the three valued network examples, the correlation-based approach is directly applicable to valued networks without any form of dichotomization or data reduction at any point in the analysis. For the Hlebec notesharing example, the correlation-based approach finds the identical 3-positional blockimage as that found by Žiberna using his valued blockmodeling approach. However, whereas the latter involves the introduction of a special type of valued regular block, supplemented with a specific parameter setting, the correlation-based heuristic uses the conventional ideal regular block without any parameters to set. In the second and third valued network examples, the herein proposed heuristic indeed seems apt at capturing what seems like intuitive feasible positions. In addition to using the approach for finding optimal partitions, the last EIES friendship example also demonstrates hypothesis testing and comparing goodness-of-fit measures for two separate time period.

Both the binary and valued examples illustrate how the herein proposed approach seems to distinguish between different solutions that in the conventional approach would be deemed equally optimal.

	Soc																	Anth					MathSt			Other								
	1_S	6_S	18_S	19_S	20_S	22_S	23_S	26_S	27_S	32_S	35_S	36_S	37_S	38_S	39_S	41_S	42_S	2_A	13_A	14_A	21_A	43_A	45_A	24_M	33_M	40_M	3_O	8_O	10_O	11_O	25_O	44_O		
Soc	1_S	2	2	2;3	2	2	2	3	2	2	2	2	2	2	3	2	2;3	4	2;3	2;3	2;3	2	2;3	2	2	2	2	2	2	2	2	4		
	6_S	2		2;3	2	2	2	0;2	4	2	2	2	2	2	0;2	2	2	0;2	2	2	2	2;1	0;2	0	1;2	2	2	2	2	0;2	0	0	2	
	18_S	2;3	3		2	3	1	2	0;2	2	3;4	2	4	2	2	0	2	2;3	1;0	2	2;1	0;2	2	0;1	2	2	2	0;2	3;2	2	0	1	2	0;2
	19_S	1;3	1;2	1;2		0	2	1;2	0;2	1;2	2	2	1	2	2	0	1;2	1;2	3;2	3;4	1;2	0;2	0	0;3	2	2	2	2	1;2	0;2	0	2	1;3	
	20_S	1;2	2	3	0		2	0;1	0	2	2;3	2;1	0;1	0;1	2	0	0;1	2;0	0	1	0	0	2;1	0;2	1	2	0	1	0	0	0;1	0	0;2	
	22_S	3	2;3	1;2	2	3		3	2	3	3	4	3	3	3	2	2;3	4	2	3	2;0	1;2	3	0;2	4	3	1;2	4	3	0	0	3;4	2;3	
	23_S	3	3	2	2	1;2	3		0	2	1	1;2	2	2	2	0	0	3	2	2	2	0	0	0;1	2	2	0;1	2	1	0	1;0	2	2;3	
	26_S	4	0	0	2	0	2	1;0		1;0	0	0	0	1;0	0	2	1;0	2	1;2	0;2	0	0;1	2	0;1	0	0	2	2	0	0	0;1	0	4	
	27_S	2	4	2	0	2	2	2	0		1;2	3	2	2	2	2	2	2	0	2	0	0	1	0	1	2	0	2	1	0	0	0	2	
	32_S	2	2	3	2	2	3	1	0	2		4	2	3	3	0	2	3	2	2	0	0	1	0	2	3	0	2	2	0	0	2	0	
	35_S	2	3	2	2	0	3	0;2	0	3	3		0	4	2	0	2	4	2	2	3	0	0	0;2	3	3	0	2	0	0	0	2	0	
	36_S	2;3	3	4	0	1	2;3	1	0	2;3	2	1		1;2	2	0	1;2	2;3	0;2	0;2	0	0	0;2	0	1	2;1	0	4	0	0	0	0	0;2	
	37_S	2;3	2	2	2;3	2	3	2	0;2	3;2	3	4	2		3	0	2;3	4;3	2	3	2	0;2	0;2	0;2	3	3	2	2	2;3	0;2	0	2	0;2	
	38_S	2	2;3	3	2	2	3	2	0	2	4;3	3	3	4;3		0	1	0;2	2	2;3	0	0	0	0	3	3	0	2	1	0	0	0	0	
	39_S	4	1	1	1	1	2	1	3	2	1	2	1	2	1		2	2	1	1	0	2	0	0	1	1	0	2	1	0	1	0	3	
	41_S	3	3	1	1	0	2	2	0	3	2	3	0	2	1	2		2	2	0;2	0	1;2	0	0;2	2	2	1	0	0	0	0	0	2	
	42_S	2;3	2;3	2;3	0;2	2;3	3;4	2;3	2	2	2;3	4	2	3;4	0	2	2		2	2;3	0	0;2	2	0;2	2	2	2	2	2	2	0;2	0	0;2	2;3
Anth	2_A	4	0;2	1;2	3	0	2	2	0;2	1;2	2	2	0;2	2	0;2	0;2	1;2	2		3	4	2	3	4	3;2	3;2	2	2	1	0;2	3;2	2	4	
	13_A	2;3	2	0;2	2;4	0;1	2	2	0;2	1;2	2	1;2	1;2	2	2	0	0;1	2	2		1	2	2	0;4	2	2	2	2	2	0;2	0;1	2	0;2	
	14_A	3	0	0	1;2	0	1	0	0	0	0	3;2	0	0	0	0	0	0;1	4	2		2	0	4	0;1	1;2	3	0	2	0	0	0;2	0;2	
	21_A	3	2	1	1;2	0	1	1	0;2	2	1	1;2	1	0	0	2	2	2	3	3	2		2	3	1;2	1;2	4	1	1	0	3	0	3	
	43_A	3	0;2	2	0;2	0	2	0	2	2	2	2	0	0;2	0;2	0	1;2	2;3	4;3	4;3	0	2		2;3	2	2	2;3	1	0	0;2	0	2	0;3	
45_A	3;4	1;2	0;2	1;3	0	1	0	0	1;0	1	0;3	0;1	0;1	0	0;2	0;1	0;3	3;4	3;4	4	2	3;2		1;3	1;2	2	0	2	0;2	0	0;2	3;4		
MathSt	24_M	2	1;2	0;2	2	0	3	2	0	1;2	2	3	0	3;2	2	0;2	0;1	2	2	3	1	0	0	0;2		4	0;1	2	3	0;2	0	3;2	0;2	
	33_M	3	2	2	3	2	3	0	2	2	3	2	2	3	0	2	2	2	3	3	1	0	1	1	4		1	2	2	0	0	3	0	
	40_M	2;3	2	1;2	1;2	0	1	1;2	0;2	1;2	1	0;2	0;1	0;2	0	0	0	0;2	2	2;3	2;3	4	2	0;2	1;2	1;2		1;2	1;2	0;2	0	1;2	0;2	
Other	3_O	3	4	2	4	4	4	1	2	1	2	2	4	2	0	2	1	1	1	2	0	0	1	0	2	2	0		1	0	0	2	0	
	8_O	3	2	2	2	1	2	1	0	1	2	2	0	2	1	0	2	2	0	2	3	0	0	2	2	2	1	0		0	0	2	2	
	10_O	3;4	0	0;2	0;2	0	2	0;2	0	2	0	0	0	0	0;3	0	0;2	0;2	0;2	2;3	0	0	0	0;2	1;2	1;2	2	0	0	0	0	0	2;4	
	11_O	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	2	2	3	0	0	0	0	0	0	0	0	0		
	25_O	3	0;1	1	2;3	0;1	3	2	0;1	1	3	3	0	2	0	0;1	1	1;2	2	3	2	0;1	0;2	0;2	2;3	3	0;2	3	2	0	0		2	
	44_O	4	2	0;2	2;3	0;2	2	2;3	3	2	0;2	2	2	2	0;2	2	2	2;3	4	2	2	2;3	2	0;4	1;2	1;2	2	2	2	2;3	1;2	2		

Fig. 13. Testing discipline-wise isolation hypothesis, EIES friendships time 1 and 2 (single-valued cells: same value for T1 and T2; double-valued cells: left-hand values for T1 and right-hand values for T2).

A:					B:				
	Soc	Anth	MathSt	Other		Soc	Anth	MathSt	Other
Soc	com,reg	nul,cre	reg,rre	nul,rre	Soc	reg	nul	reg	nul
Anth	nul,rre	com	cre	nul,rre	Anth	nul	com	cre	nul
MathSt	nul,rre,cre	nul	com	nul	MathSt	rre	nul	com	nul
Other	reg	reg	nul	reg,nul	Other	reg	reg	nul	nul

Fig. 14. Alternative blockimages for discipline-specific hypothetical partition, EIES friendship time 1 (A), and the optimal blockimage found for this network and partition (B).

However, rather than focusing solely on the particular solution with the highest correlation, it is recommended to also explore the partitions and blockmodels that are near-optimal, potentially deeming these as equally adequate representations of the underlying functional anatomy of a network. As stressed in the generalized blockmodeling approach, context is crucial: a substantive and theoretical understanding is of paramount importance in the interpretation of blockmodels (e.g. Doreian et al., 2005, p. 352).

A drawback with the herein proposed approach is that it is computationally more intensive than the conventional direct approach. Whereas local optimization searches are NP-hard, irrespective of criteria functions, the inconsistency-minimizing approach allows for the optimization of each individual block separately, independently of how other empirical blocks have been matched to ideal blocks. This is however not the case for the correlation-based approach, where the selection of a particular ideal block as representative of a particular

empirical block also affects the fitting of other empirical and ideal blocks in the blockmodel. Thus, as exemplified by the 384 potential blockimages in the final Freeman EIES friendship analysis, each individual blockimage must be tested and measured individually.

Finally, although the herein proposed approach indeed seems to work well in the examples provided in this paper, it is imperative that the approach is tested further on a much larger set of binary and valued networks. It is also imperative that the particularities of the suggested goodness-of-fit function is scrutinized further. To better understand how the weighted correlation coefficient works for the more complicated blocks of regular and generalized blockmodeling, a first step would be to derive a weighted version of the point-biserial correlation formula, disentangling the goodness-of-fit measure into the left- and right-side factors of the conventional point-biserial correlation. This and similar endeavors are however here left for future research and development.

		Soc																Anth					MathSt			Other							
		1_S	6_S	18_S	19_S	20_S	22_S	23_S	26_S	27_S	32_S	35_S	36_S	37_S	38_S	39_S	41_S	42_S	2_A	13_A	14_A	21_A	43_A	45_A	24_M	33_M	40_M	3_O	8_O	10_O	11_O	25_O	44_O
Soc	1_S		2	2	2	2	2	2	3	2	2	2	2	2	2	3	2	2	4	2	2	2	2	2	2	2	2	2	2	2	2	2	4
	6_S	2		2	2	2	2	2	0	4	2	2	2	2	2	0	2	2	0	2	2	0	0	0	1	2	2	2	2	0	0	0	2
	18_S	2	3		2	3	1	2	0	2	3	2	4	2	2	0	2	2	1	2	0	2	0	2	2	2	0	3	2	0	1	2	0
	19_S	1	1	1		0	2	1	0	1	2	2	1	2	2	0	1	1	3	3	1	0	0	0	2	2	2	2	1	0	0	2	1
	20_S	1	2	3	0		2	0	0	2	2	2	0	0	2	0	0	2	0	1	0	0	2	0	1	2	0	1	0	0	0	0	0
	22_S	3	2	1	2	3		3	2	3	3	4	3	3	3	2	2	4	2	3	2	1	3	0	4	3	1	4	3	0	0	3	2
	23_S	3	3	2	2	1	3		0	2	1	1	2	2	2	0	0	3	2	2	2	0	0	0	2	2	0	2	1	0	1	2	2
	26_S	4	0	0	2	0	2	1		1	0	0	0	1	0	2	1	2	1	0	0	0	2	0	0	0	2	2	0	0	0	0	4
	27_S	2	4	2	0	2	2	2	0		1	3	2	2	2	2	2	2	0	2	0	0	1	0	1	2	0	2	1	0	0	0	2
	32_S	2	2	3	2	2	3	1	0	2		4	2	3	3	0	2	3	2	2	0	0	1	0	2	3	0	2	2	0	0	2	0
	35_S	2	3	2	2	0	3	0	0	3	3		0	4	2	0	2	4	2	2	3	0	0	0	3	3	0	2	0	0	0	2	0
	36_S	2	3	4	0	1	2	1	0	2	2	1		1	2	0	1	2	0	0	0	0	0	0	1	2	0	4	0	0	0	0	0
	37_S	2	2	2	2	2	3	2	0	3	3	4	2		3	0	2	4	2	3	2	0	0	0	3	3	2	2	2	0	0	2	0
	38_S	2	2	3	2	2	3	2	0	2	4	3	3	4		0	1	0	2	2	0	0	0	0	3	3	0	2	1	0	0	0	0
	39_S	4	1	1	1	1	2	1	3	2	1	2	1	2	1		2	2	1	1	0	2	0	0	1	1	0	2	1	0	1	0	3
	41_S	3	3	1	1	0	2	2	0	3	2	3	0	2	1	2		2	2	0	0	1	0	0	2	2	1	0	0	0	0	0	2
42_S	2	2	2	0	2	3	2	2	2	2	4	2	3	0	2	2		2	2	0	0	2	0	2	2	2	2	2	0	0	0	2	
Anth	2_A	4	0	1	3	0	2	2	0	1	2	2	0	2	0	0	1	2		3	4	2	3	4	3	3	2	2	1	0	3	2	4
	13_A	2	2	0	2	0	2	2	0	1	2	1	1	2	2	0	0	2	2		1	2	2	0	2	2	2	2	2	0	0	2	0
	14_A	3	0	0	1	0	1	0	0	0	0	3	0	0	0	0	0	0	4	2		2	0	4	0	1	3	0	2	0	0	0	0
	21_A	3	2	1	1	0	1	1	0	2	1	1	1	0	0	2	2	2	3	3	2		2	3	1	1	4	1	1	0	3	0	3
	43_A	3	0	2	0	0	2	0	2	2	2	2	0	0	0	0	1	2	4	4	0	2		2	2	2	2	1	0	0	0	2	0
45_A	3	1	0	1	0	1	0	0	1	1	0	0	0	0	0	0	0	3	3	4	2	3		1	1	2	0	2	0	0	0	3	
MathSt	24_M	2	1	0	2	0	3	2	0	1	2	3	0	3	2	0	0	2	2	3	1	0	0	0		4	0	2	3	0	0	3	0
	33_M	3	2	2	3	2	2	3	0	2	2	3	2	2	3	0	2	2	3	3	1	0	1	1	4		1	2	2	0	0	3	0
	40_M	2	2	1	1	0	1	1	0	1	1	0	0	0	0	0	0	0	2	2	2	4	2	0	1	1		1	1	0	0	1	0
Other	3_O	3	4	2	4	4	4	1	2	1	2	2	4	2	0	2	1	1	1	2	0	0	1	0	2	2	0		1	0	0	2	0
	8_O	3	2	2	2	1	2	1	0	1	2	2	0	2	1	0	2	2	0	2	3	0	0	2	2	2	1	0		0	0	2	2
	10_O	3	0	0	0	0	2	0	0	2	0	0	0	0	0	0	0	2	0	2	0	0	0	0	1	1	2	0	0		0	0	2
	11_O	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	2	2	2	3	0	0	0	0	0	1	0	0		0	0
	25_O	3	0	1	2	0	3	2	0	1	3	3	0	2	0	0	1	1	2	3	2	0	0	0	2	3	0	3	2	0	0		2
	44_O	4	2	0	2	0	2	2	3	2	0	2	2	2	0	2	2	2	4	2	2	2	2	0	1	1	2	2	2	2	1	2	

Fig. 15. Testing discipline-partitioned hypothesis, EIES friendships time 1 (bold: row-regular checks; underline: column-regular checks; background-shaded: complete-block checks; default: null-block checks; gray/italics: ignored).

References

- Arabie, P., Boorman, S.A., Levitt, P.R., 1978. Constructing blockmodels: How and why. *Journal of Mathematical Psychology* 17 (1), 21–63.
- Borgatti, S.P., Everett, M.G., 2000. Models of core/periphery structures. *Social Networks* 21 (4), 375–395.
- Carrington, P.J., Heil, G.H., Berkowitz, S.D., 1979. A goodness-of-fit index for blockmodels. *Social Networks* 2 (3), 219–234.
- Doreian, P., 1969. A note on the detection of cliques in valued graphs. *Sociometry* 237–242.
- Doreian, P., 2006. Some Open Problem Sets for Generalized Blockmodeling. In: Batagelj, P.D.V., Bock, P.D.H.-H., Ferligoj, P.D.A., Žiberna, A. (Eds.), *Data Science and Classification*, pp. 119–130.
- Doreian, P., Batagelj, V., Ferligoj, A., 2005. *Generalized Blockmodeling*. Cambridge University Press.
- Drabek, T.E., Tamminga, H.L., Kilijaneck, T.S., Adams, C.R., 1981. Managing multi-organizational emergency responses: Emergent search and rescue networks in natural disaster and remote area settings. University of Colorado, institute of Behavioral Science, Boulder, CA.
- Fine, G.A., 1987. *With the boys: Little league baseball and preadolescent culture*. University of Chicago Press.
- Freeman, L.C., Freeman, S.C., 1980. A semi-visible college: Structural effects on a social networks group. *Electronic communication: Technology and impacts* Vol. 52 AAAS Washington, DC.
- Hlebec, V., 1996. Metodoloske značilnosti anketnega zbiranja podatkov v analizi omrežij: Magistrsko delo. FDV, Ljubljana.
- Hubert, L.J., Baker, F.B., 1978. Evaluating the conformity of sociometric measurements. *Psychometrika* 43 (1), 31–41.
- Kemery, E.R., Dunlap, W.P., Griffith, R.W., 1988. Correction for variance restriction in point-biserial correlations. *Journal of Applied Psychology* 73 (4), 688.
- Lorrain, F., White, H.C., 1971. Structural equivalence of individuals in social networks. *The Journal of Mathematical Sociology* 1 (1), 49–80.
- Nordlund, C., 2007. Identifying regular blocks in valued networks: A heuristic applied to the St. Marks carbon flow data, and international trade in cereal products. *Social Networks* 29 (1), 59–69.
- Nordlund, C., 2016. A deviational approach to blockmodeling of valued networks. *Social Networks* 44, 160–178.
- Nordlund, C., 2018. Power-relational core-periphery structures: Peripheral dependency and core dominance in binary and valued networks. *Network Science* 6 (03), 348–369.
- Sailer, L.D., 1978. Structural equivalence: Meaning and definition, computation and application. *Social Networks* 1 (1), 73–90.
- Wasserman, S., Faust, K., 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- White, D.R., Reitz, K.P., 1983. Graph and semigroup homomorphisms on networks of relations. *Social Networks* 5 (2), 193–234.
- White, D.R., Reitz, K.P., 1985. *Measuring role distance: Structural, regular and relational equivalence*. Unpublished manuscript. University of California, Irvine.
- White, H.C., Boorman, S.A., Breiger, R.L., 1976. Social structure from multiple networks. I. Blockmodels of roles and positions. *American Journal of Sociology* 730–780.
- Williams, C.R., 1990. Deciding when, how, and if to correct turnover correlations. *Journal of Applied Psychology* 75 (6), 732.
- Žiberna, A., 2007a. Generalized blockmodeling of valued networks. *Social Networks* 29 (1), 105–126.
- Žiberna, A., 2007b. Generalized blockmodeling of valued networks. Doctoral dissertation. Retrieved from <http://www2.arnes.si/~aziber4/blockmodeling/Dissertation-final-corrected.pdf>.