

Using Social Network Information for Survey Estimation

Thomas Suesse¹ and Ray Chambers¹

Model-based and model-assisted methods of survey estimation aim to improve the precision of estimators of the population total or mean relative to methods based on the nonparametric Horvitz-Thompson estimator. These methods often use a linear regression model defined in terms of auxiliary variables whose values are assumed known for all population units. Information on networks represents another form of auxiliary information that might increase the precision of these estimators, particularly if it is reasonable to assume that networked population units have similar values of the survey variable. Linear models that use networks as a source of auxiliary information include autocorrelation, disturbance, and contextual models. In this article we focus on social networks, and investigate how much of the population structure of the network needs to be known for estimation methods based on these models to be useful. In particular, we use simulation to compare the performance of the best linear unbiased predictor under a model that ignores the network with model-based estimators that incorporate network information. Our results show that incorporating network information via a contextual model seems to be the most appropriate approach. We also show that one does not need to know the full population network, but that knowledge of the partial network linking the sampled population units to the non-sampled population units is necessary. Finally, we also provide an estimator for the mean-squared error to make an informed decision about using the contextual information, as well as the results showing that this adaptive strategy leads to higher precision.

Key words: BLUP; social network models; linear models; model-based survey estimation.

1. Introduction

Survey estimation typically focuses on estimating the total $T_Y = \sum_{i \in U} Y_i$ of the values of a variable Y defined over a finite population U . Here $i \in U$ denotes the N units making up the population U . Given a sample s of n units from U , T_Y is usually estimated by $\hat{T}_Y = \sum_{i \in s} w_i Y_i$, where the w_i are sample weights and $i \in s$ denotes the n units in the sample. Traditionally, these weights are expansion weights, that is w_i is the inverse of the selection probability of the i th population unit. However, expansion weights can be quite inefficient, and alternative weighting methods derived from model-based and model-assisted methods of survey estimation, see Chambers and Clark (2012) and Särndal et al. (1992), are used to increase the precision of \hat{T}_Y . In most cases this is done by defining the sample weights so that \hat{T}_Y is an efficient unbiased predictor of T_Y under a linear regression model for Y in terms of a multivariate auxiliary variable \mathbf{X} .

¹ National Institute for Applied Statistics Research Australia and University of Wollongong, Northfield Avenue, Wollongong, New South Wales 2522, Wollongong, Australia. Emails: tsuesse@uow.edu.au and ray@uow.edu.au
Acknowledgments: We would like to thank NIASRA for providing the High Performance Cluster to run the simulations and the referees for their helpful comments that greatly improved this paper. We would also like to acknowledge financial support from Australian Research Council (ARC) (Grant number LX0883143).

Population regression models that link an individual's value of Y to auxiliary variables corresponding to that individual's geographic location, gender, and age are commonly used in survey estimation. However, auxiliary information can be more complex than this. In particular, information about other individuals in the population that are 'linked' to a particular individual also constitutes auxiliary information about that individual. This is sometimes referred to as *network* information, and typically indicates between individual correlation in the population values of Y . In this article we describe model-based survey estimation methods that exploit auxiliary information about population networks. In particular, we describe how the specification of the Best Linear Unbiased Predictor (BLUP) of T_Y can be tailored to allow for between individual correlation induced by the presence of a population network. Such correlation or association between individuals with similar characteristics is often referred to as *homophily* in the network literature.

In order to motivate the use of network information in survey estimation, consider the case of the British Household Panel Study (BHPS, <https://www.iser.essex.ac.uk/bhps/>). This is an annual longitudinal survey of British households that has been conducted since 1991. It is based on a sample of approximately 5,500 households, covering more than 10,000 individuals. The main objective of the survey is to further the understanding of social and economic change at the individual and household level in Britain. However, in addition to information about the surveyed individual, the BHPS also provides information about a person's three closest friends. Variables collected on the three closest friends are: age, sex, ethnicity, distance to friend (< 1 mile, between 1 and < 5 miles, between 5 and 50 miles, > 50 miles), and unemployment status. This information is available in seven waves, corresponding to the even-numbered years 1992–2004.

Because friends tend to share common characteristics, it is plausible that the BHPS information on friendship ties may be of value when modelling the other survey variables, in the same way as the ties between household members are typically viewed as influential in determining the outcomes of many social and economic variables. For example, a person whose friends are older than the norm might have a higher than average income, even after adjusting for that person's age and gender. As a consequence, one might think of also controlling for the average age of friends when predicting a person's income. A model of this type is referred to as a *contextual* model in what follows since it controls for contextual effects, such as the average age of friends. Clearly, since the BHPS collects information on a person's three best friends, there is scope for applying a contextual model when estimating using BHPS data. This might lead to more precise survey estimates, as a contextual variable represents an additional source of information.

The friendship data collected in the BHPS are a special case of a general type of auxiliary data whose availability is becoming increasingly widespread, especially with the rapid uptake of modern telecommunications technology. This is network data, defined by the existence, direction and strength of relationships between individuals in a population of interest. Statistical modelling of networks is now reasonably well established, see, for example Frank and Strauss (1986), Snijders (2002), Hunter and Handcock (2006), though applications to very large networks (e.g., defined by populations similar in size to those covered by a survey like the BHPS) are still rare, with data on very large networks now considered to be part of the ubiquitous Big Data concept. Furthermore, we are not aware of any attempt to use the information in a network defined on a population of interest to

improve survey estimation for that population, although, as the argument put forward in the previous paragraph indicates, there may be value in doing so.

In order to use network information in a model linking a survey variable Y to the auxiliary variable \mathbf{X} we need to characterise the population network as the outcome of a random process. In this context, we focus in this article on a network that identifies the existence and direction of a relationship between individuals in a population of size N . It is standard to represent such a network by a matrix of zeros and ones, $\mathbf{Z} = (Z_{ij})_{i,j=1}^N$ with $Z_{ii} = 0$ by convention. If a relationship exists between two individuals i and j , then $Z_{ij} = 1$ and we refer to i and j as being linked; otherwise $Z_{ij} = 0$. Such a network is said to be undirected if $\mathbf{Z} = \mathbf{Z}^\top$, otherwise it is a directed network.

Networks are most useful when characteristics of the individuals that make up the population covered by the network are also known. In such networks one not only knows the characteristics of a particular individual, but also the characteristics of the other individuals in the population linked to that individual via the network. This *external* auxiliary information may be useful in discriminating between individuals, and hence may be useful in prediction, the ultimate goal of survey estimation. For example, the BHPS collects information about the three best friends of a surveyed individual, without identifying the friends. Given that the links corresponding to being ‘one of three best friends’ define a network, this information can be treated as auxiliary data for the surveyed individual, and, combined with a model for the network, may help with formulating a more efficient prediction model for the population.

Linear models that use a social network as additional information to model the expected value of a response variable include contextual network (CN) models (Friedkin 1990). However, this information can also be used to model between unit correlation in the population values of the response variable. Such second order models include network effects models, also known as autocorrelation (AR) models, and network disturbance (ND) models (Ord 1975; Doreian et al. 1984; Duke 1993; Marsden and Friedkin 1993; Leenders 2002).

When the network defined by \mathbf{Z} is known for all N individuals in the population, the CN, AR and ND population models can be used for survey estimation. However, in practice it is extremely unlikely that \mathbf{Z} will be fully known, and a more realistic scenario is one where one or more components of this matrix will be known. The most obvious is where only the component \mathbf{Z}_{ss} corresponding to the sub-network of relationships between the n sampled individuals in s is known. Unless the sampling fraction is large, or the sample is highly clustered, it is unlikely that this sub-network will contain much useful information. Of more use, perhaps, is the component \mathbf{Z}_{sr} , defined by the links between the sampled individuals and the remaining $N - n$ non-sampled individuals in the population, denoted collectively by r . Clearly, if the network is an undirected one, the links from the non-sampled individuals to the sampled individuals will then also be known since, under symmetry, $\mathbf{Z}_{rs} = \mathbf{Z}_{sr}^\top$. The remaining component of \mathbf{Z} is \mathbf{Z}_{rr} , which corresponds to the sub-network defined by the links between the $N - n$ non-sampled individuals in the population. This will generally be unknown. Using network information in a survey sampling context therefore implies that one has to deal with situations where partial network information is observed. This inevitably means that one needs to either use more complicated modelling methods or that one needs to somehow impute the missing network components.

The main focus of this article is on the potential use of network information in survey estimation. In particular, we aim to address three questions: (i) Is embedding network information useful for survey estimation based on linear models? (ii) If the answer to (i) is yes, then which network models are potentially useful? and (iii) How much network data needs to be collected in order to obtain potentially higher precision for survey estimation? In Section 2 we provide some context for these questions by defining a standard linear model that is often used for survey estimation. This linear model does not incorporate network information, so we then describe three widely used linear models that allow for the availability of network information in addition to standard covariate information.

In Section 3 we briefly discuss estimation of the population mean of a survey variable using the empirical version of the BLUP (typically referred to as the empirical best linear unbiased predictor or EBLUP) based on a linear model for this variable, and its application under the network models introduced in the previous Section. In Section 4, the Exponential Random Graph Model (ERGM) for a network is introduced and a simple imputation of missing network information is described, with the aim of using this imputed information in the network model-based estimators introduced in Chapter 3. These ideas are then brought together in Section 5 where we describe a simulation study that investigates the performances of the imputation-based EBLUPs defined by these different network models. In particular, we compare these estimators with the standard linear estimators that ignore network information. In Section 6 we use data from “wave N” (year 2004) of the BHPS to illustrate age by sex by region estimation of population means based on a model that includes age by sex effects and a contextual variable corresponding to the maleness proportion of an individual’s three best friends. Section 7 completes the article with a discussion of our findings as they relate to the three questions raised above.

2. Linear Models on Networks

In this section we describe a number of population level linear models that use network information. Throughout, we use a friendship social network structure for simplicity of exposition. In order to develop our notation, the starting point is the linear model that assumes uncorrelated errors.

2.1. The Standard Model

The classical linear model for a population of N individuals can be written in matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I}), \quad (1)$$

where $\mathbf{Y} = (Y_1, \dots, Y_N)^\top$ is a population vector of responses, $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_N)^\top$ with $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^\top$ is the model design matrix for the population with p columns defined by a set of covariates that depend on auxiliary population information, $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_N)^\top$ is the vector of population model residuals with $\epsilon_i \sim N(0, \sigma^2)$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is the vector of regression coefficients. The population mean vector and population covariance matrix of \mathbf{Y} are then $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{V} = \sigma^2 \mathbf{I}_N$. Here \mathbf{I}_N denotes the identity matrix of order N .

It is assumed that the matrix \mathbf{X} defined by the auxiliary population information does not include variables related to social networks, so (1) does not use social network information.

Survey populations are often hierarchical and can be characterised as grouped into clusters, with each cluster j accounted for by a cluster-specific random effect u_j in the model. This can be modelled by a linear mixed model to incorporate dependence of units in the same cluster. See [Chambers and Clark \(2012, Chapter 6\)](#) for more details.

2.2. The Contextual Network (CN) Model

Consider an educational modelling exercise where Academic Performance (AP) is the response variable and Socioeconomic Status (SES) of the student is the explanatory variable. A classical contextual approach might then lead one to include the average SES of the student's school as another explanatory variable. [Friedkin \(1990\)](#) adapts this idea to network data by considering models where the response for a particular subject also depends on the characteristics of other subjects that are linked to the one of interest. In our example this would correspond to modelling AP in terms of both the student's SES as well as the SES values of the student's friends. Since a student will generally have several friends, a student's AP could then be modelled in terms of his/her SES as well as the average SES of his/her friends.

In general, such a CN model can be written in matrix form as

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}\gamma + \boldsymbol{\epsilon}, \quad (2)$$

where \mathbf{Y} and \mathbf{X} have the same meaning as for Model (1), but the columns of \mathbf{U} correspond to statistics derived from the variables that are measured on the network. In particular, the i th row of \mathbf{U} contains appropriate summary characteristics of those other individuals on the network that are linked to individual i . Thus, in the preceding example, assuming that SES is the only covariate measured on the network, then \mathbf{U} is the column vector of length N whose i th value is $\overline{\text{SES}}_i$, the average SES of all friends of student i . More generally, let $\tilde{\mathbf{X}}$ denote the population matrix of covariates measured on the network. The matrix $\tilde{\mathbf{X}}$ can be a subset of \mathbf{X} but can also include other variables that are not in \mathbf{X} .

Then one way of defining \mathbf{U} is via the identity

$$\mathbf{U} = \mathbf{W}\tilde{\mathbf{X}}, \quad (3)$$

where $\mathbf{W} = \mathbf{Z}/\mathbf{Z}\mathbf{1}_{N,N}$ is a row-normalised version of \mathbf{Z} , that is the rows of \mathbf{W} sum to one. In general, $\mathbf{U} = g(\mathbf{Z}, \tilde{\mathbf{X}})$ is a function of the network \mathbf{Z} and $\tilde{\mathbf{X}}$. A contextual variable for person i often includes the value for this person, for example a household contextual effect is computed over all household members including person i . However, the contextual value for person i defined by (3) excludes person i , because $Z_{ii} = 0$ by definition.

2.3. The Autocorrelation (AR) Model

The matrix $\tilde{\mathbf{X}}$ introduced in the preceding description of the CN model can be any set of measurements on the individuals in the network. In particular, it can be \mathbf{Y} . This leads to another class of models, called Autocorrelation (AR) models, and also known as network effects models, that incorporate network information into a linear structure. See, for

example, [Doreian et al. \(1984\)](#), [Duke \(1993\)](#), [Marsden and Friedkin \(1993\)](#), and [Leenders \(2002\)](#), and in the context of spatial models, [Ord \(1975\)](#). Under an AR model,

$$\mathbf{Y} = \theta \bar{\mathbf{Y}} + \mathbf{X}\beta + \boldsymbol{\epsilon}, \quad (4)$$

where $\bar{\mathbf{Y}} = (\bar{Y}_1, \dots, \bar{Y}_N)^\top$ and \bar{Y}_i is the average response of the individuals in the network that are linked to individual i , so $\bar{\mathbf{Y}} = \mathbf{W}\mathbf{Y}$, with \mathbf{W} defined in the previous subsection. The conditional (on \mathbf{X}) mean and variance of \mathbf{Y} under (4) are $\mu = \mathbf{D}^{-1}\mathbf{X}\beta$ and $\mathbf{V} = \sigma^2(\mathbf{D}^\top\mathbf{D})^{-1}$, where $\mathbf{D} = \mathbf{I}_N - \theta\mathbf{W}$. Note that \mathbf{W} can be defined in a variety of ways, see [Leenders \(2002\)](#), though typically it is defined as the row-normalised version of \mathbf{Z} , that is, $\sum_{j=1}^N W_{ij} = 1$. The parameter θ is restricted $\theta \notin \{\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_l}\}$ as a necessary condition for \mathbf{V} to exist, where λ_i are the eigenvalues of the row-normalised \mathbf{W} . Often θ is restricted to $(-1, 1)$.

In the context of the academic performance example introduced in the previous subsection we see that (4) implies that a student's AP score now depends on his/her SES value as well as the average AP scores of his/her friends.

2.4. The Network Disturbance (ND) Model

Models of this type have been considered by [Ord \(1975\)](#) and [Leenders \(2002\)](#) among others, and correspond to imposing an AR structure on the error term in the standard linear model (1). They are referred to as Network Disturbance (ND) models and are specified by

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} = \theta\bar{\boldsymbol{\epsilon}} + \mathbf{v}, \mathbf{v} \sim N(0, \sigma^2\mathbf{I}_N). \quad (5)$$

Here $\bar{\boldsymbol{\epsilon}} = (\bar{\epsilon}_1, \dots, \bar{\epsilon}_N)$ where $\bar{\epsilon}_i$ is the average error of those individuals in the network linked to individual i . Returning to the academic performance example introduced in Subsection 2.2, the model can be interpreted as implying that if a student's friends have a below/above average AP value (as predicted by their SES values), then the student is more likely to have an AP value that is also below/above average.

Note that the Model (5) can be rewritten as

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim N(0, \sigma^2(\mathbf{D}^\top\mathbf{D})^{-1}), \quad (6)$$

where \mathbf{D} was defined in Subsection 2.3, with the same restrictions in place on θ as for the AR model. The parameter θ is an indicator of the strength of the between individual correlations generated by the network. For $\theta = 0$, the correlation between the Y values of any two individuals in the network is zero after one adjusts for their respective values of \mathbf{X} . Under (6), the conditional (on \mathbf{X}) mean and variance of \mathbf{Y} are $\mu = \mathbf{X}\beta$ and $\mathbf{V} = \sigma^2(\mathbf{D}^\top\mathbf{D})^{-1}$ respectively.

It is worth pointing out that under the ND model, $\mu = \mathbf{X}\beta$ is unaffected by the social network, whereas under the AR model (4), $\mu = \mathbf{D}^{-1}\mathbf{X}\beta$ depends on the network through \mathbf{D} . That is, under the ND model, the expected value of Y for an individual only depends on the values of that individual's covariates. Unbiased prediction of Y can therefore ignore the network. Of course efficient prediction depends on the second order moments of (6), and so requires network information – as does prediction variance and mean squared error estimation. This is analogous to estimation under a multi-level model, where one can

ignore the multi-level structure of the data if unbiased estimation is the aim, but one needs to take this structure into account for efficient inference.

3. Prediction of Population Totals Using Network Models

The models discussed in the previous section are predictive models, that is, when second order moments are known, they can be used to compute efficient predictions of unknown values of the response variable. We now describe how these models can be fitted, and how predicted values derived from them can be used to estimate the population total $T_Y = \sum_{i \in U} Y_i$ given the sample values $\{Y_i : i \in s\}$, the population matrix of model covariates \mathbf{X} and either part of or all of the network matrix \mathbf{Z} . Throughout we assume that inclusion in sample does not depend on \mathbf{Z} and that there is non-informative sampling given \mathbf{X} , see Section 1.4 in [Chambers and Clark \(2012\)](#). Consequently, all unknown parameter values for the standard model (1) can be estimated from the sample data and predicted values of Y for the non-sampled population individuals can be computed. We start by summarising known results from finite population estimation theory.

3.1. The Empirical Best Linear Unbiased Predictor

Let $E(\mathbf{Y}) = \boldsymbol{\mu} = \mathbf{H}\boldsymbol{\lambda}$, where \mathbf{H} is a known matrix with N rows and q columns and $\boldsymbol{\lambda}$ is an unknown parameter vector of length q . Also, suppose that $\text{Var}(\mathbf{Y}) = \mathbf{V}$ is a positive definite matrix of order N whose value is known up to a constant of proportionality. Examples of \mathbf{H} and \mathbf{V} are given in the following subsection. The best linear unbiased predictor or BLUP of the population total $T_Y = \sum_{i \in U} Y_i$ is then an efficient estimator of this quantity, see [Royall \(1976\)](#). In order to specify the BLUP, let s and r denote the n sampled and $N - n$ non-sampled population individuals respectively, and put $\mathbf{H} = (\mathbf{H}_s^\top, \mathbf{H}_r^\top)^\top$ and $\mathbf{Y} = (\mathbf{Y}_s^\top, \mathbf{Y}_r^\top)^\top$. The matrix \mathbf{V} can then be partitioned conformably as

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{ss} & \mathbf{V}_{sr} \\ \mathbf{V}_{rs} & \mathbf{V}_{rr} \end{pmatrix}.$$

A standard expression for the BLUP is its so-called predictive form

$$\hat{T}_Y^{BLUP} = \sum_{i \in s} Y_i + \sum_{i \in r} \mathbf{H}_i \hat{\boldsymbol{\lambda}} + \sum_{i \in s} \tau_i (Y_i - \mathbf{H}_i \hat{\boldsymbol{\lambda}}), \quad (7)$$

where \mathbf{H}_i is the i th row of \mathbf{H} , $\hat{\boldsymbol{\lambda}} = (\mathbf{H}_s^\top \mathbf{V}_{ss}^{-1} \mathbf{H}_s)^\top \mathbf{H}_s^\top \mathbf{V}_{ss}^{-1} \mathbf{Y}_s$ is the best linear unbiased estimator (BLUE) of $\boldsymbol{\lambda}$, and τ_i is the i th element of the vector $\mathbf{V}_{ss}^{-1} \mathbf{V}_{sr} \mathbf{1}_{N-n}$, with $\mathbf{1}_{N-n}$ denoting a vector of ones of size $N - n$.

Note that the BLUP can also be expressed as a weighted sum $\hat{T}_Y^{BLUP} = \sum_{i \in s} w_i Y_i = \mathbf{w}_s^\top \mathbf{Y}_s$ of the sample values of Y , where

$$\mathbf{w}_s = \mathbf{1}_n + \mathbf{M}^\top (\mathbf{H}^\top \mathbf{1}_N - \mathbf{H}_s^\top \mathbf{1}_n) + (\mathbf{I}_n - \mathbf{M}^\top \mathbf{H}_s^\top) \mathbf{V}_{ss}^{-1} \mathbf{V}_{sr} \mathbf{1}_{N-n} \quad (8)$$

is the vector of BLUP weights. Here $\mathbf{1}_n$ is a vector of ones of size n and matrix \mathbf{M} is defined as $\mathbf{M} = (\mathbf{H}_s^\top \mathbf{V}_{ss}^{-1} \mathbf{H}_s)^\top \mathbf{H}_s^\top \mathbf{V}_{ss}^{-1}$.

A key assumption of the BLUP is that the variance matrix \mathbf{V} is known up to a constant of proportionality. This is often unrealistic, since \mathbf{V} can depend on unknown parameters,

which must then be estimated. Methods for doing this are described in the next section. Substituting these estimates into \mathbf{V} defines its plug-in estimator $\hat{\mathbf{V}}$, which can be used in (8) instead of \mathbf{V} . The resulting estimator of the population total is called the empirical BLUP or EBLUP.

3.2. Calculating the EBLUP under Network Models

In order to use the EBLUP with the different network models defined in the previous section, we need to specify \mathbf{H} and \mathbf{V} as well as estimators of the unknown parameters that underpin these matrices. These are defined as follows:

Standard Model: Here $\mathbf{H} = \mathbf{X}$ and $\mathbf{V} = \sigma^2 \mathbf{I}_N$. The residual mean squared error defines an unbiased estimator of σ^2 .

CN Model: For this model $\mathbf{H} = [\mathbf{X}, \mathbf{U}]$ and $\mathbf{V} = \sigma^2 \mathbf{I}_N$. We can unbiasedly estimate σ^2 using the residual mean squared error.

AR Model: In this case $\mathbf{H} = \mathbf{D}^{-1} \mathbf{X}$ with $\mathbf{D} = \mathbf{I}_N - \theta \mathbf{W}$ and $\mathbf{V} = \sigma^2 (\mathbf{D}^\top \mathbf{D})^{-1}$. Estimates of σ^2 and θ can be obtained by Maximum Likelihood (ML). Restricted ML (REML) is often used to obtain unbiased variance estimates but it cannot be applied here, because both the mean and variance depend on the parameter θ . The EBLUP uses the plug-in estimates of \mathbf{H} and \mathbf{V} defined by the ML estimates of σ^2 and θ .

ND Model: Here $\mathbf{H} = \mathbf{X}$ and $\mathbf{V} = \sigma^2 (\mathbf{D}^\top \mathbf{D})^{-1}$. ML estimation of σ^2 and θ can be carried out, and the resulting plug-in estimate of \mathbf{V} is used to calculate the EBLUP.

ML estimation of σ^2 and θ for the AR and ND models is not straightforward. Both models are not reproducible, that is, they do not share the property that the model for a subset of units of the population has the same form as the model for the whole population. To see this, note that the variance of the population response vector \mathbf{Y} under both models is $\sigma^2 (\mathbf{D}^\top \mathbf{D})^{-1}$ so that the variance for the sample response vector \mathbf{Y}_s is $\sigma^2 [(\mathbf{D}^\top \mathbf{D})^{-1}]_{ss}$. In general, this will not equal $\sigma^2 (\mathbf{D}_{ss}^\top \mathbf{D}_{ss})^{-1}$, which is the assumed variance if the model is fitted via ML at the sample level. This misspecification can lead to biased estimates of the model parameters. A modified approach that yields unbiased estimates of the fixed effects in the model is described in [Suesse \(2012a\)](#). However this is computationally intensive. An alternative approach replaces \mathbf{D}^{-1} by a fourth order Taylor series approximation. This speeds up computation considerably since it effectively replaces matrices of dimension $N \times N$ by matrices of dimension $n \times n$. See [Suesse \(2012a\)](#) where it is shown that ML estimates based on this approximation are essentially identical to those obtained using the modified ML method. An alternative exact ML method that is computationally more demanding was considered by [Suesse and Zammit Mangion \(2017\)](#).

3.3. Variance Estimation for the EBLUP

The prediction variance of the BLUP is

$$\text{Var}(\hat{T}^{BLUP} - T) = \tilde{\mathbf{w}}^\top \mathbf{V} \tilde{\mathbf{w}} \quad (9)$$

with $\tilde{\mathbf{w}}^\top = (\mathbf{w}_s^\top - \mathbf{1}_n^\top, -\mathbf{1}_{N-n}^\top)^\top$. This formula assumes that the vector of survey weights \mathbf{w}_s is fixed. We can use the same formula for the EBLUP, although from (8) it is clear that

the EBLUP weights are not fixed in general because the plug-in estimates of \mathbf{H} and \mathbf{V} used to calculate them will depend on estimated parameters. However, the increase in the prediction variance due to ML estimation of these parameters will be small for large sample sizes, and can be ignored, see [Chambers et al. \(2011\)](#).

Using (9) to estimate the prediction variance of the EBLUP depends on correct specification of the second order moments of Y . For the standard model and the CN model, we can avoid this by using an alternative prediction variance estimator that does not rely on specification of these second order moments, see Section 9.2 of [Chambers and Clark \(2012\)](#). This estimator is given by

$$\widehat{\text{Var}}(\hat{t}_{BLUP} - t) = \sum_{i \in s} (w_{is} - 1)^2 (Y_i - \hat{\mu}_i)^2 + (N - n) \hat{\sigma}^2 \quad (10)$$

where $\hat{\mu}_i$ is the estimated mean for $i \in s$, that is $\hat{\mu}_i = \mathbf{X}_i \hat{\beta}$ for the standard model and $\hat{\mu}_i = \mathbf{X}_i \hat{\beta} + \mathbf{U}_i \hat{\gamma}$ for the CN model, with $\hat{\sigma}^2$ corresponding to the usual unbiased estimator of σ^2 under each model.

For the AR and ND models we use Equation (9) with a plug-in estimator $\hat{\mathbf{V}}$. In this context, we note that ML estimates of variance parameters are known to be biased, which could therefore lead to a bias in $\hat{\mathbf{V}}$ and in the resulting plug-in estimator defined by (9). The standard approach to dealing with this issue is to apply REML instead of ML. Unfortunately, the AR model does not allow the application of REML, and furthermore REML is computationally more complex when fitting these population models. Consequently a bias-corrected version of ML was applied, based on the approach set out in [Goldstein \(1989\)](#), which adjusts Iterative Generalized Least Squares (IGLS) to obtain estimates that are equivalent to REML. The details of this are outlined in the Appendix of [Suesse and Chambers \(2014\)](#).

4. Modelling of Networks

Our EBLUP development in the previous section assumed that the matrix \mathbf{Z} defining the network is known. This is rather unlikely to be the case. It is far more likely that we will know either just that part of the network defined by the sampled individuals (i.e., \mathbf{Z}_{ss}) or that part of the network defined by the sampled individuals and their corresponding network links (i.e., \mathbf{Z}_{ss} and \mathbf{Z}_{sr}).

An implementation of a ‘network-based’ EBLUP in this situation must therefore take account of this incomplete network data. In this section we describe simple model-based imputation methods that can be used to approximate the impact of the unknown full network (i.e., \mathbf{Z}) on this EBLUP. In turn, this requires that we have a way of modelling \mathbf{Z} , given that we see only a part of this matrix. We start with a brief overview of models for networks.

4.1. Exponential Random Graph Models

A popular class of models that is able to describe dependencies in a network \mathbf{Z} is the class of (curved) exponential random graph models (ERGMs), these are discussed in [Wasserman and Faust \(1994\)](#) and [Carrington et al. \(2005\)](#). Under an ERGM, the

distribution of \mathbf{Z} is characterised by

$$\Pr(\mathbf{Z} = \mathbf{z}) = \exp(\eta(\boldsymbol{\zeta})' \mathbf{G}(\mathbf{z}) - \kappa(\boldsymbol{\zeta})), \quad (11)$$

where $\boldsymbol{\zeta}$ is the vector of model parameters, $\eta(\boldsymbol{\zeta})$ is a mapping from the p -dimensional to the q -dimensional space with $p \leq q$, and $\kappa(\boldsymbol{\zeta})$ is the normalising constant. Here $\mathbf{G}(\mathbf{z})$ is a vector of q ‘network statistics’ which, together with $\boldsymbol{\zeta}$, completely characterises the distribution of \mathbf{Z} . Simple examples of network statistics are the number of ‘edges’ in the network (i.e., the number of observed links, usually expressed as a fraction of the total number $N(N-1)$ of possible links) and the number of triangles (a triangle is said to exist between individuals i, j and k , if $Z_{ij} = Z_{jk} = Z_{ik} = 1$). A more complicated, but widely used network statistic is GWESP, or the geometrically weighted edgewise shared partner statistic. Roughly speaking, this corresponds to a weighted sum, over possible values of m , of counts of the number of links ‘connecting’ any two individuals in the network who are themselves linked to exactly m other individuals. Like interaction terms in regression, such statistics allow one to model networks whose ‘connectivity’ structure is extremely complicated.

Fitting an ERGM via ML is usually not possible, mainly because direct calculation of the normalising constant $\kappa(\boldsymbol{\zeta})$ is infeasible. One way of circumventing this problem is to sample from the network distribution (11) using a Markov-Chain-Monte-Carlo (MCMC) algorithm in order to obtain a stochastic approximation to the maximum likelihood estimate of $\boldsymbol{\zeta}$. Such estimates are called MCMC ML estimates (Hunter and Handcock 2006). Describing the network distribution via simple network statistics, such as the number of triangles then becomes problematic, because such specifications often lead to degenerate MCMC samples. Some authors (Snijders 2002; Snijders et al. 2006) have therefore proposed the use of more complex network statistics, such as the family of GWESP statistics, for which degeneracy seems less of a problem. For more details of network modelling, see Strauss and Ikeda (1990), Hunter and Handcock (2006), Hunter (2007), Hunter et al. (2008a), and Butts (2008).

4.2. Types of Partially Observed Networks

In the first case, denoted by SS in what follows, only \mathbf{Z}_{ss} is observed and so \mathbf{Z}_{sr} , \mathbf{Z}_{rs} and \mathbf{Z}_{rr} are missing. In the second case, denoted by SS+SR in what follows, \mathbf{Z}_{ss} and \mathbf{Z}_{sr} are observed but \mathbf{Z}_{rs} and \mathbf{Z}_{rr} are missing. This might appear strange, because for an undirected network $\mathbf{Z}_{sr} = \mathbf{Z}_{rs}^T$. This situation is motivated by the BHPS data set for which contextual information \mathbf{U}_s is available for the sample but not for the non-sample that is, \mathbf{U}_r is unavailable. This corresponds to knowing \mathbf{Z}_{ss} and \mathbf{Z}_{sr} , but not knowing \mathbf{Z}_{rs} and \mathbf{Z}_{rr} , because \mathbf{U}_s is function of \mathbf{Z}_{ss} , \mathbf{Z}_{sr} and $\tilde{\mathbf{X}}$ and \mathbf{U}_r is function of \mathbf{Z}_{rs} , \mathbf{Z}_{rr} and $\tilde{\mathbf{X}}$. See Appendix C for more details on the relationship between \mathbf{U} and \mathbf{Z} and $\tilde{\mathbf{X}}$ and simple estimators.

The third case, denoted SS+SR+RS is where \mathbf{Z}_{ss} , \mathbf{Z}_{sr} and \mathbf{Z}_{rs} are observed, with \mathbf{Z}_{rr} missing. The second and third cases are more realistic from the viewpoint of having usable network information, since here we at least have complete network information for all sampled individuals. In this context, we note that the second case provides a scenario which is related to the situation of the BHPS, where the network is not directly available but where contextual variables defined by the sample, that is \mathbf{U}_s , are known.

4.3. Imputation of Partly Observed Networks

An estimate $\hat{\mathbf{Z}}$ of the full network is necessary for calculation of the EBLUP under the network models considered in this article. However, in practice only part of the network will be observed, say \mathbf{Z}^{obs} , and another part will be missing, say \mathbf{Z}^{mis} . For example, for the scenario SS the observed network \mathbf{Z}^{obs} is \mathbf{Z}_{ss} and the missing network \mathbf{Z}^{mis} is $\mathbf{Z}_{sr} \cup \mathbf{Z}_{rs} \cup \mathbf{Z}_{rr}$. Note that here we focus on single-value imputation of \mathbf{Z}^{mis} . Our approach can be extended to multiple imputation.

We apply a simple, robust and computationally feasible approach for imputation. Standard ERGMs, for example the ERGM with EDGES and GWESP, imply a fixed marginal probability of the form $P(Z_{ij} = 1) = p$. However, this probability cannot in general be analytically determined from the model parameters, and must instead be estimated separately, for example either via simulation from the underlying ERGM using plug-in estimates of the ERGM parameters, or more simply by the moment estimator $\hat{p} = \frac{1}{|\mathbf{Z}^{obs}|} \sum_{i,j,i \neq j} Z_{ij}^{obs}$. In what follows, we apply the latter approach as this is a standard estimator for a proportion, replacing all Z_{ij}^{mis} by \hat{p} . This approach is simple and clearly can be improved upon as it uses the unconditional expectation $E(\mathbf{Z}^{mis})$ as an estimator of the conditional expectation $E(\mathbf{Z}^{mis} | \mathbf{Z}^{obs} = \mathbf{z}^{obs})$. However, estimating the latter expectation requires first fitting an ERGM to the incomplete network data obtained from the sample, calculating model parameter estimates $\hat{\theta}$, and then applying model-based imputation methods, for example multiple imputation. This approach was infeasible for the simulation study reported next, as fitting an ERGM with a large portion of the network missing took more than four hours on a single core of type Intel Xeon E5-2620 v2 – 2.10GHz for a single data set with the latest available version of *ergm* (Hunter et al. 2008b). In comparison, when no missing data are present, this fitting process took only a few seconds. Note that such considerations may not be relevant in an application where imputation of a single network is required. In this case, adopting a more sophisticated imputation method may be advisable, for example by applying the approach of Pattison et al. (2013). However even if its estimation could be improved, sampling a large number of networks is needed to obtain an estimate of $E(\mathbf{Z}^{mis} | \mathbf{Z}^{obs} = \mathbf{z}^{obs})$ or even only $E(\mathbf{Z})$. But when N is large, for example $N = 100,000$, this also may not be feasible since sampling one network took 31 hours, meaning that for large N using a more sophisticated method is impractical even if estimation is not an issue. That is, in practice, for large N the simple method mentioned above appears to be the only feasible method.

5. Mean Squared Error Estimators under the Standard and Contextual Linear Model

The AR and the ND models are difficult to fit for large N when the networks are imputed, because then the contiguity matrices are not sparse any more, making it near impossible to calculate the log-likelihood. From a practical perspective, only the contextual model is feasible to fit for large N when the network is imputed. We consider now mean squared error estimation under the standard linear model and under the contextual model when the network is only partially available (situation SS+SR) but when $\tilde{\mathbf{X}}$ is available for the whole population. For this situation, the population total $\mathbf{T}_U = \mathbf{1}_N^\top \mathbf{U}$ with $\mathbf{U} = (\mathbf{U}_s^\top, \mathbf{U}_r^\top)^\top$ referring to the total of the contextual variable must be estimated, that is $\hat{\mathbf{U}} = g(\hat{\mathbf{Z}}, \tilde{\mathbf{X}})$.

In [Appendix C](#), we show under SS+SR (\mathbf{U}_s known) that $\hat{\mathbf{U}}_r$ has a simple form that does not require the network to be observed, but only requires an estimate of the constant $P(Z_{ij} = 1) = p$, $\hat{\mathbf{X}}$ and the parameters N and n . Hence even for the BHPS for which p is known, this estimator $\hat{\mathbf{U}}$ can be calculated relatively easily and the estimated total $\hat{\mathbf{T}}_{\mathbf{U}}$ obtained. [Appendix C](#) shows the estimator $\hat{\mathbf{T}}_{\mathbf{U}}$ along with a (co)variance estimator denoted by $\hat{\mathbf{V}}_{\hat{\mathbf{T}}_{\mathbf{U}}}$.

Let the mean squared error (MSE) of the EBLUP \hat{T} be denoted by $MSE[\hat{T}]$ and let the EBLUP for the linear model without contextual information be denoted by \hat{T}_I and the corresponding EBLUP defined by the contextual model be denoted \hat{T}_C .

Put $\Delta = MSE[\hat{T}_C] - MSE[\hat{T}_I]$. See [Appendix B](#) for an expression for Δ depending on $\hat{\mathbf{T}}_{\mathbf{U}}$ and $\hat{\mathbf{V}}_{\hat{\mathbf{T}}_{\mathbf{U}}}$ along with an estimator $\hat{\Delta}$. In practice we propose to use these estimators to choose between the models without and with contextual information. When $\hat{\Delta} < 0$ we propose to use the contextual model and when $\hat{\Delta} \geq 0$, we propose to use the standard model. [Appendix B](#) also shows expressions for Δ and an estimator for $\hat{\Delta}$ for the unrealistic situation that \mathbf{U} is fully known, using the estimator $\hat{\Delta}$ proposed by [Clark and Chambers \(2008\)](#).

6. Simulation Study

6.1. Study Design

This section contains the results from a simulation study whose aim was to investigate the effect of using networks as an additional source of information when estimating the population total T_Y of a survey variable Y . A networked population of size $N = 1,000$ was independently simulated 2,000 times, balancing computation time against the number of different scenarios that were explored in the study, and independent simple random samples of size $n = 100$ and $n = 200$ were independently selected without replacement from each simulated population. This study comprises all four models under investigation. We also consider larger N , but then estimation of the AR and ND models becomes infeasible. To illustrate the effect of large N , only the contextual model is compared with the standard linear model.

6.1.1. Network Generation

We mainly consider undirected networks in our simulations. The literature on network analysis suggests that such networks are often well characterised by an ERGM defined in terms of an EDGES (number of edges) statistic and a GWESP statistic ([Hunter et al. 2008b](#)). Consequently, \mathbf{Z} was generated as a random draw from an ERGM with an EDGES statistic equal to θ on the logit scale and a weight parameter of 1.0 for the GWESP statistic. In what follows we use $ERGM(m)$ to denote such an ERGM, where m is the network density, that is the average number of links per individual. The values of θ were then chosen in order to generate a network with a density of about $m = 3, 15, 50$ network links respectively for each individual, that is $m \approx P(Z_{ij} = 1) \times N$ with $P(Z_{ij} = 1) \approx \text{expit}(\theta)$. Note that with this specification the number of network links for an individual is random, with only the approximate population average number of links fixed.

In this study we consider the three types of partially observed networks mentioned before, namely SS, SS+SR and SS+SR+RS. Finally, we also considered the situation

where no network data are used (the standard model) and also the case where the population network is fully known.

6.1.2. Parameter Specification for Linear Network Models

We generated data under the CN, AR and ND linear network models, see Section 2. Population data were simulated assuming $\sigma^2 = 3^2 = 9$, $\beta_0 = 1$ and $\beta_1 = 2$. Furthermore, the auxiliary variable X was defined so that it took values randomly in the set $\{1, \dots, 9\}$. This model has medium to high predictive power, since the Standard model implies a theoretical value of approximately $R^2 = 0.75$.

CN Model:

$$Y_i = \beta_0 + X_i\beta_1 + U_i\gamma + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$

Here $\gamma = 2$ and the contextual variable U_i is defined as the average value of X for all other individuals in the network that individual i has links with, that is $U = \mathbf{W}\mathbf{X}$, where \mathbf{W} is the row-normalised version of \mathbf{Z} and \mathbf{X} denotes the vector of population values of X .

AR Model:

$$\mathbf{Y} = \theta\mathbf{W}\mathbf{Y} + \beta_0 + \mathbf{X}\beta_1 + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2\mathbf{I}_N)$$

with $\theta = 0.5$.

ND Model:

$$\mathbf{Y} = \beta_0 + \mathbf{X}\beta_1 + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} = \theta\mathbf{W}\boldsymbol{\epsilon} + \mathbf{v}, \quad \mathbf{v} \sim N(0, \sigma^2\mathbf{I}_N)$$

with $\theta = 0.5$.

6.2. Simulation Results

Results for the $n = 100$ case are presented. Table 1 show the Monte Carlo relative mean squared errors of the estimates of T when the network is generated under an ERGM where the total number of friends is random, with expectations three and ten respectively. Corresponding simulation results for $X \sim N(0, 25)$ can be found in Suesse and Chambers (2014). Results for the $n = 200$ case are similar. Note that we do not show Monte Carlo bias, since these values were effectively zero for all methods. The results displayed in each table include the two cases where the network is ignored (the ‘standard’ model) and when the population network matrix \mathbf{Z} is fully known (‘True Model and \mathbf{Z}_U known’). For partially observed network data we show results for the SS case (only \mathbf{Z}_{ss} known), the SS+SR case (\mathbf{Z}_{ss} and \mathbf{Z}_{sr} known, simple imputation) and the SS+SR+RS case (\mathbf{Z}_{ss} , \mathbf{Z}_{sr} and \mathbf{Z}_{rs} known, simple imputation). All results are shown relative to those for the BLUP, which uses complete network information as well as knowledge of θ . Although the level of knowledge required to compute the BLUP is unrealistic in practice, its performance provides us with a benchmark against which to gauge the relative benefit of putting more effort into collecting more network information and in carrying out more intensive network modelling for imputation of the unknown parts of the network. Furthermore, comparisons with the ‘Standard’ case allow us to assess how much efficiency is lost by ignoring network information.

Table 1. $n = 100$: Undirected ERGM(3) and ERGM(10) networks with X drawn randomly from $\{1, \dots, 9\}$. Ratio of MSE(EBLUP) to MSE(BLUP).

	Population data generated under model					
	ERGM(3)			ERGM(10)		
	CN	AR	ND	CN	AR	ND
Actual MSE	86, 474	101, 148	101, 082	86, 537	89, 483	89, 705
<i>Relative EBLUP to actual MSE based on</i>						
True model and \mathbf{Z}_U known	1.00	1.00	1.01	1.00	0.99	1.02
CN						
SS+SR+RS	1.11	1.09	1.04	1.03	1.02	1.00
SS+SR	1.11	1.09	1.04	1.03	1.02	1.00
SS	2.33	1.37	1.03	1.32	1.07	1.01
AR						
SS+SR+RS	1.34	1.05	1.05	1.09	1.01	1.00
SS+SR	1.20	1.06	1.05	1.07	1.01	1.00
SS	2.42	1.39	1.03	1.33	1.06	1.01
ND						
SS+SR+RS	2.21	1.31	1.03	1.36	1.06	1.00
SS+SR	2.36	1.35	1.01	1.34	1.07	1.01
SS	2.42	1.40	1.02	1.34	1.07	1.01
Standard	2.40	1.40	1.03	1.33	1.07	1.00

It is clear from the results shown in Table 1 that ignoring the network (i.e., using the ‘Standard’ model for estimation) can lead to a large loss in efficiency if in fact either the AR or the CN models are true. Interestingly, our results also seem to indicate that adopting the CN model when in fact the AR model is true seems as good as using the correctly specified AR model when the number of friends is not small. Note that when the ND model is true, ignoring the network information in the data only leads to a marginal loss in efficiency. In fact, the EBLUPs based on the different network models are all almost fully efficient in this case, irrespective of whether the assumed network model is true.

When \mathbf{Z} is known, but not θ , we see a loss of efficiency under the AR model, mainly because the pseudo-design matrix $\mathbf{D}^{-1}(\theta)\mathbf{X}$ for this model depends on the estimated value of θ . As the number of friends increases, this loss of efficiency associated with having to estimate θ from the sample data decreases in importance. This problem is much less of an issue for the ND model because in this case the design matrix does not depend on θ . Obviously, there is no impact under the CN model.

In order to see why the CN model yields similar results as the AR model when in fact the AR model holds, we note that the mean of the AR model is $\mu = \mathbf{D}(\theta)^{-1}\mathbf{X}\beta$. If we approximate $\mathbf{D}(\theta)^{-1}$ by a first order Taylor series around zero, that is

$\mathbf{D}(\theta)^{-1} = (\mathbf{I}_N - \theta \mathbf{W})^{-1} \approx \mathbf{I}_N + \theta \mathbf{W}$, then

$$\mu \approx \mathbf{X}\beta + \theta \mathbf{W}\mathbf{X}\beta = \mathbf{X}\beta + \mathbf{U}\gamma$$

with $\gamma = \theta\beta$ and $\mathbf{U} = \mathbf{W}\mathbf{X}$. That is, the implied mean structure under the AR model is approximately the same as that under a CN model.

When \mathbf{Z}_{ss} and \mathbf{Z}_{sr} are observed, the EBLUP based on the CN model appears to perform well generally. This is because the EBLUP under this model does not depend on either \mathbf{Z}_{rs} or \mathbf{Z}_{rr} and hence is unaffected by imputation of this part of the network. This is in contrast to the performance of this EBLUP when only \mathbf{Z}_{ss} is observed. Here we see that the need to impute \mathbf{Z}_{sr} leads to a significant loss of efficiency. Since estimation of θ in the pseudo-design matrix $\mathbf{D}(\theta)^{-1}\mathbf{X}$ under the AR model has a larger negative effect than the approximation of the AR model by the CN model, we conclude that the EBLUP based on the CN model seems a generally more robust method for estimating the population total than the EBLUP based on the AR model.

It is interesting to also observe that the EBLUP based on the AR model and SS+SR network data performs generally better than the same EBLUP with access to more extensive SS+SR+RS network data when the CN model holds, reflecting the interaction of model misspecification and imputation biases. However, this effect is reversed when a ND-based EBLUP is used and a CN model underpins the network.

When we focus on where the expected number of friends per subject is small, here equal to three (see Table 1), we note that there is only a small gain associated with using imputation method SS compared to ignoring the network information and basing estimation on the ‘Standard’ model. In this situation network imputation based on SS+SR or SS+SR+RS provides the largest gains relative to ignoring the network when the contextual CN or AR model is fitted.

To investigate the effect for larger N only the contextual model and the standard model are compared, as the AR and ND models are infeasible to fit. Table 2 shows the MSE of the BLUP and various EBLUPs including the adaptive strategy that chooses either model depending on the sign of $\hat{\Delta}$ for an undirected ERGM network with approximately ten friends per person. Table 3 shows results for a directed ERGM network with approximately three friends per person. The tables also show the empirical mean of $\hat{\Delta}$ denoted by $E(\hat{\Delta})$ and Δ . Note that the estimator $\hat{\Delta}$ has only a small bias.

The results also show that the adaptive strategy is effective and that this strategy also works for large N and small sampling fractions. Surprisingly the efficiency gains even increase as the sampling ratio increases. This might be due to the fact that the simple estimate \hat{p} , the proportion of links in the observed network, is more precise with increasing N because the number of dyads in \mathbf{Z}^{obs} is $n(N-1)$ and increases with N .

Average lengths and associated coverages for nominal 95% Gaussian confidence intervals generated by the estimates of the mean squared errors of the different estimators are set out in Table 6, see the Appendix. Results for $X \sim N(0, 25)$ can be found in Suesse and Chambers (2014). Monte Carlo coverages in all cases are close to the nominal level. However, the average confidence interval length in the SS+SR/SS+SR+RS case is considerably shorter than that for the SS case when estimation is carried out under the AR

Table 2. Undirected ERGM(10) network with X drawn randomly from $\{1, \dots, 9\}$. Ratio of $MSE(EBLUP)$ to $MSE(BLUP)$ and estimated and true MSE -difference Δ .

N n	Settings of N and n					
	1,000 100	10,000 100	10,000 200	100,000 100	100,000 200	100,000 1,000
BLUP -actual MSE multiplied by $100/N^2$	8.653	9.095	8.940	9.310	9.160	9.064
<i>Relative MSE of EBLUP based on</i>						
\mathbf{Z}_U known	1.000	1.000	1.000	1.000	1.000	1.000
SS+SR+RS	1.092	1.004	1.006	1.000	1.001	1.001
SS+SR	1.075	1.004	1.006	1.000	1.001	1.001
SS	1.334	22.69	1.334	1658	734.1	1.321
Standard	1.337	1.318	1.340	1.335	1.317	1.323
Adaptive strategy	1.078	1.004	1.006	1.000	1.001	1.001
$-\Delta$ relative to BLUP	0.249	0.314	0.328	0.334	0.316	0.322
$E(-\hat{\Delta})$ relative to BLUP	0.253	0.316	0.322	0.318	0.317	0.318

and CN models. This provides further support for the conclusion reached above, that basing an EBLUP on a CN model seems a generally robust approach to using network data when estimating a population total, even though one must keep in mind that the simple linearization-based prediction variance estimator (10) used with the CN model slightly

Table 3. Directed ERGM(3) network with X drawn randomly from $\{1, \dots, 9\}$. Ratio of $MSE(EBLUP)$ to $MSE(BLUP)$ and estimated and true MSE -difference Δ .

N n	Settings of N and n					
	1,000 100	10,000 100	10,000 200	100,000 100	100,000 200	100,000 1,000
BLUP -actual MSE multiplied by $100/N^2$	8.401	9.062	8.931	9.293	9.156	9.051
<i>Relative MSE of EBLUP based on</i>						
\mathbf{Z}_U known	1.000	1.000	1.000	1.000	1.000	1.000
SS+SR	1.122	1.013	1.024	1.001	1.002	1.011
SS	2.202	529.1	21.59	2349	2936	2.300
Standard	2.201	2.259	2.274	2.267	2.262	2.310
Adaptive strategy	1.122	1.013	1.024	1.001	1.002	1.011
$-\Delta$ relative to BLUP	1.079	1.246	1.250	1.266	1.261	1.299
$E(-\hat{\Delta})$ relative to BLUP	1.095	1.288	1.265	1.252	1.241	1.262

underestimates random variation due to its assumption of fixed sample weights, resulting in too narrow confidence intervals with slight undercoverage.

7. Illustrative Example

The British Household Panel Study (BHPS) is an annual multi-purpose household panel survey in the United Kingdom that focuses on gaining insight into the social and economic change at the individual and household level in Britain and the UK, see <https://www.iser.essex.ac.uk/bhps/> for more details.

We focus on an individual’s annual income (in pounds sterling) as the variable of interest. Our aim is to investigate how the use of network information available in BHPS impacts on average income estimates for the cross-classification age by gender by region, using six categories for age 15–18 (1), 19–21 (2), 22–30 (3), 31–50 (4), 51–64 (5), 65+ (6) (in years), two for gender (1: male, 0: female) and five regions defined as: (1) ‘E/North’ consisting of East Midlands, West Midlands Conurbation, Rest of West Midlands, Greater Manchester, Merseyside, Rest of North West, South Yorkshire, West Yorkshire, Rest of Yorks & Humberside, Tyne & Wear, Rest of North; (2) ‘E/South’ containing Rest of South East, South West and East Anglia; (3) ‘London’ includes inner and outer London; and finally (4) ‘Scotland’ and (5) ‘Wales’. Northern Ireland is excluded from our analysis because BHPS sample sizes in Northern Ireland were too small to cross-classify by age and gender. We also exclude persons who did not report a positive income.

Incomes estimates for the cross-classification age by gender by region based on the linear model with two-way interaction effects age by sex are shown in Table 4. The BHPS also collects information from a respondent on his/her three best friends, consisting of the genders and ages of these friends, duration of friendships, frequency of contact, distances to the friends, their job/employment statuses, and their ethnicities. A contextual model can take a contextual effect based on the three friends and the collected variables into account.

Table 4. Gender by age by region cross-classification of estimated mean annual income in pounds sterling, using BHPS data and with weighting based on model with age by sex interactions.

Gender	Age	Region				
		London	E/North	E/South	Scotland	Wales
Female	≤ 18	2,668	3,666	2,279	1,678	4,120
	19–21	6,989	7,134	6,818	7,036	8,976
	22–30	17,068	12,759	13,403	14,147	12,774
	31–50	20,266	14,881	16,514	16,043	14,565
	51–64	12,129	10,725	10,931	11,822	11,279
	≥ 65	8,582	7,283	7,952	7,851	8,793
Male	≤ 18	1,257	3,896	2,180	2,578	4,897
	19–21	10,102	7,600	9,735	6,735	16,270
	22–30	15,617	15,617	18,294	20,719	15,960
	31–50	23,884	23,884	29,216	29,216	21,947
	51–64	20,186	20,186	23,293	27,237	19,305
	≥ 65	11,775	11,775	14,540	12,055	12,613

Table 5. Change in estimated mean annual income when BHPS data are weighted using the CN model based on age by sex interactions plus a main effect for maleness.

Gender	Age	Region				
		London	E/North	E/South	Scotland	Wales
Female	≤ 18	162	− 10 ²	− 110 ⁴	458 ⁴	391 ⁴
	19– 21	324 ¹	47 ⁴	0	− 127 ⁴	− 372 ⁴
	22– 30	22 ²	2 ⁴	− 3 ⁴	− 10 ⁴	− 7 ⁴
	31– 50	77 ²	− 2 ⁴	− 6 ⁴	− 3 ⁴	− 29 ⁴
	51– 64	− 8	− 1 ⁴	5 ⁴	− 6 ⁴	− 2 ⁴
	≥ 65	− 7	5 ⁴	− 1	4 ⁴	4 ⁴
Male	≤ 18	− 176 ³	− 20 ⁴	99 ⁴	106 ⁴	− 771 ⁴
	19– 21	139 ¹	− 5 ⁴	− 15 ⁴	71 ⁴	− 710 ⁴
	22– 30	− 23	41 ⁴	− 47 ⁴	− 89 ⁴	20 ⁴
	31– 50	− 134 ¹	30 ⁴	− 60 ⁴	114 ⁴	− 14 ⁴
	51– 64	− 214 ¹	40 ⁴	− 50 ⁴	209 ⁴	56 ⁴
	≥ 65	119 ¹	− 5 ⁴	− 2 ³	− 74 ⁴	12 ⁴

1, 2, 3, 4 Difference larger than 1, 2, 3, 4 estimated standard errors.

This situation of having contextual variables \mathbf{U}_s available from the sample s corresponds to the SS+SR case, as \mathbf{U}_s is defined by $\mathbf{U}_s = \mathbf{W}_s \tilde{\mathbf{X}}$ with $\mathbf{W}_s = \text{Diag}(\mathbf{Z}_s \mathbf{1}_N)^{-1} \mathbf{Z}_s$ and $\mathbf{Z}_s = (\mathbf{Z}_{ss}, \mathbf{Z}_{sr})$, see Equation (3). As the number of friends is fixed, $P(Y_{ij} = 1) = p$ does not need to be estimated, but is known to be $p = 3/(N - 1)$, as there are $N - 1$ candidate friends.

Table 5 shows the difference in estimates to Table 4 when estimation is based on a contextual model with effects age by sex and a contextual gender effect, where gender is a binary variable indicating whether a friend is male. In this context the model implies that a person’s income is not only predicted by age by sex but also by the average income of the person’s three best friends. For further details on how these estimates were obtained, see Suesse and Chambers (2012).

Table 5 shows that the application of the contextual model leads to substantially different results, as many differences are larger than 1, 2, 3, and often 4 standard errors.

We calculated the value of $\hat{\Delta}$ for all cells presented in Tables 4 and 5. For most of the cells, $\hat{\Delta} > 0$ and hence the standard model should be used, however there are some cells for which indeed $\hat{\Delta} < 0$ and the contextual model is deemed as better in terms of a lower predicted MSE. For the London region the cells with $\hat{\Delta} < 0$ are with the estimated improvement of the MSE relative to the standard model in brackets: female and 19–21 years (6%), female and 31–50 years (0.1%), and male and ≥ 65 years (1%). These results are based on reconstructed data using the publicly available survey weights, so these statements are to be treated with caution, as the real data might yield different results.

8. Discussion

At the end of Section 1, we stated that our aim in this article is to address the questions: (i) Is embedding network information useful for survey estimation? (ii) If the answer to (i) is

yes, then which models are potentially useful? and (iii) How much network data needs to be collected in order to obtain potentially higher precision for survey estimation? Given the simulation results that we present in Section 6, our tentative answer to (i) is yes, and our corresponding answer to (ii) is the CN and AR models when either model is true, because in both cases the mean of the response depends on the network. Our simulation results provide some evidence that this conclusion may hold more generally.

However when the mean does not depend on the network, as is the case under the ND model, our results suggest that ignoring the network does not result in a significant loss of efficiency. We have also investigated this for other ‘network covariance’ models, where the mean structure is unaffected by the network, and we have observed similar results, see [Suesse and Chambers \(2014\)](#). In effect, ignoring the network under the CN and AR models leads to a misspecification of the mean model, but this does not apply for the ND (and similar) models. Finally, our answer to (iii) is that in realistic applications it will usually be impossible to collect the full network, and our simulation results are some evidence that when either the CN model or the AR model is true then both \mathbf{Z}_{ss} and \mathbf{Z}_{sr} must be collected or alternatively the contextual sample information \mathbf{U}_s along with an estimate of $p = P(Z_{ij} = 1)$ must be available, as for the BHPS data set, in order to obtain efficiency gains. Knowledge of \mathbf{Z}_{ss} alone is not enough.

In practice, we suggest a careful model fitting exercise be carried out before attempting to use either the CN model or the AR model for survey estimation. Given the numerical difficulties with fitting the AR model, see [Suesse \(2012a\)](#), we recommend that the CN model be used if it is a reasonable fit to the data and when $\hat{\Delta} < 0$, otherwise caution is warranted and ignoring the network might be the best option.

Clearly, more extensive information on networks needs to be collected in conjunction with standard survey data to gain further insight into the usefulness of network models for survey estimation. However, it is extremely unlikely that in practical applications complete network data will be available, in which case the issue of imputation for missing network data arises. In this article we base this imputation on the fact that the sample proportion of links per individual is a simple nonparametric estimator of the marginal probability of an unobserved link. A reasonable question to ask then is whether it is better to use an imputation method based on $E(\mathbf{Z}^{mis} | \mathbf{Z}^{obs} = \mathbf{z}^{obs})$? The numerical intensity of the MCMC methods used to fit network models like the ERGM when population sizes are large meant that we could not fully explore this issue here. There is current research that tries to address some of these issues ([Koskinen et al. 2010](#)), but more is required, because even if the time issue of fitting a partially observed network is solved, simulating many ERGM networks to obtain $E(\mathbf{Z}^{mis} | \mathbf{Z}^{obs} = \mathbf{z}^{obs})$ or even just $E(\mathbf{Z})$ for large $N > 100,000$ still appears infeasible.

[Suesse and Chambers \(2014\)](#) considered the case of known ERGM parameters (without estimating them) and then applied a more sophisticated method, however even this method was consistently worse than the simple method, despite having full knowledge of the ERGM parameters. Based on these results we anticipate that more sophisticated imputation methods are unlikely to lead to substantial efficiency gains in most cases. Our simulation results indicate marginal differences between the SS+SR case and where \mathbf{Z}_U is known. We therefore hypothesize that more sophisticated imputation methods will also only provide marginal gains for estimation of a population total, and not alter the

conclusions of this article. Efficient imputation methods when the target of inference is more complex require further research.

All network models considered in this article assume that the value of the response variable Y for an individual in the study population depends on a linear combination of the values of this variable for the other individuals in the population that are linked to this person in the network. If there is an implicit ordering in the strength of these links, then this can be allowed for in the network model for Y . For example, in the case of a ‘best friend’ network, where the friendships are ordered by their strength, one can modify the CN model so that there is a separate parameter for each level of ‘best friend’, see [Friedkin \(1990\)](#) for similar examples. To illustrate, in the BHPS application, when this extended contextual model is fitted, a Wald test for equality of these effects supports the assumption of a common effect.

The use of ERGMs to model the network and the use of the three main regression models in Section 2 using the network as additional information might be restrictive. There are many other approaches to model networks, for example adding latent variables to a logistic regression model, as proposed by [Handcock et al. \(2007\)](#), but also many extensions to include network information in a regression model, see [Leenders \(2002\)](#) for some model extensions and the various options to define the weight matrix \mathbf{W} based on the network \mathbf{Z} . The use of particular models might be beneficial or detrimental and exploring the use of alternative models could be useful. The same holds for using particular network structures in the modelling approaches. Conducting other simulation studies to investigate the merits of different models is subject to future research.

Finally, we note that throughout this paper we have assumed that the method of sampling is independent of the network structure given the available population auxiliary information. In effect, we assume that measurement of the network is something that is done on the sample (as in our BHPS application), rather than sampling being something that is carried out on the network. However, there are important applications, see [Thompson and Seber \(1996\)](#), where inclusion in sample depends on being linked to another sampled individual via a network. It is clear that in these cases we cannot treat the observed network structure in \mathbf{Z}_{ss} and \mathbf{Z}_{sr} in the same way as we have in this article, and this ‘informative’ method of sampling needs to be taken into account when we attempt to impute the unknown components of \mathbf{Z} . Work on this problem is continuing.

Appendix

A. Further Simulation Results

Table 6. $n = 100$: ERGM(3) and ERGM(10) network with X drawn randomly from $\{1, \dots, 9\}$. Ratio of average lengths of nominal 95% Gaussian CIs (EBLUP/BLUP), with % actual coverage in subscript.

	Population Data Generated Under Model					
	ERGM(3)			ERGM(10)		
	CN	AR	ND	CN	AR	ND
Actual BLUP av(length)	1, 128 _{94.2}	1, 273 _{95.0}	1, 273 _{95.0}	1, 127 _{94.1}	1, 160 _{95.2}	1, 160 _{95.2}
Relative av(length) EBLUP based on						
True Model \mathbf{Z}_U known	0.98 _{93.2}	0.98 _{94.2}	0.98 _{94.1}	0.98 _{93.4}	0.98 _{94.4}	0.98 _{94.3}
CN						
SS+SR+RS	0.98 _{91.7}	1.02 _{94.1}	1.01 _{94.9}	0.98 _{93.2}	0.99 _{94.2}	0.99 _{94.7}
SS+SR	0.98 _{91.9}	1.02 _{94.1}	1.01 _{94.9}	0.98 _{93.1}	0.99 _{94.4}	0.99 _{94.6}
SS	1.47 _{92.9}	1.16 _{94.8}	1.01 _{94.6}	1.11 _{93.4}	1.03 _{94.9}	0.99 _{94.9}
AR						
SS+SR+RS	1.08 _{92.5}	0.99 _{93.6}	1.00 _{94.2}	0.98 _{92.8}	0.98 _{94.0}	0.98 _{94.5}
SS+SR	1.01 _{92.0}	0.99 _{92.9}	1.00 _{94.4}	0.98 _{93.1}	0.98 _{94.0}	0.98 _{94.5}
SS	1.46 _{92.4}	1.13 _{94.2}	0.99 _{94.8}	1.11 _{93.3}	1.02 _{94.5}	0.98 _{94.1}
ND						
SS+SR+RS	1.08 _{83.9}	0.99 _{91.5}	1.00 _{94.6}	0.98 _{89.1}	0.98 _{94.1}	0.98 _{94.3}
SS+SR	1.42 _{92.4}	1.12 _{94.0}	0.98 _{94.2}	1.11 _{93.2}	1.02 _{94.7}	0.98 _{94.5}
SS	1.49 _{92.8}	1.14 _{93.7}	0.98 _{94.2}	1.12 _{93.7}	1.02 _{94.7}	0.98 _{94.5}
Standard	1.49 _{93.6}	1.17 _{94.5}	1.01 _{94.8}	1.13 _{93.7}	1.03 _{95.2}	0.99 _{94.6}

B. Derivations

Suppose the contextual model (CN Model) holds $Y_i = \mathbf{X}_i\beta + \mathbf{U}_i\gamma + \epsilon_i$ such that $E(Y_i) = \mathbf{X}_i\beta + \mathbf{U}_i\gamma$, $\text{Var}(Y_i) = \sigma_i^2 = v_i\sigma^2$ and $\text{Cov}(Y_i, Y_j) = 0$ for $i \neq j$. The column vector β is of length p and γ of length q . When not using the contextual information (standard model), then $\gamma = 0$ and $E(Y_i) = \mathbf{X}_i\beta$.

Define $\mathbf{H}_i = (\mathbf{X}_i, \mathbf{U}_i)$ and $\lambda = (\beta^\top, \gamma^\top)^\top$. Under independence, the weights of the BLUP $\hat{T} = \mathbf{w}_H^\top \mathbf{Y}_s$ given by (8) simplify to \mathbf{w}_H

$$\mathbf{w}_H = \mathbf{1}_s + \mathbf{D}_v^{-1} \mathbf{H}_s \{ \mathbf{H}_s^\top \mathbf{D}_v^{-1} \mathbf{H}_s \}^{-1} \mathbf{T}_{Hr}^\top, \quad (12)$$

where $\mathbf{T}_{Hr} \equiv \sum_{i \in r} \mathbf{H}_i = \mathbf{H}_r \mathbf{1}_r$ is a row vector of length $q + p$, similarly the population totals are defined, for example $\mathbf{T}_H \equiv \sum_{i \in U} \mathbf{H}_i = \mathbf{H} \mathbf{1}$.

We aim at comparing the MSE of the CN and the standard model, but also under the situation SS+SR, that is when contextual information needs to be estimated for the non-sample.

$$\begin{aligned}
 \text{MSE}[\hat{T}] &= E[(\hat{T} - T_Y)^2] \\
 &= E \left\{ \sum_{i \in s} w_{Hi} Y_i - \sum_{i \in U} Y_i \right\}^2 + \text{Var} \left[\sum_{i \in s} (w_{Hi} - 1) Y_i - \sum_{i \in r} Y_i \right] \\
 &= E \{ \mathbf{w}_H^\top \mathbf{Y}_s - \mathbf{Y}^\top \mathbf{1} \}^2 + \text{Var} [(\mathbf{w}_H - \mathbf{1}_s)^\top \mathbf{Y}_s - \mathbf{Y}_r^\top \mathbf{1}_r]
 \end{aligned} \tag{13}$$

Then according to [Clark and Chambers \(2008\)](#), $\text{MSE}[\hat{T}]$ can be re-written as

$$\begin{aligned}
 \text{MSE}[\hat{T}] &= \left\{ \left(\sum_{i \in s} w_{Hi} \mathbf{H}_i - \sum_{i \in U} \mathbf{H}_i \right) \boldsymbol{\lambda} \right\}^2 + \sum_{i \in s} (w_{Hi} - 1)^2 \text{Var}(Y_i) \\
 &\quad + \sum_{i \in r} \text{Var}(Y_i)
 \end{aligned} \tag{14}$$

$$= \mathbf{d}_H (\boldsymbol{\lambda} \boldsymbol{\lambda}^\top) \mathbf{d}_H^\top + \sum_{i \in s} (w_{Hi} - 1)^2 \sigma_i^2 + \sum_{i \in r} \sigma_i^2 \tag{15}$$

where $\mathbf{d}_H = \sum_{i \in s} w_{Hi} \mathbf{H}_i - \mathbf{T}_H$. The term $\boldsymbol{\lambda} \boldsymbol{\lambda}^\top$ can be estimated by $\hat{\boldsymbol{\lambda}} \hat{\boldsymbol{\lambda}}^\top - \widehat{\text{Var}}(\hat{\boldsymbol{\lambda}})$.

Let the EBLUP under the CN model using \mathbf{H}_i be denoted by \hat{T}_C and that of standard model using only \mathbf{X}_i for $i \in U$ by \hat{T}_I .

Then the difference $\Delta \equiv \text{MSE}[\hat{T}_C] - \text{MSE}[\hat{T}_I]$ can be estimated by

$$\begin{aligned}
 \hat{\Delta} &= \mathbf{d}_H (\hat{\boldsymbol{\lambda}} \hat{\boldsymbol{\lambda}}^\top - \widehat{\text{Var}}(\hat{\boldsymbol{\lambda}})) \mathbf{d}_H^\top - \mathbf{d}_X (\hat{\boldsymbol{\lambda}} \hat{\boldsymbol{\lambda}}^\top - \widehat{\text{Var}}(\hat{\boldsymbol{\lambda}})) \mathbf{d}_X^\top \\
 &\quad + \sum_{i \in s} (w_{Hi} - 1)^2 \hat{\sigma}_i^2 - \sum_{i \in s} (w_{Xi} - 1)^2 \hat{\sigma}_i^2.
 \end{aligned}$$

Since we assume that the contextual model holds $\mathbf{d}_H = 0$ and

$$\hat{\Delta} = -\mathbf{d}_X (\hat{\boldsymbol{\lambda}} \hat{\boldsymbol{\lambda}}^\top - \widehat{\text{Var}}(\hat{\boldsymbol{\lambda}})) \mathbf{d}_X^\top + \sum_{i \in s} (w_{Hi} - 1)^2 \hat{\sigma}_i^2 - \sum_{i \in s} (w_{Xi} - 1)^2 \hat{\sigma}_i^2.$$

According to [Clark and Chambers \(2008\)](#) the contextual model is chosen when $\hat{\Delta} < 0$.

For the simple case of one contextual variable U_i the result of [Clark and Chambers \(2008\)](#) applies and $\hat{\Delta}$ simplifies to

$$\hat{\Delta} = T_{Ur}^2 (2\hat{\sigma}_u^2 S_u^{-1} - \hat{\gamma}^2), \tag{16}$$

and the contextual model is chosen when $\hat{\gamma}^2 > 2\hat{\sigma}_u^2 S_u^{-1}$, where $S_c \equiv \sum_{i \in s} c_i$ and $c_i \equiv U_i - C^\top \mathbf{X}_i^\top$ with $C \equiv (\sum_{i \in s} \mathbf{X}_i^\top \mathbf{X}_i)^{-1} \mathbf{X}_i^\top U_i$.

Now suppose that \mathbf{X}_i is known for all units $i \in U$ and \mathbf{U}_i is only known for $i \in s$, that is the population totals $\mathbf{T}_X \equiv \sum_{i \in U} \mathbf{X}_i$ of the covariates \mathbf{X}_i are known, whereas $\mathbf{T}_{Ur} \equiv \sum_{i \in r} \mathbf{U}_i$ is unknown and must be estimated and its estimate is denoted by $\hat{\mathbf{T}}_{Ur}$.

The weights w_i are now a function of the sample values of \mathbf{X}_i and \mathbf{U}_i but also of the known population total \mathbf{T}_X and the estimated non-sample total of the contextual variables $\hat{\mathbf{T}}_{U_r}$. The weights depending on $\hat{\mathbf{T}}_{U_r}$ are denoted by \hat{w}_i . The EBLUP based on the estimated $\hat{\mathbf{T}}_{U_r}$ is denoted by \hat{T} .

Now the MSE can be expressed as follows

$$\begin{aligned} \text{MSE}[\hat{T}] &= E[(\hat{T} - T_Y)]^2 \\ &= (E(\hat{T} - T_Y))^2 + \text{Var}(\hat{T} - T_Y) \\ &= (E[E(\hat{T} - T_Y | \hat{\mathbf{T}}_{U_r})])^2 + \text{Var}[E(\hat{T} - T_Y | \hat{\mathbf{T}}_{U_r})] + E[\text{Var}(\hat{T} - T_Y | \hat{\mathbf{T}}_{U_r})] \end{aligned} \quad (17)$$

Previously with known \mathbf{T}_{U_r}

$$\begin{aligned} E(\hat{T} - T_Y) &= \left(\sum_{i \in s} w_i \mathbf{X}_i - \sum_{i \in U} \mathbf{X}_i \right) \beta + \left(\sum_{i \in s} w_i \mathbf{U}_i - \sum_{i \in U} \mathbf{U}_i \right) \gamma \\ &= \left(\sum_{i \in s} w_i \mathbf{H}_i - \sum_{i \in U} \mathbf{H}_i \right) \lambda. \end{aligned}$$

Now the conditional expectation given $\hat{\mathbf{T}}_{U_r}$ gives

$$E(\hat{T} - T_Y | \hat{\mathbf{T}}_{U_r}) = \left(\sum_{i \in s} \hat{w}_i \mathbf{X}_i - \sum_{i \in U} \mathbf{X}_i \right) \beta + \left(\sum_{i \in s} (\hat{w}_i - 1) \mathbf{U}_i - \hat{\mathbf{T}}_{U_r} \right) \gamma$$

The outer expectations/variances are always with respect to the distribution of $\hat{\mathbf{T}}_{U_r}$ and are usually suppressed, unless necessary. Now assuming that $\hat{\mathbf{T}}_{U_r}$ is an unbiased estimate of \mathbf{T}_{U_r} , equivalently $\hat{\mathbf{T}}_U$ is unbiased estimate of \mathbf{T}_U with estimated (co)variance matrix $\mathbf{V}_{\hat{\mathbf{T}}_U}$. It follows that the estimated weights \hat{w}_i depending on $\hat{\mathbf{T}}_{U_r}$, see Equation (12), are also unbiased, that is $E(\hat{w}_i) = w_i$. First we obtain a variance estimate for $\hat{\mathbf{w}}$ using $\hat{\mathbf{T}}_{H_r} = (\mathbf{T}_{X_r}, \hat{\mathbf{T}}_{U_r})$

$$\begin{aligned} \text{Var}(\hat{\mathbf{w}}) &= \text{Var} \left(\mathbf{1}_s + \mathbf{D}_V^{-1} \mathbf{H}_s \{ \mathbf{H}_s^\top \mathbf{D}_V^{-1} \mathbf{H}_s \}^{-1} \hat{\mathbf{T}}_{H_r}^\top \right) \\ &= \mathbf{D}_V^{-1} \mathbf{H}_s \{ \mathbf{H}_s^\top \mathbf{D}_V^{-1} \mathbf{H}_s \}^{-1} \mathbf{V}_{\hat{\mathbf{T}}_H} \{ \mathbf{H}_s^\top \mathbf{D}_V^{-1} \mathbf{H}_s \}^{-1} \mathbf{H}_s^\top \mathbf{D}_V^{-1} \end{aligned}$$

with

$$\mathbf{V}_{\hat{\mathbf{T}}_H} = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{V}_{\hat{\mathbf{T}}_U} \end{pmatrix}.$$

Some blocks are zero because \mathbf{T}_X is known, hence there is no variability with respect to the distribution of $\hat{\mathbf{T}}_{U_r}$. By defining the $n \times (p + q)$ matrix $\mathbf{B} = \mathbf{D}_V^{-1} \mathbf{H}_s \{ \mathbf{H}_s^\top \mathbf{D}_V^{-1} \mathbf{H}_s \}^{-1}$ and partitioning as

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_p & \mathbf{B}_q \end{pmatrix},$$

such that \mathbf{B}_p is of dimension $n \times p$ and \mathbf{B}_q of dimension $n \times q$. We can write now

$$\text{Var}(\hat{\mathbf{w}}) = \mathbf{B}_q \mathbf{V}_{\hat{\mathbf{T}}_U} \mathbf{B}_q^\top. \quad (18)$$

We obtain

$$\begin{aligned} E(E(\hat{T} - T_Y | \hat{\mathbf{T}}_{Ur})) &= E \left\{ \left(\sum_{i \in s} \hat{w}_i \mathbf{X}_i - \sum_{i \in U} \mathbf{X}_i \right) \beta + \left(\sum_{i \in s} (\hat{w}_i - 1) \mathbf{U}_i - \hat{\mathbf{T}}_{Ur} \right) \gamma \right\} \\ &= \left(\sum_{i \in s} w_i \mathbf{X}_i - \sum_{i \in U} \mathbf{X}_i \right) \beta + \left(\sum_{i \in s} (w_i - 1) \mathbf{U}_i - \sum_{i \in r} \mathbf{U}_i \right) \gamma \\ &= \left(\sum_{i \in s} w_i \mathbf{H}_i - \sum_{i \in U} \mathbf{H}_i \right) \boldsymbol{\lambda}. \end{aligned}$$

By ignoring terms that do not vary with respect to $\hat{\mathbf{T}}_{Ur}$ and collecting the remaining terms we obtain

$$\begin{aligned} \text{Var}(E(\hat{T} - T_Y | \hat{\mathbf{T}}_{Ur})) &= \text{Var} \left\{ \left(\sum_{i \in s} \hat{w}_i \mathbf{X}_i - \sum_{i \in U} \mathbf{X}_i \right) \beta + \left(\sum_{i \in s} (\hat{w}_i - 1) \mathbf{U}_i - \hat{\mathbf{T}}_{Ur} \right) \gamma \right\} \\ &= \text{Var} \{ \hat{\mathbf{w}}^\top \mathbf{H}_s \boldsymbol{\lambda} - \hat{\mathbf{T}}_{Ur} \gamma \} \\ &= (\mathbf{H}_s \boldsymbol{\lambda})^\top \text{Var}(\hat{\mathbf{w}}) \mathbf{H}_s \boldsymbol{\lambda} + \gamma^\top \mathbf{V}_{\hat{\mathbf{T}}_U} \gamma - 2 \text{Cov}(\hat{\mathbf{w}}^\top \mathbf{H}_s \boldsymbol{\lambda}, \hat{\mathbf{T}}_{Ur} \gamma) \\ &= (\mathbf{H}_s \boldsymbol{\lambda})^\top \mathbf{B}_q \mathbf{V}_{\hat{\mathbf{T}}_U} \mathbf{B}_q^\top \mathbf{H}_s \boldsymbol{\lambda} + \gamma^\top \mathbf{V}_{\hat{\mathbf{T}}_U} \gamma - 2 \text{Cov}(\hat{\mathbf{w}}^\top \mathbf{H}_s \boldsymbol{\lambda}, \hat{\mathbf{T}}_{Ur} \gamma) \\ &= (\mathbf{B}_q^\top \mathbf{H}_s \boldsymbol{\lambda})^\top \mathbf{V}_{\hat{\mathbf{T}}_U} \mathbf{B}_q^\top \mathbf{H}_s \boldsymbol{\lambda} + \gamma^\top \mathbf{V}_{\hat{\mathbf{T}}_U} \gamma - 2 (\mathbf{H}_s \boldsymbol{\lambda})^\top \text{Cov}(\hat{\mathbf{w}}, \hat{\mathbf{T}}_{Ur}) \gamma \\ &= (\mathbf{B}_q^\top \mathbf{H}_s \boldsymbol{\lambda})^\top \mathbf{V}_{\hat{\mathbf{T}}_U} \mathbf{B}_q^\top \mathbf{H}_s \boldsymbol{\lambda} + \gamma^\top \mathbf{V}_{\hat{\mathbf{T}}_U} \gamma - 2 (\mathbf{H}_s \boldsymbol{\lambda})^\top \mathbf{B}_q \mathbf{V}_{\hat{\mathbf{T}}_U} \gamma \\ &= (\mathbf{B}_q^\top \mathbf{H}_s \boldsymbol{\lambda})^\top \mathbf{V}_{\hat{\mathbf{T}}_U} \mathbf{B}_q^\top \mathbf{H}_s \boldsymbol{\lambda} + \gamma^\top \mathbf{V}_{\hat{\mathbf{T}}_U} \gamma - 2 (\mathbf{B}_q^\top \mathbf{H}_s \boldsymbol{\lambda})^\top \mathbf{V}_{\hat{\mathbf{T}}_U} \gamma \\ &= (\mathbf{B}_q^\top \mathbf{H}_s \boldsymbol{\lambda} - \gamma)^\top \mathbf{V}_{\hat{\mathbf{T}}_U} (\mathbf{B}_q^\top \mathbf{H}_s \boldsymbol{\lambda} - \gamma) \end{aligned}$$

Using $E(X - 1)^2 a_i = (E(X) - 1)^2 a_i + \text{Var}(X) a_i$ for any r.v. X and constant a_i , we obtain

$$\begin{aligned} E(\text{Var}(\hat{T} - T_Y | \hat{\mathbf{T}}_{Ur})) &= E \left(\sum_{i \in s} (\hat{w}_i - 1)^2 \sigma_i^2 + \sum_{i \in r} \sigma_i^2 \right) \\ &= \sum_{i \in s} (w_i - 1)^2 \sigma_i^2 + \sum_{i \in r} \sigma_i^2 + \sum_{i \in s} \sigma_i^2 \text{Var}(\hat{w}_i) \\ &= \sum_{i \in s} (w_i - 1)^2 \sigma_i^2 + \sum_{i \in r} \sigma_i^2 + (\sigma^2)^\top \text{Diag}(\mathbf{B}_q \mathbf{V}_{\hat{\mathbf{T}}_U} \mathbf{B}_q^\top), \end{aligned}$$

where $\sigma^2 = (\sigma_1^2, \dots, \sigma_n^2)^\top$ and $\text{Diag}(\mathbf{A})$ gives the vector on the diagonal of matrix \mathbf{A} .

Finally using that under the CN model $\mathbf{d}_H = 0$ and using $\mathbf{b} = \mathbf{B}_q^\top \mathbf{H}_s \boldsymbol{\lambda} - \gamma = [(\mathbf{H}_s^\top \mathbf{H}_s)^{-1} \mathbf{H}_s^\top \mathbf{H}_s]_q \boldsymbol{\lambda} - \gamma = \gamma - \gamma = 0$ ($[\mathbf{A}]_q$ refers to the last q rows of matrix \mathbf{A} ,

we obtain

$$\begin{aligned}
 MSE[\hat{T}] &= \left\{ \left(\sum_{i \in s} w_i \mathbf{H}_i - \sum_{i \in U} \mathbf{H}_i \right) \boldsymbol{\lambda} \right\}^2 + \sum_{i \in s} (w_i - 1)^2 \sigma_i^2 + \sum_{i \in r} \sigma_i^2 \\
 &\quad + \mathbf{b}^\top \mathbf{V}_{\hat{T}_U} \mathbf{b} + (\sigma^2)^\top \text{Diag}(\mathbf{B}_q \mathbf{V}_{\hat{T}_U} \mathbf{B}_q^\top) \\
 &= \mathbf{d}_H (\boldsymbol{\lambda} \boldsymbol{\lambda}^\top) \mathbf{d}_H^\top + \sum_{i \in s} (w_i - 1)^2 \sigma_i^2 + \sum_{i \in r} \sigma_i^2 \\
 &\quad + 0^\top \mathbf{V}_{\hat{T}_U} 0 + (\sigma^2)^\top \text{Diag}(\mathbf{B}_q \mathbf{V}_{\hat{T}_U} \mathbf{B}_q^\top) \\
 &= \sum_{i \in s} (w_i - 1)^2 \sigma_i^2 + \sum_{i \in r} \sigma_i^2 + (\sigma^2)^\top \text{Diag}(\mathbf{B}_q \mathbf{V}_{\hat{T}_U} \mathbf{B}_q^\top),
 \end{aligned} \tag{19}$$

which differs from (14) by the additional last term.

In practice $MSE[\hat{T}]$ can be estimated by

$$\widehat{MSE}[\hat{T}] = \sum_{i \in s} (\hat{w}_{Hi} - 1)^2 \sigma_i^2 + \sum_{i \in r} \sigma_i^2 + (\sigma^2)^\top \text{Diag}(\mathbf{B}_q \hat{\mathbf{V}}_{\hat{T}_U} \mathbf{B}_q^\top), \tag{20}$$

Now Δ can be estimated by

$$\begin{aligned}
 \hat{\Delta} &= -\mathbf{d}_X (\hat{\boldsymbol{\lambda}} \hat{\boldsymbol{\lambda}}^\top - \widehat{\text{Var}}(\boldsymbol{\lambda})) \mathbf{d}_X^\top + \sum_{i \in s} (w_{Hi} - 1)^2 \hat{\sigma}_i^2 - \sum_{i \in s} (w_{Xi} - 1)^2 \hat{\sigma}_i^2 \\
 &\quad + (\sigma^2)^\top \text{Diag}(\mathbf{B}_q \hat{\mathbf{V}}_{\hat{T}_U} \mathbf{B}_q^\top).
 \end{aligned}$$

We propose to use the contextual model when $\hat{\Delta} < 0$ and the model without the contextual effects otherwise. When only one contextual variable is provided and $\sigma_i^2 = \sigma^2$, then the first three terms are replaced as in [Clark and Chambers \(2008\)](#) and we obtain

$$\hat{\Delta} = \hat{T}_{Ur}^2 (2\hat{\sigma}^2 S_u^{-1} - \hat{\gamma}^2) + \hat{\sigma}^2 \mathbf{1}_s^\top \text{Diag}(\mathbf{B}_q \mathbf{B}_q^\top) \hat{V}_{\hat{T}_U},$$

where the last term disappears when the total T_U is known, as then $\hat{V}_{\hat{T}_U} = 0$, and then the formula coincides with (16).

C. Estimation of Contextual Population Information

For notational convenience, we use $\mathbf{1}_s$ as the vector of length n , $\mathbf{1}_r$ as the vector of length $N - n$, $\mathbf{1}_N$ as the vector of length N and similarly the matrices of ones $\mathbf{1}_{s,N}$, $\mathbf{1}_{r,N}$, $\mathbf{1}_{N,N}$ with appropriate sizes.

The contextual population information is often not available. As defined before $\tilde{\mathbf{X}}$ is a vector of population covariates, then the vector of contextual population information can be obtained by the formula

$$\mathbf{U} = \mathbf{W} \tilde{\mathbf{X}} \tag{21}$$

$$\mathbf{W} = \mathbf{Z} / \mathbf{Z} \mathbf{1}_{N,N}, \tag{22}$$

where the division refers to element-wise division of the matrices, it is an expression for dividing each row of \mathbf{Z} by the number of friends of person i (the sum of row i). Often the number of friends is fixed, for example for the BHPS, each person has exactly three friends, then $\mathbf{W} = \frac{1}{3}\mathbf{Z}$.

Suppose that \mathbf{U} is available for the sample, that is \mathbf{U}_s , and that the full \mathbf{Z} is not available. We outline here a simple method to obtain $\hat{\mathbf{T}}_{\mathbf{U}}$ and $\hat{\mathbf{V}}_{\mathbf{T}_{\mathbf{U}}}$ under a simple ERGM and a simple estimator for $\hat{\mathbf{Z}}$.

The ERGM under consideration is outlined in Sections 4 and 6 and has an edges statistic and a GWESP statistic. One can show that (e.g., when simulating under this ERGM) that $P(Z_{ij} = 1) = p$ is a constant (irrespective of whether the network is undirected or directed), that is for each dyad the same marginal probability applies. However dyads are usually not independent.

Under the situation SS+SR we can estimate p by

$$\hat{p} = \frac{1}{n_s(N-1)} \sum_{i \in s} \sum_{j \neq i, j \in s, r} Z_{ij} = \mathbf{1}_s^\top \mathbf{Z}_s \mathbf{1}_N / (n_s(N-1)).$$

Under independence and from general properties of a Binomial random variable $\text{Var}(\hat{p}) = \hat{p}(1-\hat{p})/(n_s(N-1))$. Due to dependence the real variance will be larger (as the covariances between dyads are usually positive) and might be estimated by simulating under an ERGM. However from simulations, see for example [Suesse \(2012b\)](#) where a correlation matrix has to be estimated by simulating a large number of ERGM networks to fit a certain class of network models, it can also be shown that most correlations between dyads are near zero and that only some correlations for dyads sharing a node, for example Y_{ij} and Y_{ik} , are positive but small in magnitude, for example 0.02. Hence the correlation structure under the independence assumption (meaning that all correlations are exactly zero) does not deviate much from the true correlation structure and we expect the estimated variance derived under the binomial distribution to hold approximately.

Now we use \hat{p} and the simple estimator $\text{Var}(\hat{p})$ to estimate $\mathbf{T}_{\mathbf{U}}$ and $\mathbf{V}_{\hat{\mathbf{T}}_{\mathbf{U}}}$. Using (21) and (22) we can write under SS + SR using $\mathbf{Z}_r = (\mathbf{Z}_{rs}, \mathbf{Z}_{rr})$

$$\mathbf{T}_{\mathbf{U}} = \mathbf{1}_N^\top \mathbf{U} = \mathbf{1}_s^\top \mathbf{U}_s + \mathbf{1}_r^\top \mathbf{U}_r \quad (23)$$

and

$$\mathbf{U}_r = \frac{\mathbf{a}_1}{\mathbf{a}_2} = \frac{\mathbf{Z}_r \tilde{\mathbf{X}}}{\mathbf{Z}_r \mathbf{1}_N}.$$

Now \mathbf{Z}_r is not observed and must be estimated, here by $\mathbf{P}_r = E(\mathbf{Z}_r)$. Note that \mathbf{P}_r is a $(N-n) \times N$ matrix, but has zeros along the off-diagonal that correspond to the diagonal of \mathbf{Z} because \mathbf{Z} has zero diagonal entries. The other entries are constant and equal \hat{p} .

Now

$$\mathbf{a}_1 = \hat{\mathbf{P}}_r \tilde{\mathbf{X}} = \hat{p} \{N \mathbf{1}_r \tilde{\mathbf{X}} - \tilde{\mathbf{X}}_r\}$$

with the population average of $\tilde{\mathbf{X}}$ denoted by $\bar{\tilde{\mathbf{X}}}$ and

$$\mathbf{a}_2 = \hat{\mathbf{P}}_r \mathbf{1}_N = \hat{p}(N-1) \mathbf{1}_r,$$

provided the number of friends is not fixed and unknown. If fixed, as for the BHPS, the network is not required to estimate p , because p is known and equals $\frac{3}{N-1}$, as each person has three friends in the population with $N - 1$ possible candidates. Then \mathbf{a}_2 has simple structure, it is a vector with elements equal to this number, that is 3.

The estimator of \mathbf{U}_r is

$$\hat{\mathbf{U}}_r = \frac{\hat{p}\{N\mathbf{1}_r\bar{\mathbf{X}} - \tilde{\mathbf{X}}_r\}}{\hat{p}(N-1)\mathbf{1}_r} = \frac{N\mathbf{1}_r\bar{\mathbf{X}} - \tilde{\mathbf{X}}_r}{(N-1)\mathbf{1}_r}$$

and hence an estimator $\hat{\mathbf{T}}_U$ of \mathbf{T}_U is obtained by replacing \mathbf{U}_r by $\hat{\mathbf{U}}_r$ in (23).

Assume for simplicity the denominator is fixed (as for the BHPS) and we have only one contextual variable, then we approximate $\mathbf{V}_{\hat{T}_U}$ by

$$\hat{\mathbf{V}}_{\hat{T}_U} = \frac{\{N\mathbf{1}_r\bar{\mathbf{X}} - \tilde{\mathbf{X}}_r\}^\top \{N\mathbf{1}_r\bar{\mathbf{X}} - \tilde{\mathbf{X}}_r\}}{\hat{p}^2(N-1)^2} \widehat{\text{Var}}(\hat{p}) \quad (24)$$

Otherwise when the denominator is not fixed, then the variance of the ratio might be obtained by the delta method or by simulation under a network model or re-sampling methods, as the parametric bootstrap method. In the following we ignore the variability of the denominator and apply naively Equation (24).

9. References

- Butts, C. 2008. "Network: A Package for Managing Relational Data in R." *Journal of Statistical Software* 24(2): 1–36. Doi: <http://dx.doi.org/10.18637/jss.v024.i06>.
- Carrington, P., J. Scott, and S. Wasserman. 2005. *Models and Methods in Social Network Analysis*. New York: Cambridge University Press.
- Chambers, R., H. Chandra, and N. Tzavidis. 2011. "On Bias-Robust Mean Squared Error Estimation for Pseudo-Linear Small Area Estimators." *Survey Methodology* 37: 153–170. Available at: <http://www5.statcan.gc.ca/olc-cel/olc.action?ObjId=12-001-X201100211604&ObjType=47&lang=en> (accessed September 2017).
- Chambers, R.L. and R.G. Clark. 2012. *An Introduction to Model-Based Survey Sampling with Applications*. Oxford: Oxford University Press.
- Clark, R.G. and R.L. Chambers. 2008. "Adaptive Calibration for Prediction of Finite Population Totals." *Survey Methodology* 34: 163–172. Available at: <http://www.statcan.gc.ca/pub/12-001-x/2008002/article/10757-eng.pdf> (accessed September 2017).
- Doreian, P., K. Teuter, and C.H. Wang. 1984. "Network Auto-Correlation Models – some Monte-Carlo Results." *Sociological Methods & Research* 13(2): 155–200. Doi: <https://doi.org/10.1177/0049124184013002001>.
- Duke, J.B. 1993. "Estimation of the Network Effects Model in a Large Data Set." *Sociological Methods & Research* 21(4): 465–481. Doi: <https://doi.org/10.1177/0049124193021004003>.
- Frank, O. and D. Strauss. 1986. "Markov Graphs." *Journal of the American Statistical Association* 81(395): 832–842. Doi: <http://dx.doi.org/10.1080/01621459.1986.10478342>.
- Friedkin, N.E. 1990. "Social Networks in Structural Equation Models." *Social Psychology Quarterly* 53(4): 316–328.

- Goldstein, H. 1989. "Restricted Unbiased Iterative Generalized Least-Squares Estimation." *Biometrika* 76(3): 622–623. Doi: <https://doi.org/10.1093/biomet/76.3.622>.
- Handcock, M.S., A.E. Raftery, and J.M. Tantrum. 2007. "Model-Based Clustering for Social Networks." *Journal of the Royal Statistical Society Series A* 170: 301–322. Doi: <http://dx.doi.org/10.1111/j.1467-985X.2007.00471.x>.
- Hunter, D., M. Handcock, C. Butts, S. Goodreau, and M. Morris. 2008b. "ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks." *Journal of Statistical Software* 24(2): 1–29. Doi: <http://dx.doi.org/10.18637/jss.v024.i03>.
- Hunter, D.R. 2007. "Curved Exponential Family Models for Social Networks." *Social Networks* 29(2): 216–230. Doi: <http://dx.doi.org/10.1198/106186006X133069>.
- Hunter, D.R., S.M. Goodreau, and M.S. Handcock. 2008a. "Goodness of Fit of Social Network Models." *Journal of the American Statistical Association* 103(481): 248–258. Doi: <http://dx.doi.org/10.1198/016214507000000446>.
- Hunter, D.R. and M.S. Handcock. 2006. "Inference in Curved Exponential Family Models for Networks." *Journal of Computational and Graphical Statistics* 15(3): 565–583. Doi: <http://dx.doi.org/10.1198/106186006X133069>.
- Koskinen, J., G. Robins, and P. Pattison. 2010. "Analysing Exponential Random Graph (p-star) Models with Missing Data Using Bayesian Data Augmentation." *Statistical Methodology* 7(3): 366–384. Doi: <https://doi.org/10.1016/j.stamet.2009.09.007>.
- Leenders, R. 2002. "Modeling Social Influence Through Network Autocorrelation: Constructing the Weight Matrix." *Social Networks* 24(1): 21–47. Doi: [https://doi.org/10.1016/S0378-8733\(0100049-1\)](https://doi.org/10.1016/S0378-8733(0100049-1)).
- Marsden, P.V. and N.E. Friedkin. 1993. "Network Studies of Social-Influence." *Sociological Methods & Research* 22(1): 127–151. Doi: <https://doi.org/10.1177/0049124193022001006>.
- Ord, K. 1975. "Estimation Methods for Models of Spatial Interaction." *Journal of the American Statistical Association* 70(349): 120–126. Doi: <http://amstat.tandfonline.com/> <https://doi.org/10.1080/01621459.1975.10480272>.
- Pattison, P., G. Robins, T. Snijders, and P. Wang. 2013. "Conditional Estimation of Exponential Random Graph Models from Snowball Sampling Designs." *Journal of Mathematical Psychology* 57(6): 284–296. Doi: <https://doi.org/10.1016/j.jmp.2013.05.004>.
- Royall, R.M. 1976. "Linear Least-Squares Prediction Approach to 2-Stage Sampling." *Journal of the American Statistical Association* 71(355): 657–664. Doi: <http://dx.doi.org/10.2307/2285596>.
- Särndal, C., B. Swensson, and J. Wretman. 1992. *Model Assisted Survey Sampling*. Springer series in statistics. New York: Springer-Verlag.
- Snijders, T. 2002. "Markov Chain Monte Carlo Estimation of Exponential Random Graph Models." *Journal of Social Structure* 1–40. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.20.5323&rep=rep1&type=pdf> (accessed September 2017).
- Snijders, T., P. Pattison, G. Robins, and M. Handcock. 2006. "New Specifications for Exponential Random Graph Models." *Sociological Methodology* 36: 99–153. Doi: <http://dx.doi.org/10.1111/j.1467-9531.2006.00176.x>.

- Strauss, D. and M. Ikeda. 1990. "Pseudolikelihood Estimation for Social Networks." *Journal of the American Statistical Association* 85(409): 204–212. Doi: <http://dx.doi.org/10.2307/2289546>.
- Suesse, T. 2012a. "Estimation in Autoregressive Population Models." In Proceedings of Fifth Annual ASEARC Research Conference, 11–14. Applied Statistics and Research Collaboration (ASEARC). 2–3 February 2012, Wollongong, Australia. Available at: <http://eis.uow.edu.au/asearc/5thAnnResCon/index.html> (accessed September 2017).
- Suesse, T. 2012b. "Marginalized Exponential Random Graph Models." *Journal of Computational and Graphical Statistics* 21(4): 883–900. Doi: <http://dx.doi.org/10.1080/10618600.2012.694750>.
- Suesse, T. and R. Chambers. 2012. *Using Social Network Information for Survey Estimation*. Report 11/12, National Institute of Applied Statistics Research Australia, University of Wollongong. Available at: <http://niasra.uow.edu.au/content/groups/public/@web/@inf/@math/documents/doc/uow137689.pdf> (accessed September 2017).
- Suesse, T. and R. Chambers. 2014. *Using Social Network Information for Survey Estimation*. Report 13/14, National Institute of Applied Statistics Research Australia, University of Wollongong. Available at: <http://niasra.uow.edu.au/content/groups/public/@web/@inf/@math/documents/mm/uow182447.pdf> (accessed September 2017).
- Suesse, T. and A. Zammit Mangion. 2017. "Computational Aspects of the em Algorithm for Spatial Econometric Models with Missing Data." *Journal of Statistical Computation and Simulation* 87: 1767–1786. Doi: <http://dx.doi.org/10.1080/00949655.2017.1286495>.
- Thompson, S. and G. Seber. 1996. *Adaptive Sampling*. Wiley Series in probability and mathematical statistics. New York: Wiley.
- Wasserman, S. and K. Faust. 1994. *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.

Received July 2015

Revised July 2017

Accepted September 2017