Absent Ties in Social Networks, their Treatments, and Blockmodeling Outcomes

Anja Žnidaršič[1], Patrick Doreian[2], and Anuška Ferligoj[3]

Abstract

An absent tie is one for which we have no information regarding its nature. Absent ties for a network is a set of such ties. This lack of information can be present anywhere in network data and has the potential to compromise the results of all network analytic tools. To assess this impact, we used real networks and based simulations on them by introducing varying amounts of absent ties. They were treated with four treatments of absent ties. Blockmodeling, using structural equivalence, was applied to the known networks and then to every treated network. The results were compared. The amount of absent ties, their treatments, the block structure of a network, and the level of reciprocity all have an effect of the adequacy of the results of blockmodeling. Reconstruction combined with imputation based on modal values was the best overall treatment. However, treatments of absent ties can work for some networks but not others and we recommend treatments of absent ties based on the form of networks.

1 Introduction

Social networks are fundamental to social life and much is gained by studying them. A social network is a finite set (or sets) of actors and a social relation (or relations) defined over them. One goal of social network analyses is to detect, from network data, simple and useful descriptions of the fundamental underlying structures of networks. One widely used technique for finding such structural patterns is blockmodeling (Doreian et al., 2005). It is necessary to consider the issue of blockmodeling results being affected by measurement errors.

Wasserman and Faust (1998) distinguished five different techniques for gathering social network data: questionnaires, interviews, observations, archival records, and experiments. Marsden (2005, pg. 10) emphasized that "in the absence of archival records, surveys are often the most practical alternative: they make much more modest demands on participants than do diary methods and observation". Questionnaires are most commonly used when actors are people (Wasserman and Faust, 1998; Marsden, 2005) and

---

[1] Faculty of Organizational Sciences, University of Maribor, Kidričeva 55a, 4000 Kranj, Slovenia; anja.znidarsic@fov.uni-mb.si

[2] Department of Sociology, University of Pittsburgh, 2G03 Forbes Quad, Pittsburgh PA 15260, USA and the Faculty of Social Sciences, University of Ljubljana, Slovenia; pitpat@pitt.edu

[3] Faculty of Social Sciences, University of Ljubljana, Kardeljeva pl. 5, 1000 Ljubljana, Slovenia; anuska.ferligoj@fdv.uni-lj.si

are often in a form of a roster where respondents have to recognize the network members rather than recall them. Huisman (2009) emphasized that absent data on particular ties may occur especially in whole network studies where rosters are used as a questionnaire format.

We limit this study to binary networks where relational ties between two actors are present or not present. In terms of the adjacency matrix representation of a network, the corresponding elements are either 0 or 1 where a 0 represents a 'null tie' and a 1 represents a 'tie' that is present. The network is composed of ties *and* null ties. When we use the term 'absent tie', for a particular tie, we mean that we have no information regarding it. It could be a tie or it could be a null tie: we simply do not know which[2]. Therefore, we use the term *absent tie* as a shorthand for a 'tie for which we have no information'. If data are gathered with surveys, there is the term *tie (or item) non-response* and it occurs when an actor participates in the study but provides no information on particular tie(s). Information is absent when respondents provide no indication regarding the presence or absence a particular tie or ties (Rumsey, 1993; Borgatti and Everett, 2006; Huisman and Steglich, 2008; Huisman, 2009).

Here, we distinguish absent ties from *actor non-response* (Stork and Richards, 1992; Costenbader and Valente, 2003; Kossinets, 2006; Knoke and Yang, 2008; Huisman, 2009, Žnidaršič et al., 2012) which can be easily recognized because it is manifest as a row(s) of absent ties in the relational matrix. We emphasize that actor non-response is more frequent than absent ties when data are gathered by surveys, while with other data collection techniques actor non-response is less frequent.

The partial information for the incompletely observed actors are available and should be used to "obtain (better) estimates of the structural properties of the actors and the network, and may give information on the nature of the missing data mechanism" (Huisman, 2009, pg. 3). Rumsey (1993) observed that it is common practice for researchers of social networks to discard information about actors who do not respond or respond only partially. As a result, they were unable to find actual data sets reporting any type of absent data. This suggests that the phenomenon of absent ties is under-reported seriously in the literature.

Huisman (2009) emphasized that if there is no information about a network tie, there is limited capacity to describe the network context of the missing relationship between two actors and also the context of their neighbours. He showed that absent ties can have large effects on the structural properties of social networks (degree, reciprocity, transitivity, assortativity, inverse geodesic distance) where absent ties are even slightly more biased compared to actor non-response. Similarly, Borgatti and Everett (2006) emphasized that absent ties "can lead to a radically different understanding of the network and misleading measurements of network indices such as centrality".

De Leeuw et al. (2003, pg. 158) distinguished three forms of item non-response in classical surveys: (i) information is not provided by a respondent for a certain question (e.g. a respondent refused to respond for the item or overlooked a question); (ii) information provided by a respondent is not usable (e.g. the answer is not readable, could not

---

[2]We do not use the term of 'missing tie' because one connotation of this term is that the tie is missing. Instead, what is missing is information regarding the nature of that tie regardless of whether is a tie or a null tie.

be coded, or is out of the range of possible answers); and (iii) information is lost (usable information is lost in data entry).

Here, we focus on the impact of absent ties (or absent ties in case of survey data) for the outcomes of blockmodeling procedures. Our broad study design takes empirical networks whose blockmodel structure is known, introduces different amounts of absent ties, and treats them in several ways, applies blockmodeling to these treated networks, and compares the resulting blockmodeling with the blockmodel of the initial networks. This is done multiple times for each network to study the impact of both absent ties and the treatments of absent ties.

2   Treatment of absent ties

We focus our attention on absent ties and assume that there is no actor non-response[3]. In contrast to actor non-response (Žnidaršič et al., 2012), the network matrix for absent ties is characterized by (potentially) having absent ties (where we do not know their values) in all rows of the matrix. As the goal of this paper is to study the impact of treatments of absent ties on the stability of blockmodeling, the complete case approach (where both row and corresponding column are deleted) is not applicable and we excluded it from this study[4].

The four treatments incorporated in the study are the following. A *null tie imputation* procedure can be used where an unobserved tie is recorded as zero. It seems that this may well be the most commonly used procedure if only by default. To the extent that the data collection process is inattentive to explicitly recording absent data, the absent data will be recorded as 0. When a tie, r_{ij}, is missing and the modal value of incoming ties of actor j is used, this is *imputations based on modal values (of incomming ties)* which, in the figures below, is labeled simply as *modal values*[5]. The use of the *reconstruction* procedure for absent ties takes the form of replacing the unreported tie, r_{ij}, with then observed tie, r_{ji} (Stork and Richards, 1992; Huisman, 2009). If both ties r_{ij} and r_{ji} are unobserved, the reconstruction procedure is not possible. In the simplest case, a zero is imputed (and this treatment is labeled *reconstruction* in the figures reported below). Finally, *reconstruction plus imputations based on modal values* can be used (and this treatment is labeled *reconstruction plus modal values* in the figures). These four simple treatment methods can be used and all amount to adapting (or restricting) the treatments of actor non-response to deal with absent ties.

---

[3]Tackling the presence of both actor non-response and absent ties is a future agenda item.

[4]In surveys, if there are missing data for a small number of cases out of hundreds or thousands of cases, then case deletion is a reasonable but not perfect data analytic response. However, for small networks this is no longer the case. The absent data are about ties between actors and can be distributed across the entire network. They are not missing values on a single variable leading to cases being dropped one at a time. Rather, they are absent data about pairs of actors and dropping case removes both actors. Complete case deletion is likely to remove a high proportion of cases.

[5]Of course, especially in networks with high number of outgoing ties also outgoing ties could be used in imputations. In networks with small numbers of outgoing ties (in relation to the size of networks) the imputations based on modal value of outgoing ties will be the same or very similar to the null tie imputations. In our set of treatments we decided to use the modal values of incoming ties because popular individuals with many contacts are more likely to be chosen by their friends (Feld, 1991). Therefore, the incoming ties seem to be more relevant than number of outgoing ties for predicting or imputing absent ties.

3   Blockmodeling and indices for comparing two block-models

The goal of blockmodeling is to delineate the underlying structure of a network by transforming it to a smaller and more comprehensible form (Batagelj et al., 2004). This transformation is achieved by partitioning the network actors into clusters (subsets called *positions*), and, at the same time, partitioning the set of ties into *blocks* which determine the ties between clusters (positions) (Wasserman and Faust, 1998; Doreian et al., 2005). A block is defined as the set of relational ties between two clusters of actors. The actors within a cluster should have the same (or a similar) pattern of ties to each other and to actors of the other clusters. Given a selected equivalence, the result of blockmodeling is a compact representation of a network in the form of a blockmodel which delineates the essential structure of a network and is represented by a reduced graph or by an image matrix. The units in the reduced network are positions (composed of equivalent actors) and arcs in a reduced graph represent relations between positions (Doreian et al., 2005)[6]. The potential hazard with absent ties is that their presence can affect the composition of positions and change the characterization of blocks.

Here, we focus on blockmodeling of binary networks based on structural equivalence using direct methods (Batagelj et al., 1992b; Doreian et al., 2005). Actors are structurally equivalent if they are connected in exactly the same way to same neighbors (Lorrain and White, 1971; Faust, 1988), a formal definition is presented in Doreian et al. (2005, pg. 172). Batagelj et al. (1992b) proved that there are just two possible (ideal) blocks consistent with structural equivalence: null and complete blocks.

Generalized blockmodeling uses a direct approach where the optimal partition(s) is (are) identified based on minimal values of a compatible criterion function defined by the difference between empirical blocks and corresponding ideal blocks. The criterion function was first presented in in Batagelj et al. (1992a,b) and extended in Doreian et al. (2005, pg. 185-187, 223-226). When the value of a compatible criterion function is 0, then the obtained network partition perfectly matches the selected equivalence. In most empirical situations, the minimized value of the criterion function exceeds 0. These blockmodeling procedures have been implemented in the program Pajek (Batagelj and Mrvar, 2012), and in the R-package called Blockmodeling (Žiberna, 2008). Both were used in this study.

The concepts of networks and blockmodeling are used in constructing Figure 1. In the left panel of Figure 1 is a network presented by a graph where vertices represent boys and girls in a classroom and ties represent a liking relationship between them (Doreian et al., 2005, pg. 237). These data were collected by a student in a social network analysis class at the University of Pittsburgh for a class assignment. The data come from a pre-school where the student was a volunteer. The measured relation was 'plays with' and was constructed by observing play activity and later interpreted as 'liking'. Ties with an arrowhead are arcs that represent directed ties and ties without arrowheads are reciprocated (symmetric). There are many reciprocated ties and the reciprocity measure is 0.79. Also clear is a gender based split of the children. Using structural equivalence, the same

---

[6]The blockmodeling concepts of partitions and blocks can be viewed also in terms of positions and roles (Faust and Wasserman, 1992; Ferligoj et al., 2011).

two groups are obtained (where circles and squares represent boys and girls, respectively). The fitted blockmodel, with complete blocks on the diagonal and null blocks off the diagonal are shown on the right of Figure 1. The middle panel shows the matrix array and it has 12 inconsistencies, all of which are ties missing according to the ideal complete blocks.
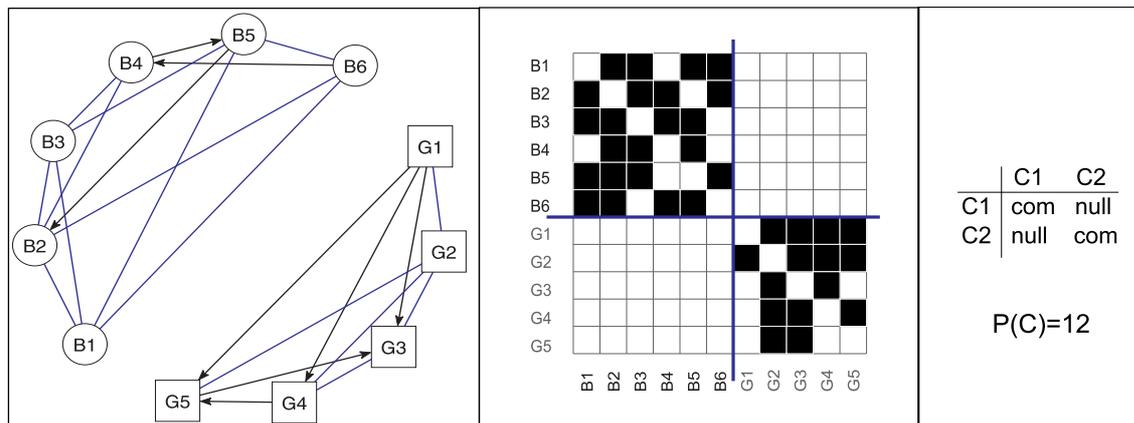

Figure 1: The boy-girl liking network (left), two clusters based on structural equivalence (middle), and image matrix (right).

The concepts of blockmodeling of binary networks can be extended to signed networks (Doreian and Mrvar, 1996, 2009), valued networks (Žiberna, 2007) and to 2-mode (or more) networks (Borgatti and Everett, 1992; Doreian et al. 2004; Mrvar and Doreian 2009; Doreian et al. 2012).

The focus of this study is on the impact of different treatments of absent ties on the identified blockmodel structures. We start with a whole (or known) network, impose different regimes of absent ties on it, treat the resulting absent data with the four different treatments described above, establish blockmodels of these 'treated' networks and compare the blockmodel structures for the known and treated networks by two indices.

The first index to compare two blockmodels is the Adjusted Rand Index (ARI) which measures the concordance between two partitions in terms of their composition (Hubert and Arabie, 1985). The Adjusted Rand Index has an expected value zero and maximal value one. Therefore, the lower the ARI measure, the worse is the correspondence of the position memberships for two partitions. Steinley (2004), based on an extended simulation study, provided guidelines for assessing the correspondence of two partitions of a set of units. For $ARI \geq 0.9$, the correspondence between partitions is excellent. When $0.9 > ARI > 0.8$, the correspondence is acceptable and the memberships of the two partitions are deemed to be close enough be taken as the same. For lower values of ARI, the correspondence is unacceptable. In terms of what we study here, the comparison is between the position membership of a known blockmodel with the position memberships of the blockmodel of the treated networks. For each network we consider, for all amounts of absent ties, and all treatments, we average the values of ARI and denote this by $mARI$. When $mARI \geq 0.9$ the blockmodel results are *fully stable* and when $0.9 > mARI > 0.8$, the correspondence is acceptable and blockmodel results are *stable*. But when $mARI < 0.8$ then the positions of the known blockmodel and the treated

blockmodel are not the same and the blockmodel results are *unstable*. Blockmodeling results cannot be accepted for configurations where this holds.

The second index, one that focuses on the form of the blockmodel, is the proportion of incorrect block types denoted by $ErrB$. It compares the distribution of the block types in the image matrices of the known and treated networks. If the two blockmodels agree completely then $ErrB = 0$. Otherwise $ErrB > 0$ and, somewhat arbitrarily, we take the two blockmodels to be essentially the same when $ErrB$ is below 0.2. Once this threshold is reached, then the blockmodel structure of the treated network cannot be accepted as an adequate surrogate for the blockmodel of the known network. For each set of simulations we denote the average value of $ErrB$ by $mErrB$.

These indices focus on different features of a blockmodel. In terms of characterizing, and understanding, the locations of actors, $mARI$ is the more consequential measure. As far as delineating the overall structure of the network is concerned, $mErrB$ is more important. To the extent that we want to consider *both* criteria, as we must if our goal is to use the blockmodel to interpret the behaviour of actors given the blocks, then a poor performance on one of these measure means that the blockmodel of the treated network is unacceptable.

4   Simulations based on four real networks for absent ties

The results of an extensive study of the impact of actor non-response treatments on the stability of blockmodels on real and simulated network data were presented in Žnidaršič et al. (2012). The main finding was that the selection of the best actor non-response treatment depends on the level of symmetry of the network. This is measured directly by reciprocity defined as $reciprocity = \frac{2 \cdot M}{M+A}$, where M indicates the number of mutual dyads and A the number of asymmetric dyads (Huisman, 2009).

Based on the extended simulations on real and artificial networks, two broad recommendations were made. First, (i) for symmetric (or largely symmetric) networks, the best treatments are reconstruction and a combination of reconstruction with imputations based on modes, and (ii) for non-symmetric (or largely non-symmetric) networks, the best approaches are complete case and imputations based on modes. The treatments that are the best for symmetric networks perform the worse in the case of non-symmetric networks and vice versa.

The second important recommendation was not to use either null tie imputations nor the complete case approach for networks with high levels of symmetry. Null tie imputation is the simplest treatment but it is highly likely to destroy both partition memberships and the blockmodel structure. When the complete case approach is used, we lose all of the partially observed data (between respondents and non-respondents).

Here, we examine the impact of absent ties and the effectiveness (or not) of treatments of absent ties on the stability of the blockmodeling results. We use four real networks with different levels of reciprocity. We start with a network having the highest reciprocity value (0.79) and finish with a network having the lowest reciprocity value (0.26).

The detailed outline of our simulation study is straightforward: (i) take a real whole network; (ii) establish the blockmodel of this real 'known' network; (iii) select a percentage of ties as absent ties (including both zeroes and ones); (iv) mark this percentage

of randomly selected ties as absent ties; (v) treat the absent ties with each of four treatments presented in Section 1; (v) establish the blockmodel of the 'treated' networks; and (vi) compare the 'known' and 'treated' blockmodels using the stability indices $mARI$ and $mErrB$ presented in Section 3. The percentage of absent ties ranges from 1% to 50%. For each starting real network and combinations of percent of missing ties and non-response treatments, the generation of absent ties was repeated 100 times.

4.1 The boy-girl liking network

The results of simulation study for absent ties for the boy-girl liking network (Figure 1) are presented in Figures 2(a) and 2(b). On the left, the mean values of ARI ($mARI$) are plotted against the percentage of absent ties for each of the four treatments. When there is 15% or fewer absent ties, all four methods lead to results that are fully stable for the correspondence of partitions, $mARI = 1$ for all treatments. However, when the incidence of absent ties is above 15%, there are two distinct pairs of trajectories. By far, the best treatments are reconstruction (marked with circles) and the combination of reconstruction and modal imputations (marked by triangles). These two trajectories drop together slowly. Letting q denote the percentage of absent ties, for $1 \leq q \leq 30$, $mARI$ is 1 or close to 1 showing that not only are the blockmodels of the treated networks close to those of the original network, they are most often identical. In short, the correspondence is excellent. Even for $30 < q < 43$, the values of $mARI$ remain in the acceptable range[7].

In contrast, for $q > 15$, the other two treatments, null tie imputations (marked with plus signs) and imputations based on modal values of incoming ties (marked with squares), begin to deteriorate. For $q > 15 > 24$, the values of $mARI$ remain above 0.8. However, once $q > 25$ the values of $mARI$ drop precipitously, especially so for imputations based on modal values. Neither treatment method works well once there is 25% or more absent ties.

Turning to the stability of blockmodels delineated, measured by the proportion of incorrectly identified block types, a similar set of results emerge (See Figure 2(b).) Regardless of the amount of absent ties (up to 50%), both reconstruction and the combination of reconstruction and imputations based on modal values perform very well. Not only are their values of $mErrB \leq 0.2$ throughout the range, they are often close to zero. At about 26% of absent ties, imputations based on modal values lead to unacceptable blockmodels and null tie imputations follows suit soon thereafter.

4.2 The Transatlantic Industries (TI) team network

The second example we consider, the Transatlantic Industries (TI) team network, has a reciprocity value 0.54. TI is one of two successful Little League baseball teams of boys reported by Fine (1987). The boys were asked to name three their best friends in a team. This network was extensively studied in terms of generalized blockmodeling by Doreian et al. (2005, pg. 196-199, 216-220). Starting on the left of Figure 3(a) we show the network, the formatted matrix array and the blockmodel (featuring structural equivalence) that was used. The formatted array is based on the best obtained blockmodel. The value

---

[7]Note that, for this network, 43% of randomly absent ties amounts to 47 ties out of 110 possible ties.
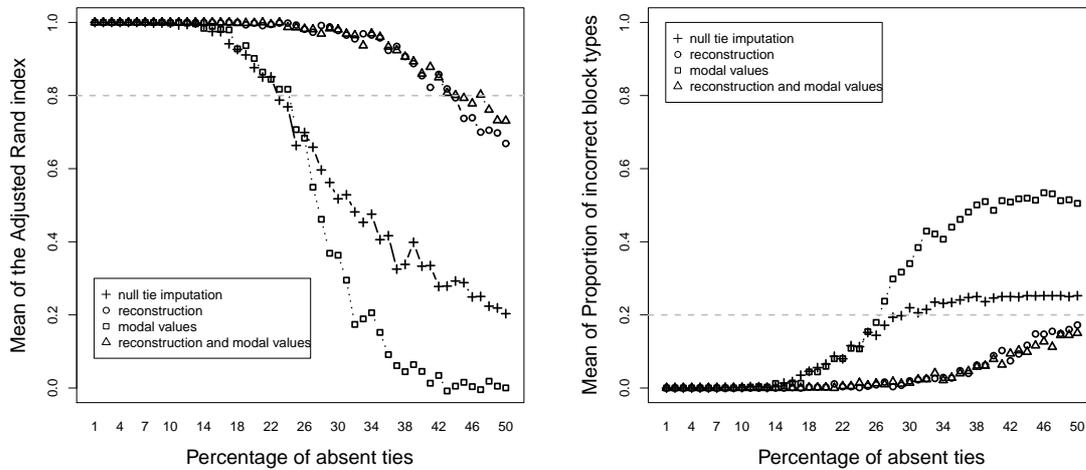
(a) Mean of the Adjusted Rand Index, $mARI$ (b) Mean of Incorrect block types, $mErrB$

Figure 2: Results of the simulation study for absent ties on the boy-girl liking network.

of the criterion function is equal to 29 (composed of 27 ties in null blocks and the absence of 2 ties in the complete block).

Figure 3(b) shows the results for agreement between partitions for the TI network. Even for 1% of absent ties, none of the treatment methods lead to $mARI = 1$. All four trajectories drop immediately for $q > 1$ and, by $q = 7$, all trajectories have dropped below the threshold value of 0.8. Although there are minor differences, all four treatment methods fail for levels of absent ties above 7%.

In contrast, the results regarding the structure of the blockmodel are, in general, very good (see Figure 3(c)). For all four treatment methods, the value of $mErrB$ is 0 (perfect correspondence) or close to 0 for absent ties levels up to about 15%. Once $q > 15$, the performance of imputations based on modal values deteriorates and when the percentage of absent ties reaches about 25%, its performance is unacceptable. According to this ($mErrB$) criterion, the performance of both reconstruction and null tie imputations are exceptionally good. And until $q = 34$ is reached, the combination of reconstruction and modal imputations performs equally well. Even at higher levels of absent ties, the values of $mErrB$ remain well below the threshold of 0.2.

The contrasting performance of $mARI$ and $mErrB$ for this network is instructive. In terms of locating actors in positions, the blockmodel results are very sensitive to absent ties. Any attempts to interpret actor memberships in positions are likely to be poor except when there are very modest levels of the number of absent ties. However, if the primary focus is on identifying the blockmodel structure of the network, then the results of block-modeling are very stable following the use of three of the absent ties treatment methods considered here.
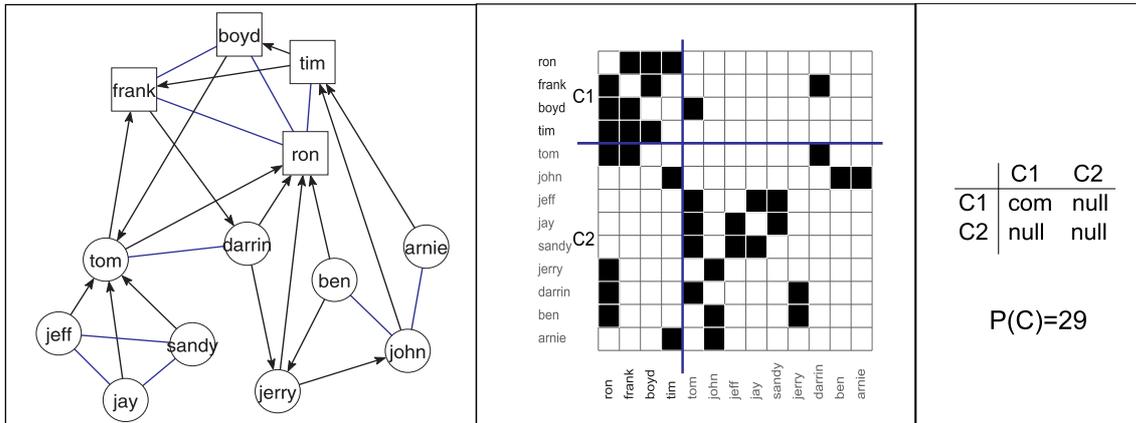
(a) The TI network (left), two clusters based on structural equivalence (middle), and image matrix (right)
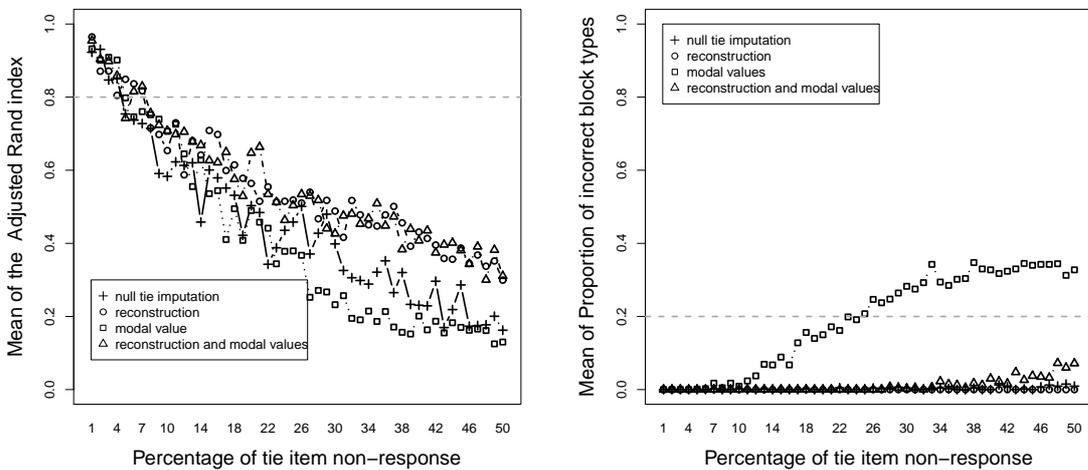
(b) Mean of the Adjusted Rand Index, $mARI$

(c) Mean of Incorrect block types, $mErrB$

Figure 3: Results of the simulation study for absent ties on the Transatlantic Industries network.

4.3 The note borrowing network

The third (note borrowing) network has a reciprocity value of 0.46. The data were collected by Valentina Hlebec from 15 undergraduate students attending lectures of a course (Hlebec and Ferligoj, 2002) and were used by Batagelj et al. (2004) for blockmodeling. The students were asked (without limitations on the number of nominations) "from whom would you borrow learning materials". The note borrowing network is presented in left panel in Figure 4(a). The shapes are used to depict the three clusters based on structural equivalence. A circle indicates membership in cluster $C1$, a triangle shows membership in $C2$ and a square indicates membership in $C3$. The best obtained blockmodel is shown on the right in with the formatted network array in the middle. The fitted blockmodel has 28 inconsistencies.

The results of our simulation study with non-symmetric blockmodel structure are presented in Figures 4(b) and 4(c). For levels of absent ties up to about 12% all four methods

return acceptable results regarding cluster (position) composition. With a higher incidence of absent ties, null tie imputations no longer performs adequately and it is the worst of the four treatment methods. At about a level of 16% absent ties, reconstruction also fails, and shortly thereafter, as q rises, the results following the use of the other two methods are not acceptable.
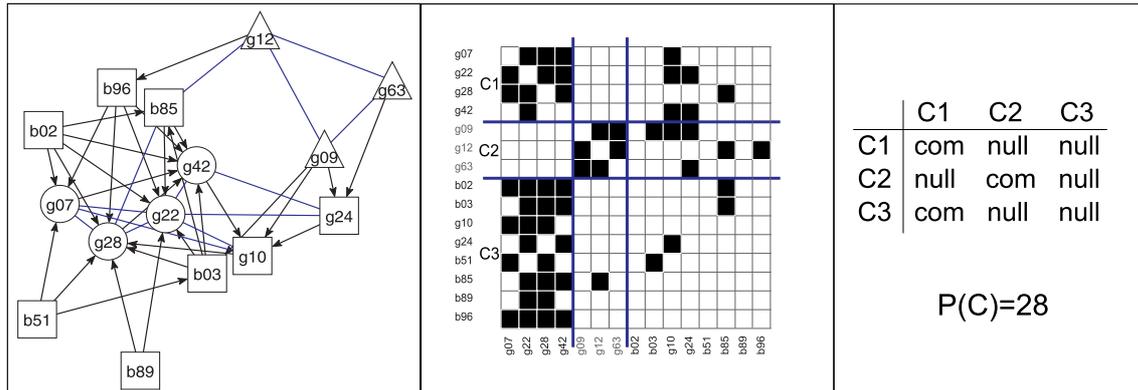
(a) The note borrowing network (left), three clusters based on structural equivalence (middle), and image matrix (right)
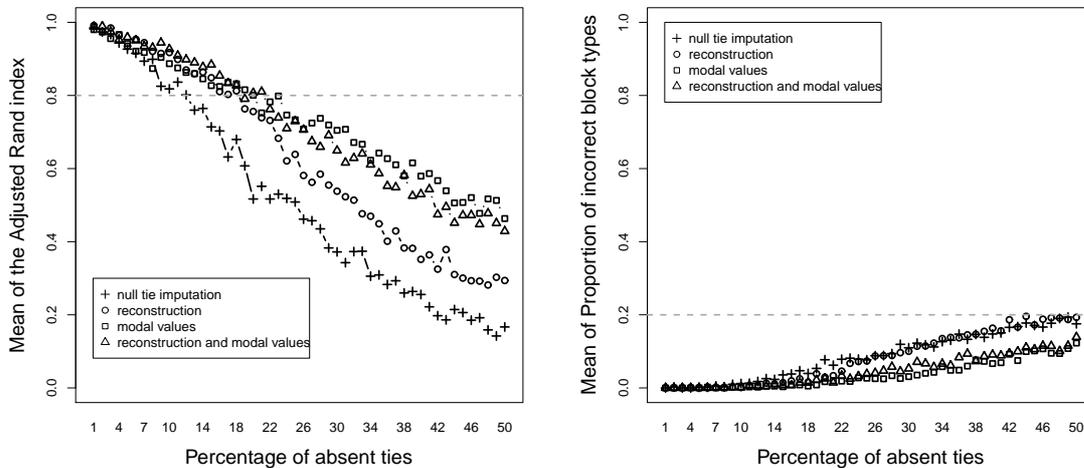

(b) Mean of the Adjusted Rand Index, $mARI$

(c) Mean of Incorrect block types, $mErrB$

Figure 4: Results of the simulation study for absent ties on the note borrowing network.

All four treatment methods perform extremely well for $q < 10$ and the mean values $mErrB$ are 0 or close to 0. As q increases above a level of 10% absent ties, $mErrB$ rises but remains below 0.2 for the whole range of introduced absent ties (see Figure 4(c)). However, the best two treatments are imputation based on modal values and the combination of reconstruction and modal imputations.

4.4 The Sharpstone Auto (SA) team network

The last network that we consider also comes from the baseball Little League collection reported by Fine (1987). This network, for the Sharpstone Auto (SA) team has the lowest value of reciprocity, 0.26, considered here. The network is presented in the left panel of Figure 5(a). The blockmodel based on structural equivalence into two positions (with three and seven actors) has 17 inconsistencies. The image matrix with two complete blocks and two null blocks shows a core-periphery structure.
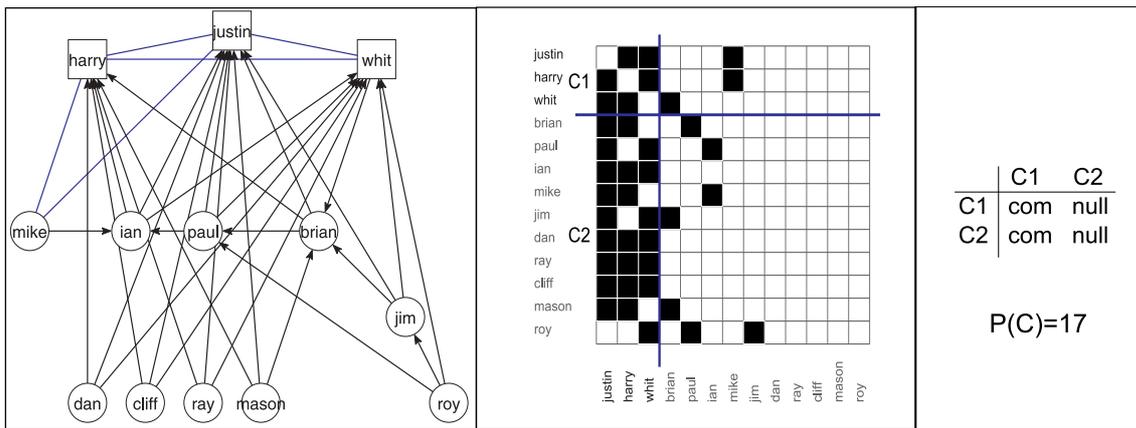
(a) The SA network (left), two clusters based on structural equivalence (middle), and image matrix (right)
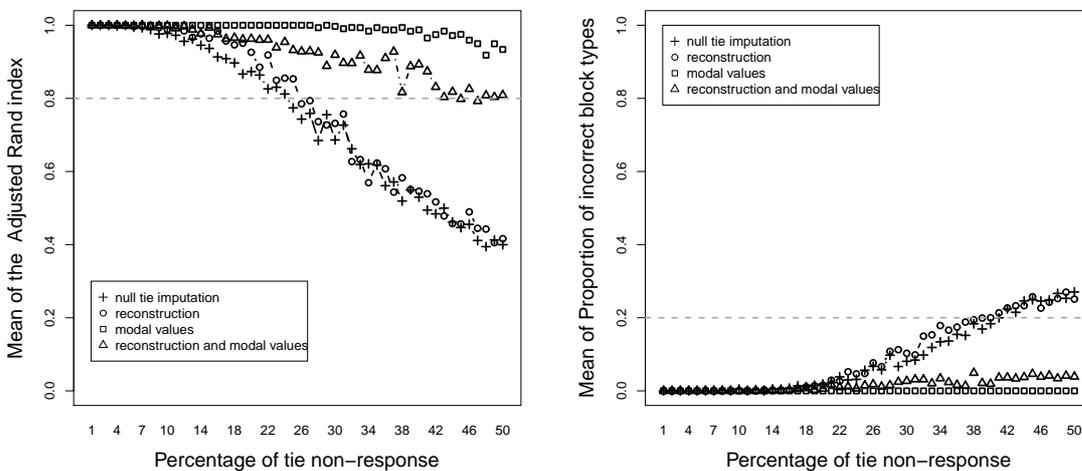

(b) Mean of the Adjusted Rand Index, $mARI$

(c) Mean of Incorrect block types, $mErrB$

Figure 5: Results of the simulation study for absent ties on the Sharpstone Auto network.

Figures 5(b) and 5(c) present the simulation results of the simulations based on this network. The best performance, as far comparing the composition of clusters is concerned, follows the use of imputations based on modal values of incoming ties. Throughout half of the range for imposed absent ties, the values of $mARI$ are 1, indicating perfect correspondence. Only towards the high end of the range for the percentage of absent ties $mARI$ drops below this value but it still remains well above the 0.8 threshold for stable

blockmodels. The second best performance follows the use the combination of reconstruction and modal imputations although for high levels of percentage of absent ties ($q > 34$), it is barely adequate. The other two treatment methods, null tie imputations and reconstruction, fail for $q > 0.25$.

The agreement between image matrices is perfect for imputations based on modes ($mErrB = 0$) throughout the whole range of percentage of absent ties. The performance following the use of the combination or reconstruction procedure and modal imputations is excellent over the whole range of percentage of absent ties with $mErrB$ values always below 0.05. The null tie imputations and reconstruction procedure produce perfect agreement for 16% of absent ties or less, then the $mErrB$ values start to increase and at 40% of missing ties $mErrB$ values cross the 0.2 threshold and their performance becomes unacceptable.

4.5   A summary across the four simulation studies

The results for the Sharpstone Auto network are the easiest to interpret. The underlying true network has a clear core-periphery structure. This is reflected by two inter-related features: (i) the columns are either mostly ones or mostly zeros and (ii) the two complete blocks are dense. It is not surprising that imputation based on the modal values is the best treatment because the modal values of the columns are clearly one or zero with little ambiguity. As far as identifying the blockmodel structure, this treatment method is perfect in its performance, even with the highest level ($q = 50$) of absent ties. With regard to identifying the composition of the positions, it performs almost as well. The combination of reconstruction and using modal values performs the next best for both $mARI$ and $mErrB$. One obvious recommendation is that if the true network is thought to have a core-periphery structure, and absent ties are suspected, these two methods of treating absent ties are preferable. This network structure is one where reciprocity does not seem that important (except for ties within the core position).

For the boy-girl network, the results are straightforward also. Reciprocity is important for this network in the sense that there are two well separated positions whose internal ties are dense. For levels of up to about $q = 25$, it does not matter which treatment for absent ties is used. Both position membership and blockmodel structure will be fully stable. Thereafter, the best two treatment methods are reconstruction and the combination of reconstruction with the use of modal values and lead to stable blockmodeling results (although performance diminishes as $q > 35$). The null tie imputation gets worse with higher values of q because null ties are inserted in the complete blocks. Imputations based on modes also fails but for slightly different reasons. The ties are concentrated in nearly complete blocks and using the imputation based on modal values will locate ties in otherwise null blocks. The null ties are also concentrated but in null blocks and this imputation treatment will put null ties in otherwise complete blocks. All blocks will become more similar as q gets larger and neither the position membership nor the block model structure will be stable. The best two absent ties treatments are reconstruction and the mixture of reconstruction with the use of modal values (where the latter plays a lesser role) when there are clearly separated blocks and few (or no) ties between actors in different positions.

The structure of the TI network is one that is particularly sensitive to the presence of

absent ties with regard the position composition. For $q > 7$, all treatments for absent ties lead to unstable blockmodel results regarding position membership. However, the blockmodel structure is fully stable or stable for all treatment methods except imputations based on modes. Remarkably, the results are stable even with a 50% level of absent ties. The one complete block is internally dense with the remaining ties scattered across the three null blocks. When some of these ties involve actors from the core this leads to the instability of position composition. Unless these are few absent ties, the safest recommendation is that any method other than the modal imputation will return the stable blockmodel structure and none of the treatments that we have considered can return stable results for position membership.

In some respects, the note borrowing network results are similar but less severe than those for the TI network (except that imputations based on the modal value) and do not fail for higher levels of q. In terms of getting the blockmodel structure correctly (in a stable fashion), the best treatment methods are imputation based on the mode and the combination of reconstruction with use of modes. In terms of position membership, when $q < 18$ all treatments other than null tie imputation are adequate.

The networks studied in Section 4 are small but the impacts of absent ties and their treatments are quite complex. Therefore, we decided to use ANCOVA to investigate the effects of percentage of absent ties or absent ties ($pTie$), absent tie treatment (T), reciprocity of the network (R), and size of the network (S).

Table 1 presents the ANCOVA results for both the Adjusted Rand Index (left panel) and the proportion of incorrect block types (right panel) where main effects and all interactions (two, and three-way) are ordered according to their partial η^2 values.

The highest effect on the Adjusted Rand Index is the percentage of absent ties ($\eta^2 = 0.3464$) and it is clear from the presented figures that higher percentage of absent ties leads to lower values of ARI which indicates poorer identification of position membership of actors. The second largest effect has reciprocity in combination with size of a network ($\eta^2 = 0.1577$), where the smallest network in our study has the highest reciprocity and its performance according to the position membership is the best compared to other networks in the study. The third largest effect has treatment in combination with reciprocity ($\eta^2 = 0.0998$), where both reconstruction treatments perform better for symmetric networks with higher reciprocity and imputations based on modal values are better for less symmetric networks with lower reciprocity. However, reciprocity alone has the lowest effect among all other main effects ($\eta^2 = 0.0135$). The fourth largest effect has a size of a network ($\eta^2 = 0899$) and this is also the largest main effect among. The position membership is better preserved in small networks.

The largest effect on the proportion of incorrect blocks ($ErrB$) is again the percentage of absent ties ($\eta^2 = 0.3450$) and its impact is similar to that for ARI. The second largest effect is the treatment of absent data in interaction with reciprocity, where use of reconstruction and combination of reconstruction with imputations based on modal values produce better results for networks with higher reciprocity. The second largest main effect is treatment of absent data ($\eta^2 = 0.1413$), while reciprocity has a lower effect ($\eta^2 = 0.0614$). The lowest main effect is size of a network ($\eta^2 = 0.0033$) which is, in fact, the lowest effect among all 15 effects. Tu summarize, the size of a network has larger effect on the position membership of actors than on the obtained blockmodel structure according to types and position of blocks.

Table 1: Analyses of covariance for the Adjusted Rand Index and the Proportion of incorrect block types with absent ties data.

Adjusted Rand Index				Proportion of incorrect block types			
Effect	Df_1	F	Partial η^2	Effect	Df_1	F	Partial η^2
pTie	1	42389	0.3464	pTie	1	42111	0.3450
R*S	1	14971	0.1577	T*R	3	7438	0.2182
T*R	3	2954	0.0998	pTie*T*R	3	5486	0.1707
S	1	7904	0.0899	T	3	4386	0.1413
pTie*T*R	3	2088	0.0726	T*R*S	3	3893	0.1274
T	3	2078	0.0723	pTie*T	3	2243	0.0776
pTie*T	3	844	0.0307	R	1	5233	0.0614
R	1	1092	0.0135	pTie*T*R*S	3	1457	0.0518
pTie*R	1	1069	0.0132	pTie*R	1	3791	0.0453
T*R*S	3	227	0.0085	pTie*R*S	1	2573	0.0312
pTie*S	1	582	0.0072	R*S	1	1603	0.0197
pTie*T*S	3	163	0.0061	T*S	3	467	0.0172
T*S	3	157	0.0059	pTie*T*S	3	352	0.0130
pTie*T*R*S	3	140	0.0052	pTie*S	1	680	0.0084
pTie*R*S	1	0.2	0.0000	S	1	261	0.0033

Residuals degrees of freedom - $Df_2 = 79968$
pTie - percentage of absent ties   T - treatment of absent ties
R - reciprocity of the networks   S - size of the networks

4.6   The impact of inserting inaccurate data

When an absent ties is treated, a tie value (0 or 1) is placed into its network matrix array. We know the true value of every tie in the known network that was marked as an absent tie. Therefore, we can assess the performance of each treatment of absent ties . When a treatment imputes the right value of a tie, the tie was recovered accurately. But if there is a change in the tie value, an inaccurate data value has been inserted. For each treatment procedure, each network and all levels of percentage of absent ties we kept track of the insertion of inaccurate data retrievals. Figure 6 plots the percentage of ties whose values were changed from the original values against the percentage of absent ties.

For cases where the absent ties treatments work well (as described below), for a 50% level of absent ties the levels of inaccurate data introduction are 25% or lower and suggest that at least 75% of the tie values are returned correctly. There are corresponding values for lower levels of absent ties and we pursue the impact of inaccurate data imputations.

For the BG (Boy-Girl) network, imputation based on modes introduced the highest levels of inserting inaccurate data and the problem gets worse as the level of percentage of absent ties (q) increases. See Figure 6(a). Null tie imputation has the second worst performance. For $q < 22$, reconstruction and the combination of reconstruction with imputation based on modes have the same performance. For $q > 22$, the combined treatment performs less well than reconstruction and the latter is clearly the best treatment as far as retrieving blockmodels is concerned. The higher levels of introduced inaccurate data for imputations based on the mode helps account for its poor performance. The same holds for null tie imputation. In terms of getting the positions right ($mARI > 0.8$) and getting the block structure right ($mErrB < 0.2$) the slight differences between the two
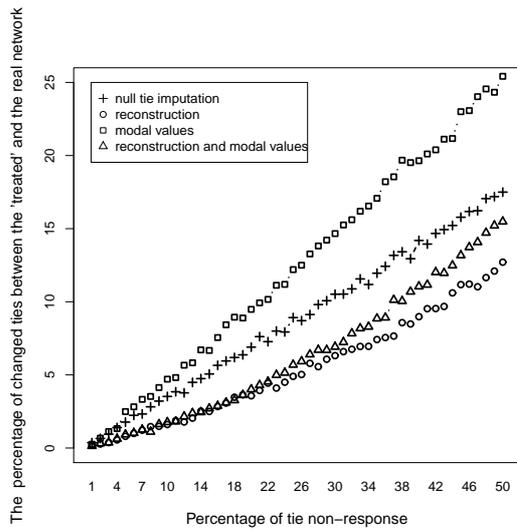
(a) The boy-girl liking ties network
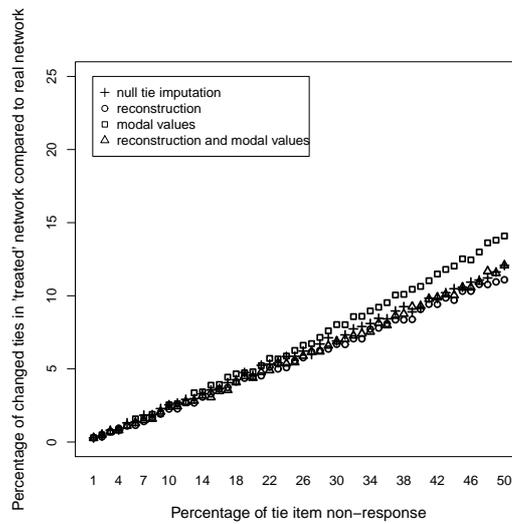
(b) Transatlantic Industries network
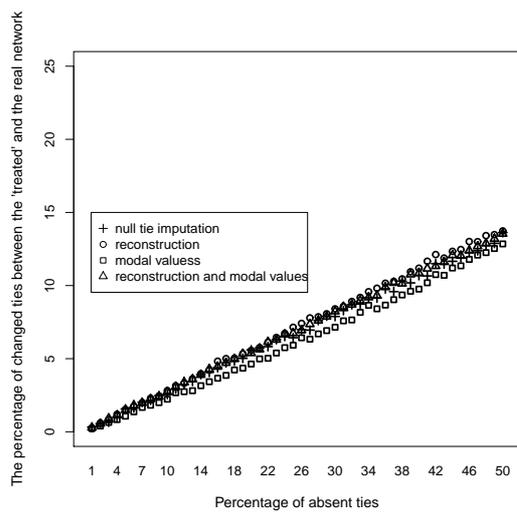
(c) The note borrowing network
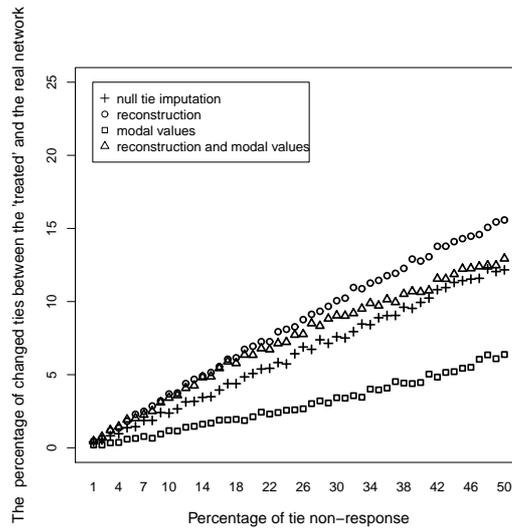
(d) The Sharpstone Auto network

Figure 6: The average percentage of changed ties in the treated network for absent ties compared to the whole network

best treatments seems inconsequential until very high levels of q are reached.

Compared to the BG network, the trajectories in Figure 6(d) for the Sharpstone Auto network show the reverse pattern. Imputation based on modes produces fewer data inaccuracies. Next is null tie imputation followed by the combined treatment. By a wide margin, especially for higher levels of q, the highest level of data inaccuracy comes from using reconstruction. For both $mARI$ and $mErrB$, the blockmodeling results are outstanding following the use of imputations based on modes and suggest that the lower levels of data inaccuracies helps account for this. For $q < 18$, the combined use reconstruction and imputation based on modes is the second best treatment with regard to blockmodeling results yet it is the second worst in terms of introducing inaccurate data.

This suggests that, in addition to the amount of data inaccuracy, it is necessary also to consider the pattern of inaccuracies across block types.

At face value, the trajectories in Figure 6(b) for the TI network look very similar, consistent the trajectories for $mARI$ in Figure 3(a) showing poor performance for all treatment methods. But for higher values of q, imputations based on modes introduces higher levels of data inaccuracies. The trajectory for $mErrB$ in Figure 3(a) suggests that this level of data inaccuracies is particularly damaging with regard to block type identification. The other three treatments all perform well of the identification of block types even for high levels of absent ties.

The trajectories of introduced data inaccuracies in Figure 6(c) for the note borrowing network show the closest correspondence across treatments. But this close correspondence is not reflected (for $q > 7$) in Figure 4 for $mARI$ and, to a lesser extent, for $mErrB$. Again, this suggests that the pattern of inaccuracies across block types merits attention.

5   Summary

We used the term *absent tie* as one for which we have no information regarding the nature of a tie regardless of whether it is present or not. This lack of information can be present in every row of a network adjacency matrix and we called the set of such ties *absent ties*. To the extent that researchers are inattentive to the presence of absent ties and record them as null ties, this is a major problem for social network analysis with the potential to compromise the results obtained from using most of the techniques available in the literature.

We considered four different known real networks and performed simulations based on them. Different amounts (from 1% to 50%) of absent ties were applied to these networks which were then treated with four treatments of absent ties : reconstruction, imputations based on the modal values of incoming ties; a combination of reconstruction with imputations based on the modal values, and null tie imputation. Regarding network methods, we focused on blockmodeling based on structural equivalence. Blockmodeling was applied to the known networks and also to each treated network. These blockmodeling results were compared. For each combination of network, amount of introduced absent ties, and treatment of absent ties, our simulations were based on 100 repetitions. Two criteria were used to compare the partition of the known network with the partitions of the treated networks: the correspondence of the partitions of networks into positions, measured by the Adjusted Rand Index, and the proportion of incorrectly identified block types in the blockmodels of the treated networks.

The percentage of absent ties, the treatments of absent ties, the block structure of a network, reciprocity of a network, and size of a network all have an effect of the stability of the results of blockmodeling. Despite this potential complexity, we draw the following conclusions:

(i) With regard to blockmodel structure, blockmodeling in terms of structural equivalences fares very well and the exceptions can be accounted for (in the other conclusions).

(ii) The combination of reconstruction and imputations based on modes is the best overall treatment method for absent ties. The blockmodel structure is correctly discovered in all four networks and the membership of positions is returned well for three of the four networks. The one exception is the TI network where the performance is borderline.

(iii) Both reciprocity and blockmodel structure matter in systematic ways. The results following the use of imputation based on modes are good when reciprocity is low but they are unacceptable for networks with high reciprocity. Imputations based on modes fares badly for core-periphery structures while reconstruction works well for them.

(iv) Null tie imputation is the worst treatment for absent ties and its use never succeeds with regard to obtaining correctly the membership of positions.

(v) The criteria of getting the position membership and the blockmodel structure correctly do not always lead to the same implications with regard to blockmodeling outcomes. In general, performances are better for the blockmodel structure than for position membership. Put differently, performance is better with regard to the macro-structure of the networks and worse with regard to micro-structural details. This leads us to think that all methods focusing primarily on micro-structure details are threatened also by absent ties.

Regardless the fact that different criterion functions are not completely comparable due to different network sizes the results clearly show that the value of criterion function (how good is the blockmodel structure) also affects the blockmodeling results when the absent ties are somehow treated. Blockmodels with larger values of the criterion function are less stable in the presence of absent ties.

6 Conclusion

The results reported in this paper have two direct broad implications. One is that blockmodeling results are vulnerable to the kinds of measurement error that we study. The second is that if we know something about the measurement error, it can be treated in ways that reduce the vulnerability. Of course, if nothing is known about the measurement errors they cannot be treated: therefore our practical recommendations are therefore limited. Even though network data are often collected in ways that preclude obtaining the required knowledge about the actual measurement errors, this does not imply that network data must be collected in this way. A third implication of our results is that instruments need to be designed to facilitate the collection of this type of knowledge. If classical data collection procedures preclude awareness of missing ties they need to be replaced.

There are some obvious additional open problems and we list five of them as having higher priority. It will be fruitful to consider:

(i) extending this type of evaluation to larger networks;

(ii) looking at different types of block patterns for structural equivalence starting from well constructed artificial networks;

(iii) extending this type of examination to other equivalence types and other blockmodel structures;

(iv) moving beyond considering random absent ties; and (v) looking at valued networks (including signed networks).

References

[1] Batagelj, V., Doreian, P., and Ferligoj, A. (1992a): An optimizational approach to regular equivalence. *Social Networks*, **14**, 121 - 135.

[2] Batagelj, V., Ferligoj, A., and Doreian, P. (1992b): Direct and indirect methods for structural equivalence. *Social Networks*, **14**, 63-90.

[3] Batagelj, V. and Mrvar, A. (2012): *Pajek 3.05 - program for Large Network Analysis*. Available at: http://pajek.imfm.si/doku.php?id=download.

[4] Batagelj, V., Mrvar, A., Ferligoj, A., and Doreian, P. (2004): Generalized Block-modeling with Pajek. *Metodološki Zvezki*, **1**, 455-467.

[5] Borgatti, S. P. and Everett, M.G. (1992): Regular blockmodels of multiway, multi-mode matrices. *Social Networks*, **14**, 91120.

[6] Borgatti, S. P. and Everett, M.G. (2006): A graph-theoretic perspective on centrality. *Social Networks*, **28**, 466-484.

[7] Costenbader, E. and Valente T.W. (2003): The stability of centrality measures when networks are sampled. *Social Networks*, **25**, 283-307.

[8] Doreian, P., Batagelj, V., and Ferligoj, A. (2004): Generalized blockmodeling of two-mode network data. *Social Networks*, **26**, 29-53.

[9] Doreian, P., Batagelj, V., and Ferligoj, A. (2005): *Generalized Blockmodeling*. Cambridge University Press, New York, NY.

[10] Doreian, P. and Mrvar, A. (1996): A partitioning approach to structural balance. *Social Networks*, **18**, 149-168.

[11] Doreian, P. and Mrvar, A. (2009): Partitioning signed social networks. *Social Networks*, **31**, 1-11.

[12] Doreian, P., Lloyd, P., and Mrvar, A. (2012): Partitioning large signed two-mode networks : problems and prospects. *Social Networks*, doi: http://dx.doi.org/10.1016/j.socnet.2012.01.002

[13] Faust, K. (1988): Comparison of methods for positional analysis: Structural and general equivalences. *Social Networks*, **10**, 313-41.

[14] Faust, K., and Wasserman S. (1992): Blockmodels: Interpretation and evaluation. *Social Networks*, **14**, 5-61.

[15] Feld, Scott L. (1991): Why your friends have more friends than you do. *The American Journal of Sociology*, **96**, 464-1477.

[16] Ferligoj, A., Patrick D., and Batagelj, V. (2011): Positions and roles, 434-446. *The SAGE Handbook of Social Network Analysis*, Thousand Oaks: Cambridge Sage Publications.

[17] Fine, G.A. (1987): *With the Boys: Little League Baseball and Preadolescent Culture.* Chicago, USA: University of Chicago Press.

[18] Hlebec, V, and Ferligij, A. (2002): Reliability of social network measurement instruments. *Field method*, **14**, 288-306.

[19] Hubert, L. and Arabie, P. (1985): Comparing partitions. *Journal of Classification*, **2**, 193-218.

[20] Huisman, M. (2009): Effects of missing data in social networks. *Journal of Social Structure*, **10**. Available at: http://www.cmu.edu/joss/content/articles/volume10/huisman.pdf.

[21] Huisman, M. and Steglich, C. (2008): Treatment of non-response in longitudinal network studies. *Social networks*, **30**, 297-308.

[22] Knoke, D. and Yang, S. (2008): *Social networks analysis.* Sage Publications,Los Angeles. 2nd edition.

[23] Kossinets, G. (2006): Effects of missing data in social networks. *Social networks*, **28**, 247-268.

[24] de Leeuw, E.D., Hox, J., and Huisman, M. (2003): Prevention and Treatment of Item Nonresponse. *Journal of Official Statistics*, **19(2)**, 153-176.

[25] Lorrain, F., and White, H.C. (1971): Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology*, **1**, 49-80.

[26] Marsden, P.V. (2005): Recent Developments in Network Measurement. In Carrington P.J., Scott J. and Wasserman S. (Eds.), *Models and Methods In Social Network Analysis*, 8-30. Cambridge University Press, New York.

[27] Mrvar, A. and Doreian, P. (2009): Partitioning signed two-mode networks. *The Journal of mathematical sociology*, **33**, 196-221.

[28] Rumsey, D. J. (1993): *Nonresponse models for social network stochastic processes.* Ph.D. thesis, The Ohio State University.

[29] Steinley, D. (2004): Properties of the Hubert-Arabie Adjusted Rand index. *Psychological Methods*, **9**, 386-396.

[30] Stork, D. and Richards, W.D. (1992): Nonrespondents in communication network studies: problems and possibilities. *Group and Organization Management*, **17**, 193-209.

[31] Wasserman, S. and Faust, K. (1994): *Social Network Analysis: Methods and Applications.* Cambridge University Press, New York, NY.

[32] Žiberna, A. (2007): Generalized Blockmodeling of Valued Networks. *Social networks*, **29**, 105-126.

[33] Žiberna, A. (2008): Blockmodeling 0.1.7: An R package for Generalized and classical blockmodeling of valued networks. Available at:
http://www2.arnes.si/ ∼aziber4/

[34] Žnidaršič, A., Doreian, P., and Feligoj, A. (2012): Tie non-response in social networks, treatments of tie non-response, and blockmodeling outcomes. *Social Networks*, doi: http://dx.doi.org/10.1016/j.socnet.2012.02.002