Quant Mark Econ (2018) 16:409–440
https://doi.org/10.1007/s11129-018-9199-z

# Sequential sampling enhanced composite likelihood approach to estimation of social intercorrelations in large-scale networks

**Yan Chen[1] · Youran Qi[1] · Qing Liu[2] · Peter Chien[1]**

Published online: 13 June 2018

**Abstract** The increasing access to large social network data has generated substantial interest in the marketing community. However, due to its large scale, traditional analysis methods often become inadequate. In this paper, we propose a sequential sampling enhanced composite likelihood approach for efficient estimation of social intercorrelations in large-scale networks using the spatial model. Given a known population network, the proposed approach sequentially takes small samples from the network, and adaptively improves model parameter estimates through learnings obtained from previous samples. In comparison to population-based maximum likelihood estimation that is computationally prohibitive when the network size is large, the proposed approach makes it computationally feasible to analyze large networks and provide efficient estimation of social intercorrelations among members in large networks. In comparison to sample-based estimation that relies on information purely from the sample and produces underestimation bias in social intercorrelation estimates, the proposed approach effectively uses information from the population without compromising computation efficiency. Through simulation studies based on simulated networks and real networks, we demonstrate significant advantages of the proposed approach over benchmark estimation methods and discuss managerial implications. We also discuss extension of the proposed approach in the context of an unknown population network structure, as well as in an alternative form of the spatial model.

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s11129-018-9199-z) contains supplementary material, which is available to authorized users.

✉ Qing Liu
qliu@bus.wisc.edu

[1] Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, USA

[2] Department of Marketing, University of Wisconsin-Madison, Madison, WI 53706, USA

**Keywords** Spatial models · Spatial autocorrelation · Social intercorrelation · Composite likelihood · Sequential sampling · Active learning · Large-scale networks · Big data

**JEL Classification** C13 · C31 · C83 · M30

## 1 Introduction

In the past decades the world has been transformed by the data explosion. The amount of data collected has increased tremendously with the new technology advancements and reduced cost in data storage. In particular, with an explosive growth of social media and social networking sites such as Facebook.com and Linkedin.com, large amounts of data become available on interactions and associations among individuals. The increasing access to social network data has generated substantial interest in the marketing community. Various modeling approaches have been proposed in the marketing literature to study the intercorrelations of consumer behavior among members in social networks (see reviews by Hartmann et al. 2008; Van Den Bulte and Wuyts 2007). Whereas the simultaneous equations econometric models focus on the identification of the causal social interactions by addressing issues such as endogenous group formation, correlated unobservables and simultaneity (e.g., Hartmann 2010; Nair et al. 2010; Nam et al. 2010), statistical models treat the networks as exogenous and provide a reduced-form representation of social interactions (e.g., Wang et al. 2013; Yang and Allenby 2003).

A common challenge faced by all aforementioned modeling approaches in the analysis of large-scale social networks is computation—when the network size gets large, the analysis of the network in its entirety becomes intractable and impractical. Although statistical sampling methods can be used to take a small “representative” sample from the entire network, recent research has found that estimates of social intercorrelations tend to be biased if samples of the networks rather than the entire population are used (Chen et al. 2013). In this paper, we aim to address these issues by proposing a new method for efficient estimation of social intercorrelations in large-scale social networks.

By definition, a network is a collection of nodes (i.e., network members) in which some pairs of these nodes are connected by edges and others are not. We focus on a type of statistical models that have been used for the analysis of network data—the simultaneous autoregressive (SAR) model (Cressie 1993; Anselin 1988). It is a class of spatial models that have been widely adopted in marketing (e.g., Aravindakshan et al. 2012; Bradlow et al. 2005; Yang and Allenby 2003; Bronnenberg and Sismeiro 2002; Bronnenberg and Mahajan 2001). Through the social intercorrelation parameter $\rho$, the spatial model effectively captures social interactions among network members. A key advantage of this model is that it can capture social interactions among member who are both directly connected or indirectly connected through common neighbors (Chen et al. 2013). Extant research has demonstrated that accurate learning of the social intercorrelation parameter $\rho$ is crucial to marketers in multiple ways, including the understanding and prediction of interdependent choice decisions across consumers (Yang and Allenby 2003), brand performance across markets (Bronnenberg and Sismeiro 2002), market responses to promotions (Bronnenberg and Mahajan 2001), and optimal advertising budget allocation (Aravindakshan et al. 2012). On the other hand, biased social intercorrelation estimates can easily result in economic harm to managers who are using the information for marketing decisions, as discussed in detail by Chen et al. (2013).

To achieve consistent parameter estimates for the spatial model, it has been advocated to use the likelihood-based approach such as the maximum likelihood estimation (Ord 1975). The likelihood function for the spatial model contains the determinant of the matrix $\mathbf{I} - \rho\mathbf{W}$, where $\mathbf{I}$ is the identity matrix, $\rho$ is the social intercorrelation parameter, and $\mathbf{W}$ is the spatial weight matrix of size $N \times N$ defined based on the connections among the $N$ network members. As a result, the maximum likelihood estimation involves the evaluation of the determinant of the matrix $\mathbf{I} - \rho\mathbf{W}$ for each new candidate value of $\rho$, which is of computational complexity $O(N^3)$. For large networks where the network size $N$ easily exceeds tens of thousands, this becomes computationally prohibitive. Bayesian estimation has the same issue because it is also based on the likelihood function.

Various solutions have been proposed to address this computation problem. One way is to reduce the computational complexity by assuming a sparse structure of the spatial weight matrix $\mathbf{W}$ where there are few connections among members of the network. For example, Pace and Zou (2000) assume that only the nearest neighboring node has a direct effect and provide a closed form solution for the maximum likelihood estimates of the spatial model. Pace and Barry (1997) assume few direct relationships among members in the network and provide algorithms for quick computation of the log determinant of $\mathbf{I} - \rho\mathbf{W}$ based on the Cholesky decomposition under the assumption. Smirnov and Anselin (2001) propose a four-step divide-and-conquer algorithm by using the characteristic polynomials, which can be truncated and thus reduce the computational cost to be of linear complexity under a very sparse structure of the spatial weight matrix $\mathbf{W}$. Similarly, Barry and Pace (1999) propose to approximate the logarithm of the determinant of $\mathbf{I} - \rho\mathbf{W}$ using the Monte Carlo simulation; Pace and LeSage (2004) propose the Chebyshev approximation; LeSage and Pace (2007) propose the matrix exponential spatial specification (MESS) that relies on the assumption that the matrix $\mathbf{I} - \rho\mathbf{W}$ can be replaced by an exponential function $e^{\alpha\mathbf{W}}$. While these methods significantly reduce the computational complexity, they all rely on the assumption of a sparse structure of the spatial weight matrix $\mathbf{W}$ which imposes high restrictions on the type and number of connections among network members that may not be realistic.

An alternative approach to addressing this computation problem is to approximate the likelihood function directly. This is represented by the pseudo-likelihood approach proposed by Besag (1975). The pseudo-likelihood is an inference function defined as the product of the conditional densities of the observation at one site given those at other sites. It belongs to the general class of composite likelihood (see Varin et al. 2011 for a review). The composite likelihood is an inference function formed by multiplying a collection of component likelihoods, where each individual component is either a marginal or a conditional density. Because the components are multiplied regardless of the possible dependence between components, the composite likelihood is the true likelihood only if different components are independent marginally or conditionally. In general, the composite likelihood serves as an approximation of the true likelihood.

Recent variants of Besag's proposal include the use of subsets or blocks of observations in the conditional densities. For example, in Vecchia (1988), each component is the conditional density of a single observation given only a subset of the rest observations chosen by the spatial proximity to improve computational efficiency. However, for the inference function proposed by Besag or Vecchia, the number of components in the composite likelihood is the same as the number of observations, which is still computationally challenging if the network size $N$ is large. In addition, when the interest is on the social intercorrelations among members in the network, it is not very informative to use the conditional density of only one observation. Stein et al. (2004) further develop Vecchia's proposal by first grouping observations into blocks and then using blocks of observations instead of a single observation in the conditional densities. However, when the network size gets large, difficulties arise in the systematic partition of observations into blocks and the selection of the subsets. Most recently, Zhou et al. (2017) propose a computationally efficient approach by using a paired maximum likelihood estimator (PMLE) and approximating the PMLE with a closed-form expression for a spatial lag model without independent variables $X$. The approach focuses only on directly connected pairs of network members under the assumption that the network is sparse with few members that are indirectly connected.

In this paper, we propose an alternative approach for efficient estimation of social intercorrelations in large-scale networks using the spatial error or the spatial lag model. We call this approach the sequential sampling enhanced composite likelihood approach. Given a known population network, the proposed approach works by sequentially taking small samples from the network, and gradually improving the model parameter estimates based on the composite likelihood. Each component of the composite likelihood is the conditional density of the nodes in the corresponding sample given all the remaining nodes in the population. The sample points in the next sample are selected based on the knowledge obtained from the previous sample. The interplay between the sequential sampling and the composite likelihood estimation makes it possible for the proposed approach to achieve high computational efficiency and estimation accuracy at the same time.

Through comparative studies based on both simulated networks and real networks, we find that the proposed procedure shows significant advantages in terms of both computing time and accuracy in parameter estimation in comparison to the benchmark estimation methods. While the benchmark estimation methods are computationally intensive or even infeasible when the network size is large, the proposed approach works well on large networks with high estimation efficiency. For example, in the comparative study based on a large network with 105,938 nodes and over 2 million edges, the proposed approach recovers the true social intercorrelation parameter, and the mean computation time taken is only thirty minutes or so. In contrast, the benchmark population-based maximum likelihood estimation is infeasible computationally, and the benchmark sample-based maximum likelihood estimation produces biased social intercorrelation estimate and takes more than two hours long.

The proposed approach makes the following contribution to the literature. Methodologically, it complements existing literature on composite likelihood through a sequential sampling and estimation procedure that adaptively improves model parameter estimates through learnings obtained from previous samples. In particular, the proposed approach sequentially takes small samples from the network where the next sample is determined by estimation results and corresponding residuals from the previous sample. At each stage of the sequential procedure, a set of parameter estimates is obtained by maximizing the composite likelihood function which is the product of the conditional densities corresponding to the series of small samples selected so far. The parameter estimates are then gradually improved until the sequential procedure converges or stops according to a pre-specified stopping rule.

Substantively, the proposed approach provides a valuable tool to practitioners and researchers alike in the efficient estimation of social intercorrelations in large-scale networks. In comparison to population-based maximum likelihood estimation that is computationally prohibitive when the network size is large, the proposed approach scales well to large networks while at the same time achieves high estimation accuracy. In comparison to sample-based estimation that relies on information purely from the sample and produces underestimation bias in social intercorrelation estimates, the proposed approach effectively uses information from the population without compromising computation efficiency. It has important managerial implications for marketing decisions, such as optimal allocation of market spending based on estimates of social intercorrelations in consumer networks (Chen et al. 2013; Aravindakshan et al. 2012).

The rest of the paper is organized as follows. We review the spatial error model of social interactions in Section 2. We propose the sequential sampling approach in Section 3 in the context of a known population network. Then in Section 4 we examine the performance of the proposed approach through an extensive simulation study and provide implementation guidelines based on simulation findings. We compare the proposed approach with benchmark estimation methods and demonstrate the marketing implication of the proposed approach in Section 5. We discuss extension of the proposed approach in Section 6 in the context of an unknown population network structure, as well as in an alternative form of the spatial model. We then conclude the paper in Section 7.

## 2 Spatial error model of social interaction

We start with a review of the spatial error model used for the analysis of network data. Consider a social network with $N$ members. For member $i$ in the network, $n_i$ is the number of edges adjacent to $i$. For the purpose of this research, we focus on undirected networks only where edges simply represent presence of connections and do not convey information on the directions. The $N \times N$ spatial weight matrix $\mathbf{W}$ for such a network is commonly defined to be the row-standardized adjacency matrix, that is,

$$\mathbf{W}(i,j) = \begin{cases} \frac{1}{n_i}, & \text{if } i \text{ and } j \text{ are connected,} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Let $\mathbf{X}$ denote the model matrix that contains $p$ independent variables, and $\mathbf{y}$ denote the vector of responses. The relationship between the response variable and the independent variables is captured through the following regression model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \tag{2}$$

where $\mathbf{y}$ and $\mathbf{e}$ are $N \times 1$ vectors, $\mathbf{X}$ is an $N \times p$ matrix, and $\boldsymbol{\beta}$ is a $p \times 1$ vector. The error term $e_i$ for member $i$ is allowed to be correlated with the error terms for other members in the network through the following autoregressive specification (e.g., Yang and Allenby 2003; Bronnenberg and Sismeiro 2002; Bronnenberg and Mahajan 2001),

$$\mathbf{e} = \rho\mathbf{W}\mathbf{e} + \boldsymbol{\epsilon}, \tag{3}$$

where $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \boldsymbol{I})$. Combining Eqs. 2 and 3, the distribution of $\mathbf{y}$ is essentially

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{I} - \rho\mathbf{W}')^{-1}). \tag{4}$$

With the spatial weight matrix $\mathbf{W}$ defined to be the row-standardized adjacency matrix in Eq. 1, the valid range of the parameter $\rho$ is such that $|\rho| < 1$ (Kelejian and Prucha 2010). As discussed in Hartmann et al. (2008), this model provides a reduced-form representation of the social interactions among members in the network through the autoregressive error structure in Eq. 3. The parameter $\rho$ captures the strength of social intercorrelations where higher magnitudes of $\rho$ indicate stronger social intercorrelations.

It is important to note that the effect of social interactions captured by this model is not limited to directly connected members (dyads). Instead, it takes into consideration of the overall network topology through the autoregressive structure in Eq. 3 such that the errors $e_i$ and $e_j$ can be correlated even if members $i$ and $j$ are not directly connected (Chen et al. 2013). This makes the spatial model distinctly different from models that focus on the study of social interactions in dyads (e.g., Hartmann 2010; Nair et al. 2010; Yang et al. 2006).

### 2.1 Maximum likelihood estimation

The maximum likelihood estimates of the model parameters $(\rho, \boldsymbol{\beta}, \sigma^2)$ in the spatial error model (2) can be obtained by maximizing the log likelihood function

$$l(\rho, \boldsymbol{\beta}, \sigma^2) = -\frac{N}{2}\log 2\pi\sigma^2 + \log(\det(\mathbf{I} - \rho\mathbf{W})) - \frac{1}{2\sigma^2}\|(\mathbf{I} - \rho\mathbf{W})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|_2^2, \tag{5}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix, $\det(\mathbf{I} - \rho\mathbf{W})$ is the determinant of $\mathbf{I} - \rho\mathbf{W}$, and $\|.\|_2^2$ denotes the $L_2$ norm such that

$$\|(\mathbf{I} - \rho\mathbf{W})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|_2^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{I} - \rho\mathbf{W}')(\mathbf{I} - \rho\mathbf{W})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Given $\rho$, the maximum likelihood estimates of $\boldsymbol{\beta}$ and $\sigma^2$ can be expressed in closed-forms as $\hat{\boldsymbol{\beta}}(\rho)$ and $\hat{\sigma}^2(\rho)$ such that

$$\hat{\boldsymbol{\beta}}(\rho) = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}}, \tag{6}$$

$$\hat{\sigma}^2(\rho) = \frac{1}{N}\|\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\hat{\boldsymbol{\beta}}(\rho)\|_2^2, \tag{7}$$

where $\tilde{\mathbf{X}} = (\mathbf{I} - \rho\mathbf{W})\mathbf{X}$ and $\tilde{\mathbf{y}} = (\mathbf{I} - \rho\mathbf{W})\mathbf{y}$. Substituting $\boldsymbol{\beta}$ and $\sigma^2$ in the log likelihood function (5) with $\hat{\boldsymbol{\beta}}(\rho)$ and $\hat{\sigma}^2(\rho)$, the maximum likelihood estimate of $\rho$ can be obtained (Ord 1975) by maximizing

$$l(\rho, \hat{\boldsymbol{\beta}}(\rho), \hat{\sigma}^2(\rho)) = -\frac{N}{2}\log 2\pi\hat{\sigma}^2(\rho) + \log(\det(\mathbf{I} - \rho\mathbf{W})) - \frac{N}{2}. \tag{8}$$

This function does not have a closed-form solution. Therefore, the maximum likelihood estimate of $\rho$ cannot be expressed in a closed form, but it can be obtained through numerical methods such as grid search. Because the evaluation of the determinant of matrix $\mathbf{I} - \rho\mathbf{W}$ is of computational complexity $O(N^3)$, this becomes computationally prohibitive for networks with large network size $N$. In addition to the large amount of time and memory required, the computation is often numerically unstable. Although various methods have been proposed to reduce the computational complexity (e.g., Pace and Barry 1997; Barry and Pace 1999; Smirnov and Anselin 2001; Pace and Zou 2000; Pace and LeSage 2004; LeSage and Pace 2007), they rely on the assumption of a sparse spatial weight matrix $\mathbf{W}$. Such assumption imposes high restrictions on the connections among network members which may not be realistic.

### 2.2 Bias in sample-based estimation

When the network size is large, researchers often resort to sampling to make the analysis feasible. The basic premise is that if a “representative” sample is taken, then working with the smaller sample network would provide insights about how the entire population network would behave. While random sampling has been a popular sampling method in marketing research, it does not work well in network settings because a random sample of the nodes from the population does not preserve the topology of the network. Therefore, researchers in sociology and marketing have used alternative sampling methods such as snowball sampling and forest fire sampling to obtain a sample from the network (e.g., Ebbes et al. 2016; Henry 2005; Salganik and Heckathorn 2004; Tepper 1994; Frenzen and Davis 1990). Originally proposed by Goodman (1961), the snowball sampling procedure starts from a randomly chosen node (called the snowball seeding node), and then sample all nodes connected to the snowball seeding node. In the next stage, the procedure considers all nodes just added in the previous stage, and sample all their connections, with the duplicated nodes excluded. The sampling procedure continues until a desired sample size is reached. Forest fire sampling method can be considered as a more general case of snowball sampling in the sense that it starts by selecting one (or more) nodes at random as the seeding nodes, and then randomly include a certain percentage of unselected neighbors of the seeding nodes and all edges among them. Regard the included neighbors as the new seeding nodes and repeat this process until a desired sample size is reached.

To estimate the parameters in the spatial error model (2) using a sample $\mathbf{s}$, let $(\mathbf{X_s}, \mathbf{y_s})$ denote the independent and response variables from the $|\mathbf{s}| = n$ sampled nodes of the network. The maximum likelihood estimate based on the sample is obtained by maximizing the likelihood

$$l(\rho, \boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log 2\pi\sigma^2 + \log(\det(\mathbf{I} - \rho\mathbf{W_s})) - \frac{1}{2\sigma^2}\|(\mathbf{I} - \rho\mathbf{W_s})(\mathbf{y_s} - \mathbf{X_s}\boldsymbol{\beta})\|_2^2, \quad (9)$$

where $\mathbf{W_s}$ is an $n$ by $n$ matrix defined by

$$\mathbf{W_s}(i, j) = \begin{cases} \frac{1}{n_{s_i}}, & \text{if } i \text{ and } j \text{ are connected in the sample,} \\ 0, & \text{otherwise} \end{cases}$$

and $n_{s_i}$ is the number of connections that node $i$ has in the sample.

Note that the information on the network structure as reflected through $\mathbf{W_s}$ is only based on the sample where some members' connections in the network are excluded from the sample, such as those members that are randomly excluded in forest fire samples, or those that are on the last stage of snowball samples. The network structure based on the sample thus would deviate from the true structure based on the population. As demonstrated in detail by Chen et al. (2013), this results in estimation bias of the social intercorrelation parameter $\rho$ in the spatial model. Although sampling methods that better preserve the network structure (e.g., snowball sampling) perform better than other sampling methods (e.g., random sampling), they all lead to underestimated social intercorrelations, especially in networks where the number of connections of network members are characterized by the scale-free power-law distribution, commonly observed in many large networks such as the network of people connected by e-mail (Barabási and Albert 1999).

## 3 The proposed SEQ-MCLE approach

We propose a sequential sampling enhanced composite likelihood approach that makes it computationally feasible to analyze large networks using the spatial model and obtain efficient estimates of social intercorrelations. We call this the SEQ-MCLE approach for simplification. Next we briefly review the definition of composite likelihood and then introduce our proposed SEQ-MCLE approach.

### 3.1 Definition of composite likelihood

The composite likelihood is an inference function defined as a product of a collection of component likelihoods, where each individual component is either a marginal or a conditional density (see Varin et al. 2011 for a review). It is used to approximate the true likelihood function in the case of computational complexity. In general, consider an $N$-dimensional random vector $\mathbf{y}$ with probability density function (p.d.f)

$f_{\boldsymbol{\theta}}(\mathbf{y})$, where $\boldsymbol{\theta} \in \Theta$ is the unknown parameter. Let $\{A_1, A_2, \ldots, A_B\}$ denote a set of marginal or conditional events with associated likelihoods $L_i(\boldsymbol{\theta}; \boldsymbol{y}) \propto f_{\boldsymbol{\theta}}(\boldsymbol{y} \in A_i)$, for $i = 1, 2, \ldots, B$. A composite likelihood $Ł_c(\boldsymbol{\theta}; \mathbf{y})$ is defined as follows

$$Ł_c(\boldsymbol{\theta}; \mathbf{y}) = \prod_{i=1}^{B} L_i(\boldsymbol{\theta}; \boldsymbol{y}). \tag{10}$$

It has been shown that the parameter estimate obtained by maximizing the composite likelihood function (MCLE) is consistent and asymptotically normal under standard regularity conditions (Lindsay 1988; Varin et al. 2011). The mean of the asymptotic normal distribution is the true parameter vector, and the variance-covariance matrix is the inverse of the Godambe information matrix (Godambe 1960). In comparison to the maximum likelihood estimate (MLE) from the true or full likelihood function, MCLE is less efficient but the efficiency loss is generally small (Xu and Reid 2011; Varin et al. 2011). On the other hand, MCLE is more robust than MLE to model misspecification and missing data (Varin et al. 2011). MCLE also has the computational robustness (Renard et al. 2004) in the sense that the composite likelihood surface is much smoother and easier to maximize than the full likelihood (Liang and Yu 2003).

The pseudo-likelihood function proposed by Besag (1975) for the analysis of spatial models is one of the first examples of composite likelihood where each individual component is the conditional density of a single observation given the rest. Note that the total number of components in the pseudo-likelihood function equals the network size $N$. Therefore it is still computationally challenging to make inferences based on the pseudo-likelihood function if the network size $N$ is large. Moreover, there is information loss on the social intercorrelation among members in the network because each conditional density in the pseudo-likelihood is only on one single node which ignores the dependence between the node and its connections in the network.

### 3.2 The proposed sequential sampling enhanced composite likelihood

Given a known population network, the proposed SEQ-MCLE approach works by sequentially taking small samples from the network where at each stage of the sequential procedure, parameter estimates are obtained and a new sample is determined by the estimation results and corresponding residuals from the previous sample. The parameter estimates are obtained by maximizing the composite likelihood function which is the product of individual components corresponding to the series of small samples selelcted so far. The estimates are gradually improved until the sequential procedure converges or stops according to a pre-specified stopping rule.

Each component in the composite likelihood function is defined as the conditional density of one sample given all the remaining nodes in the population. Specifically, let $\mathbf{s}$ denote the index of a sample drawn from the total of $N$ network members, that is, $\mathbf{s}$ is a subset of $\{1, \ldots, N\}$ and let $\mathbf{s}^c$ denote the complement of $\mathbf{s}$. Let $\mathbf{Q} = (\mathbf{I}-\rho\mathbf{W})'(\mathbf{I}-\rho\mathbf{W})$. Then for the spatial error model (2), the conditional distribution of the observations in the sample $\mathbf{y}_{\mathbf{s}}$ given the rest of the observations in the population $\mathbf{y}_{\mathbf{s}^c}$ is

$$\mathbf{y}_{\mathbf{s}}|\mathbf{y}_{\mathbf{s}^c} \sim N(\mathbf{X}_{\mathbf{s}}\boldsymbol{\beta} - \mathbf{Q}_{\mathbf{ss}}^{-1}\mathbf{Q}_{\mathbf{ss}^c}(\mathbf{y}_{\mathbf{s}^c} - \mathbf{X}_{\mathbf{s}^c}\boldsymbol{\beta}), \sigma^2\mathbf{Q}_{\mathbf{ss}}^{-1}), \tag{11}$$

where $\mathbf{Q}_{\mathbf{ss}}$ is a submatrix of $\mathbf{Q}$ with row and column indices from $\mathbf{s}$ and $\mathbf{Q}_{\mathbf{ss}^c}$ is the submatrix of $\mathbf{Q}$ with row indices from $\mathbf{s}$ and column indices from the complement set $\mathbf{s}^c$. Note that the calculation of matrix inverse and determinant in the likelihood function of Eq. 11 only occurs on $\mathbf{Q}_{\mathbf{ss}}$ which can be quickly computed given that the size of the sample $\mathbf{s}$ is small. In addition, the evaluation of the matrix-vector multiplication $\mathbf{Q}_{\mathbf{ss}^c}(\mathbf{y}_{\mathbf{s}^c} - \mathbf{x}_{\mathbf{s}^c}\boldsymbol{\beta})$ in the mean of the conditional distribution (11) only depends on a subset of $\mathbf{s}^c$ such that the corresponding elements in $\mathbf{Q}_{\mathbf{ss}^c}$ are nonzero. That is, it only depends on the subset of nodes in $\mathbf{s}^c$ that are either directly connected to the nodes in the sample $\mathbf{s}$ or indirectly connected through common neighbors. Hence, given that the size of the sample $\mathbf{s}$ is small, the evaluation of the conditional density is computationally fast even when the population size $N$ is large.

Each sample in our proposed approach contains $k$ seeding nodes and at most $n$ of their randomly chosen neighbors, where $k$ and $n$ are pre-specified small numbers. We discuss in more detail the choices of $k$ and $n$ in Section 4. The initial seeding nodes $\mathbf{j}_1$ can be chosen randomly. The nodes in the next sample are selected based on the knowledge obtained from the previous sample. Specifically, let $\mathbf{s}_1$ be the first sample with the seeding nodes $\mathbf{j}_1$, and let $\hat{\boldsymbol{\theta}}^{(1)}$ denote the parameter estimates obtained by maximizing the conditional likelihood using the first sample, that is, $L(\boldsymbol{\theta}, \mathbf{y}_{\mathbf{s}_1}|\mathbf{y}_{\mathbf{s}_1^c})$, where $L(\boldsymbol{\theta}, \mathbf{y}_{\mathbf{s}_1}|\mathbf{y}_{\mathbf{s}_1^c}) = f_{\boldsymbol{\theta}}(\mathbf{y}_{\mathbf{s}}|\mathbf{y}_{\mathbf{s}^c})$ which is the p.d.f. of the conditional Normal distrubution in Eq. 11 for $\mathbf{s} = \mathbf{s}_1$ and $\boldsymbol{\theta} = (\rho, \boldsymbol{\beta}, \sigma^2)$. Note that according to Eq. 11, the variance and the mean of the conditional distribution are $Var(\mathbf{y}_{\mathbf{s}_1}|\mathbf{y}_{\mathbf{s}_1^c}, \boldsymbol{\theta}) = \sigma^2\mathbf{Q}_{\mathbf{s}_1\mathbf{s}_1}^{-1}$, and $E(\mathbf{y}_{\mathbf{s}_1}|\mathbf{y}_{\mathbf{s}_1^c}, \boldsymbol{\theta}) = \mathbf{X}_{\mathbf{s}_1}\boldsymbol{\beta} - \mathbf{Q}_{\mathbf{s}_1\mathbf{s}_1}^{-1}\mathbf{Q}_{\mathbf{s}_1\mathbf{s}_1^c}(\mathbf{y}_{\mathbf{s}_1^c} - \mathbf{X}_{\mathbf{s}_1^c}\boldsymbol{\beta})$. If the estimates obtained are close to the true values of parameters, then based on the normal distribution (11), the standardized residuals $\mathbf{r}_{\mathbf{s}_1}$ approximately follow a standard multivariate normal distribution, where

$$\mathbf{r}_{\mathbf{s}_1} = Var[\mathbf{y}_{\mathbf{s}_1}|\mathbf{y}_{\mathbf{s}_1^c}, \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(1)}]^{-1/2}(\mathbf{y}_{\mathbf{s}_1} - E[\mathbf{y}_{\mathbf{s}_1}|\mathbf{y}_{\mathbf{s}_1^c}, \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(1)}]). \tag{12}$$

Let $\mathbf{j}_2$ contain the top $k$ nodes in the sample $\mathbf{s}_1$ that have the largest absolute standardized residuals $|\mathbf{r}_{\mathbf{s}_1}|$, excluding the initial seeding nodes $\mathbf{j}_1$. We then select $\mathbf{j}_2$ as the seeding nodes and form the next sample $\mathbf{s}_2$ by taking at most $n$ neighbors of $\mathbf{j}_2$ that have not been included in the previous sample. The rationale for this selection is that the nodes with the largest absolute standardized residuals have the highest fitting errors and require further examination. Therefore, we take these nodes as the seeding nodes for the next sample so that more information can be obtained on these nodes. This selection criterion is similar in nature to active learning in the machine learning literature where new training sample points are selected at places where predictions from the current model have the highest uncertainty (Cohn et al. 1996; Schein and Ungar 2007; Settles 2010). While the focus is on a classification problem and the uncertainty of prediction is measured by the misclassification rate such as $1 - \widehat{Prob}(y = 1|x)$ (Settles 2010) in active learning, we focus on a regression problem here and use the absolute standardized residual to measure the uncertainty of prediction.

Next we update the parameter estimates to $\hat{\boldsymbol{\theta}}^{(2)}$ by maximizing the updated composite likelihood $L(\boldsymbol{\theta}, \mathbf{y}_{\mathbf{s}_1}|\mathbf{y}_{\mathbf{s}_1^c})L(\boldsymbol{\theta}, \mathbf{y}_{\mathbf{s}_2}|\mathbf{y}_{\mathbf{s}_2^c}) = \prod_{i=1}^{2} f_{\boldsymbol{\theta}}(\mathbf{y}_{\mathbf{s}_i}|\mathbf{y}_{\mathbf{s}_i^c})$. Similarly, we find the $k$ nodes in $\mathbf{s}_2$ based on the absolute standardized residuals $|\mathbf{r}_{\mathbf{s}_2}|$ and use them as the seeding nodes $\mathbf{j}_3$ for the next sample $\mathbf{s}_3$. To avoid duplicated seeding nodes, the previous seeding nodes are excluded in the selection of the seeding nodes for the next sample. The sequential procedure continues until the parameter estimates converge, or the number of samples reaches a pre-specified number $B$. Formally, the flow of the sequential procedure is summarized as follows.

For sample $t = 1, 2, \ldots, B$, do the following steps:

- Use $\mathbf{j}_t$ as the seeding nodes and sample at most $n$ of their neighbors that have not been included in previous samples $\{\mathbf{s}_1, \ldots, \mathbf{s}_{t-1}\}$.
- Update the parameter estimates by maximizing the composite likelihood, that is,

$$\hat{\boldsymbol{\theta}}^{(t)} = \arg\max_{\boldsymbol{\theta}} \prod_{i=1}^{t} L(\boldsymbol{\theta}, \mathbf{y}_{\mathbf{s}_i}|\mathbf{y}_{\mathbf{s}_i^c}). \tag{13}$$

  Specifically, for $\boldsymbol{\theta} = (\rho, \boldsymbol{\beta}, \sigma^2)$ in the spatial error model, first obtain the MCLE estimate $\hat{\rho}^{(t)}$ of the social intercorrelation parameter $\rho$ by maximizing the composite likelihood function $\prod_{i=1}^{t} L(\rho, \boldsymbol{\beta}_c(\rho), \sigma_c^2(\rho), \mathbf{y}_{\mathbf{s}_i}|\mathbf{y}_{\mathbf{s}_i^c})$ through numerical methods such as grid search, where $\boldsymbol{\beta}_c(\rho)$ and $\sigma_c^2(\rho)$ are the following closed-form expressions of the estimates to $\boldsymbol{\beta}$ and $\sigma^2$ which maximize the composite likelihood function under given $\rho$.

$$\boldsymbol{\beta}_c(\rho) = \left(\sum_{i=1}^{t} \tilde{\mathbf{X}}_{\mathbf{s}_i}' \mathbf{Q}_{\mathbf{s}_i\mathbf{s}_i} \tilde{\mathbf{X}}_{\mathbf{s}_i}\right)^{-1} \left(\sum_{i=1}^{t} \tilde{\mathbf{X}}_{\mathbf{s}_i}' \mathbf{Q}_{\mathbf{s}_i\mathbf{s}_i} \tilde{\mathbf{y}}_{\mathbf{s}_i}\right), \tag{14}$$

$$\sigma_c^2(\rho) = \frac{1}{\sum_{i=1}^{t}|\mathbf{s}_i|} \sum_{i=1}^{t} (\tilde{\mathbf{y}}_{\mathbf{s}_i} - \tilde{\mathbf{X}}_{\mathbf{s}_i}\boldsymbol{\beta}_c(\rho))' \mathbf{Q}_{\mathbf{s}_i\mathbf{s}_i} (\tilde{\mathbf{y}}_{\mathbf{s}_i} - \tilde{\mathbf{X}}_{\mathbf{s}_i}\boldsymbol{\beta}_c(\rho)), \tag{15}$$

  $|\mathbf{s}_i|$ is the size of sample $\mathbf{s}_i$, $\tilde{\mathbf{y}}_{\mathbf{s}_i} = \mathbf{y}_{\mathbf{s}_i} + \mathbf{Q}_{\mathbf{s}_i\mathbf{s}_i}^{-1} \mathbf{Q}_{\mathbf{s}_i\mathbf{s}_i^c} \mathbf{y}_{\mathbf{s}_i^c}$, and $\tilde{\mathbf{X}}_{\mathbf{s}_i} = \mathbf{X}_{\mathbf{s}_i} + \mathbf{Q}_{\mathbf{s}_i\mathbf{s}_i}^{-1} \mathbf{Q}_{\mathbf{s}_i\mathbf{s}_i^c} \mathbf{X}_{\mathbf{s}_i^c}$. After $\hat{\rho}^{(t)}$ is obtained, the MCLE estimates to $\boldsymbol{\beta}$ and $\sigma^2$ are then given by $\boldsymbol{\beta}_c(\hat{\rho}^{(t)})$ and $\sigma_c^2(\hat{\rho}^{(t)})$ using the above expressions with $\rho = \hat{\rho}^{(t)}$.
- If $\hat{\boldsymbol{\theta}}^{(t)}$ has reached convergence based on the definition such that for a pre-specified $b$ (e.g. $b = 20$), the absolute difference between the maximum and the minimum estimates of the last $b$ consecutive estimates does not exceed 0.01, then stop the procedure. Otherwise select the seeding nodes $\mathbf{j}_{t+1}$ for the next sample as the top $k$ nodes in the sample $\mathbf{s}_t$ that have the largest absolute standardized residuals $|\mathbf{r}_{\mathbf{s}_t}|$, where $\mathbf{r}_{\mathbf{s}_t} = Var[\mathbf{y}_{\mathbf{s}_t}|\mathbf{y}_{\mathbf{s}_t^c}, \boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}]^{-1/2}(\mathbf{y}_{\mathbf{s}_t} - E[\mathbf{y}_{\mathbf{s}_t}|\mathbf{y}_{\mathbf{s}_t^c}, \boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}])$. Avoid duplication with previous seeding nodes.

### 3.3 Convergence of the proposed SEQ-MCLE procedure

The estimates of the parameters obtained by the proposed SEQ-MCLE are asymptotically consistent (details of the proof are provided in Appendix C). Thus, as the number of sequential samples increases in the proposed approach, the resulting estimates $\hat{\boldsymbol{\theta}}$ converge to the true parameters $\boldsymbol{\theta}$ in probability.

We end the section by noting that the active learning nature of the sequential sampling contributes to estimation accuracy by adaptively selecting the next sample and improving model parameter estimates based on estimation results and corresponding residuals from the previous sample. On the other hand, the use of the composite likelihood estimation at the same time enhances computational efficiency. With the small size of each sequential sample, it is computationally fast in the proposed approach to evaluate each component of the composite likelihood function and update the MCLE estimates each time a new sample is selected until convergence. Thus, the interplay between the sequential sampling and the composite likelihood estimation makes it possible for the proposed SEQ-MCLE approach to obtain accurate parameter estimates with high computational efficiency. In comparison to the population-based maximum likelihood estimation that becomes computationally prohibitive when the network size gets large, the proposed SEQ-MCLE approach makes it computationally feasible to analyze large networks and provide accurate estimation of social intercorrelations. In comparison to the sample-based estimation that relies on information purely from the sample and produces underestimation bias in social intercorrelation, the proposed approach effectively uses information from the population without compromising computation efficiency.

## 4 Simulation study

In this section, we investigate through an extensive simulation study the performance of the proposed SEQ-MCLE approach over different network structures with different magnitudes of social intercorrelations. Based on the findings, we provide guidelines on the choices of parameters for the proposed approach, including the number of seeding nodes $k$ and the maximum number of neighbors $n$ for the sequential samples, as well as the maximum number of sequential samples $B$ for the stopping rule.

### 4.1 Simulation design

We begin with simulated networks encompassing three types of network topology that have been studied extensively in the literature: the power-law networks (Barabási and Albert 1999), the small-world networks or the “WS” networks (Watts and Strogatz 1998), and the power-cluster networks (Holme and Kim 2002). For each of the three network types, we generate $2 \times 2 = 4$ representative networks with two network sizes ($N = 10000, 50000$) and two different settings of parameters that give rise to different network characteristics. Hence we have $3 \times 2 \times 2 = 12$

simulated networks, where the details of the generation algorithms are provided in the supplemental online Appendix.

#### *4.1.1 Power-law network*

Many real large networks exhibit the property of power-law networks where the number of connections of network members follows a scale-free power-law distribution. These include the web pages connected by the hyperlinks in the World Wide Web, the network of people connected by e-mail, and the network of scientific papers connected by citations (Barabási and Bonabeau 2003; Katona and Sarvary 2007). In particular, the probability that any node is connected to $d$ other nodes in a power-law network is proportional to $d^{-\gamma}$ with $\gamma > 0$. In practice, $\gamma$ is found to be usually between 2 and 3 (Saramaki and Kaski 2004). As a result, most nodes have just a few connections with the exception of some nodes that have a tremendous number of connections. In that sense, such a network with the power-law degree distribution has no "scale", and thus has the property of being scale-free.

We generate the power-law networks following the models in Barabási and Albert (1999, "BA") and Krapivsky and Redner (2001, "KR"). Whereas the BA model leads to power-law networks with $\gamma \approx 3$ in the degree distribution, the setting in the KR model gives us power-law networks with $\gamma \approx 2.1708$. We use "PL1" and "PL1L" to denote the two power-law networks generated by the BA model with network sizes $N = 10000$ and $N = 50000$, and use "PL2" and "PL2L" to denote the power-law networks generated by the KR model.

#### *4.1.2 WS network*

Introduced by Watts and Strogatz (1998), a WS network exhibits a small-world property where members in the network form subclusters but at the same time any two members of the network can reach each other within a small number of edges regardless of the size of the network. It is usually generated from a regular network by rewinding the connections randomly with a given probability. We generate the WS networks by following the algorithm in Watts and Strogatz (1998). We use "WS1" and "WS1L" to denote the two WS networks respectively of sizes $N = 10000$ and $N = 50000$ that are generated with the low rewinding probability, and use "WS2" and "WS2L" to denote the WS networks generated with the high rewinding probability.

#### *4.1.3 Power-cluster network*

The power-cluster network is another type of network topology that mimics real networks. Specifically, it has the power-law degree distribution with clustering at the same time. We generate the power-cluster networks by following the steps proposed by Holme and Kim (2002). We use "PC1" and "PC1L" to denote the two power-cluster networks respectively of sizes $N = 10000$ and $N = 50000$ that are generated with the low clustering coefficient, and use "PC2" and "PC2L" to denote the power-cluster networks generated with the high clustering coefficient.

### *4.1.4 Scientific collaboration network ("Colla")*

In addition to the simulated networks, we also examine two real networks available through Stanford Large Network Dataset Collection. The first real network is the Condense Matter Physics collaboration network from the e-print arXiv. It covers scientific collaborations between authors for papers submitted to the Condense Matter Physics category. If an author $i$ co-authored a paper with author $j$, the network contains an edge between $i$ and $j$. The network covers papers in the period from January 1993 to April 2003 (124 months). It contains a total of 23,133 authors with 93,497 connections.

### *4.1.5 Flickr image-sharing network ("Flickr")*

The second real network that we examine is the Flickr image-sharing network. It is an online network of images from Flickr.com, the photo-sharing website. Links are formed between images from the same location, submitted to the same gallery, group, or set, with common tags, etc. The network contains 105,938 nodes and 2,316,948 edges.

Table 1 summarizes the characteristics of the twelve simulated networks as well as the two real networks. Altogether they provide a good representation of networks with varying network structures and characteristics. For example, among these networks the clustering coefficient varies from 0 to 0.6969, the characteristic path length varies from 3.6462 to 9.0023, and the average degree varies from 2.0002 to 43.7416. The definitions of these network characteristics are described below.

**Table 1** Characteristics of representative simulated and real networks

| Type | Label | N | L | CC | Degree | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Min | Med. | Avg. | Max |
| PL | PL1 | 10,000 | 3.6462 | 0.0074 | 5 | 7 | 9.9970 | 413 |
| | PL1L | 50,000 | 4.0930 | 0.0020 | 5 | 7 | 9.9994 | 964 |
| | PL2 | 10,000 | 3.6884 | 0.0000 | 1 | 1 | 2.0010 | 2262 |
| | PL2L | 50,000 | 3.9888 | 0.0000 | 1 | 1 | 2.0002 | 9544 |
| WS | WS1 | 10,000 | 7.3065 | 0.6969 | 20 | 20 | 20.2030 | 23 |
| | WS1L | 50,000 | 9.0023 | 0.6969 | 20 | 20 | 20.2027 | 24 |
| | WS2 | 10,000 | 4.4102 | 0.5994 | 20 | 22 | 21.7990 | 27 |
| | WS2L | 50,000 | 5.2044 | 0.5992 | 20 | 22 | 21.7998 | 28 |
| PC | PC1 | 10,000 | 3.9011 | 0.0688 | 3 | 4 | 6.0030 | 576 |
| | PC1L | 50,000 | 4.3694 | 0.0598 | 3 | 4 | 6.0006 | 1449 |
| | PC2 | 10,000 | 4.8049 | 0.6080 | 3 | 4 | 6.0030 | 361 |
| | PC2L | 50,000 | 5.4579 | 0.6072 | 3 | 4 | 6.0006 | 1071 |
| Real | Colla | 23,133 | 5.3521 | 0.6331 | 1 | 5 | 8.0834 | 281 |
| | Flickr | 105,938 | 4.3346 | 0.0891 | 1 | 7 | 43.7416 | 5425 |

- The characteristic path length L: This is the number of links on the shortest path between two members, averaged across all pairs of members of the network (Watts and Strogatz 1998). A smaller L suggests that two randomly selected members of the network are more likely to be directly connected.
- The clustering coefficient CC: This is defined as follows (Watts and Strogatz 1998). Suppose a member $i$ has $n_i$ linked neighbors. Then at most $n_i(n_i - 1)/2$ edges can exist between them. Define $C_i$ as the fraction of the $n_i(n_i - 1)/2$ links that actually exist for network member $i$, and the clustering coefficient CC is defined as the average of $C_i$ across all members in the network. A high CC indicates that the corresponding network tends to consist of clusters, where members in each cluster are highly interconnected.
- Degree Summary Statistics: Degree for a network member is defined as the number of connections that the member has in the network. We include in the table the median and the average, as well as the minimum and maximum degrees.

Given the total of 14 networks with varying network characteristics, we generated data for each network according to the spatial error model (2), with true parameter values $\boldsymbol{\beta} = (\beta_0, \beta_1) = (1, 1)'$, $\sigma = 2$, and covariates $\mathbf{x}_i = (1, r_i)$, where $r_i \overset{iid}{\sim} N(0, 1)$ for $i = 1, \ldots, N$. For the magnitude of the social intercorrelation parameter $\rho$ which is the focus of our study, we consider 5 possible scenarios: $\{0.1, 0.3, 0.5, 0.7, 0.9\}$, ranging from low to high, for each network. Hence, in total we generated $14 \times 5 = 70$ datasets.

For each generated dataset, we then investigate the performance of our proposed approach. Note that with $k$ seeding nodes and $n$ of their neighbors in each sequential sample of the proposed approach, the evaluation of the conditional likelihood function for each sample is of computation complexity $O((n + k)^3)$. This implies that the choices of $n$ and $k$ should be small for computational feasibility. In the simulation studies we focus on two choices of $k$ ($k = 1$ or $k = 2$), and three choices of $n$ ($n = 10$, $n = 30$ or $n = 50$). To assess the convergence of the proposed approach, we fit each dataset using the proposed sequential procedure with different choices of $k$ and $n$, and repeat the procedure 30 times, each time with different seeding nodes for the initial sample. Thus, we conducted a total of $70 \times 2 \times 3 \times 30 = 12,600$[1] estimations for the simulation study.

For purposes of illustration and comparison, we let the proposed sequential procedure run for the maximum $B = 200$ samples for all estimations, even though the procedure may converge well before it reaches 200 sequential samples. The mean estimates of the social intercorrelation parameter $\hat{\rho}^{(t)}$ are calculated over 30 replications of the proposed procedure. All time-series plots of the mean estimates

[1] The $4 \times 5 \times 30 = 600$ simulation results of the WS networks under the setting of $k = 1$ and $n = 50$ turn out to be exactly the same as those under $k = 1$ and $n = 30$. This is because the maximum number of degree in the WS networks does not exceed 30. Thus, the effective number of estimations for our subsequent analysis are 12,000.

$\hat{\rho}^{(t)}$ for different networks and parameter settings are provided in the supplemental online Appendix.

## 4.2 Simulation findings

In this section, we summarize the findings and provide implementation guidelines based on regression analysis of the simulation results in Section 4.1. We start by examining the absolute error $|\hat{\rho} - \rho|$, where $\rho$ is the true social intercorrelation and $\hat{\rho}$ is the estimate from the proposed approach. We find that the mean of all absolute errors from the simulation study is 0.022 and 80% does not exceed 0.03, demonstrating good overall performance of the proposed approach in recovering true social intercorrelations.

To understand how the performance of the proposed approach varies with the different networks and parameter settings, we use the absolute error as the response variable and fit a regression model using the following predictor variables: the type of the network ("PL", "WS", "PC", "Real"), network characteristics N, L, CC, and Average Degree (AvgDeg) as shown in Table 1, settings of $k$ ("$k = 1$", "$k = 2$") and $n$ ("$n = 10$", "$n = 30$", "$n = 50$") used for the proposed approach, and the magnitude of the social intercorrelation $\rho$ in the simulated data. For the three discrete factor variables, the levels "PL", "$k = 1$" and "$n = 10$" are chosen as the reference levels respectively.

Since high correlations exist among the predictor variables (e.g., WS networks are associated with high Clustering Coefficients), we use the LASSO regression (Tibshirani 1996) which works better than the standard regression method in discovering truly significant variables in the presence of highly correlated predictors. Note that in LASSO regression, the insignificant effects are shrunk to zero.

The results from the LASSO regression using the absolute error as the response variable, as summarized in Table 2 on the left, reveal several interesting findings.

**Table 2** Results from the LASSO regression analysis

| Coefficients | Estimates in absolute error model | Estimates in convergence rate model |
|---|---|---|
| (Intercept) | $1.59 \times 10^{-2}$ | 58.56 |
| WS | $1.14 \times 10^{-2}$ | 27.72 |
| PC | 0 | 0 |
| Real | 0 | 0 |
| $k = 2$ | 0 | 0 |
| $n = 30$ | $-1.80 \times 10^{-3}$ | 0 |
| $n = 50$ | $-3.00 \times 10^{-3}$ | −2.00 |
| N | 0 | 0 |
| L | $4.45 \times 10^{-4}$ | 0 |
| CC | $8.11 \times 10^{-3}$ | 16.29 |
| AvgDeg | $1.43 \times 10^{-4}$ | 0.62 |
| $\rho$ | $-4.48 \times 10^{-3}$ | 0 |

First, there is no significant difference between the choices of $k = 1$ and $k = 2$ seeding nodes for the proposed approach, while the choice of $n = 50$ yields the smallest estimation errors in comparison to the other two choices ($n = 10$ and $n = 30$). Second, the network size $N$ has no significant influence, while networks with smaller L, CC and AvgDeg resulted in smaller estimation errors from the proposed approach. Third, the performance of the proposed approach is slightly worse (i.e., with slightly higher estimation errors) on WS networks than on other types of networks. Fourth, the proposed approach performs better in data with higher social intercorrleation $\rho$.

Next, we focus on the convergence rate. The convergence rate is defined as the first time that the estimates of $\rho$ become stable in the last 20 consecutive iterations, that is, when the absolute difference between the maximum and minimum estimates from the last 20 iterations does not exceed 0.01. If the estimates never become stable, we let the convergence rate be 200, the maximum number of iterations used in our simulation study. We find that the mean of all convergence rates from our simulation study is 77.91 and 81% of the convergence rates does not exceed 120. This indicates that the proposed procedure is capable of recovering the true social intercorrelation fairly quickly. Results from the LASSO regression using the convergence rate as the response variable are reported on the right of Table 2 and they reveal the following findings. First, there is no significant difference between the choices of $k = 1$ and $k = 2$ seeding nodes for the proposed approach, while the choice of $n = 50$ yields the fastest convergence rate. Second, the network size $N$ and the characteristic path $L$ have no significant influence, while networks with smaller CC and AvgDeg are associated with faster convergence rates from the proposed approach. Third, the proposed approach has a slightly slower convergence rate on the WS network than on other types of networks. Fourth, the magnitude of $\rho$ has no influence on the convergence rate of the proposed approach.

The above findings help shed light on the implementation guidelines of our proposed approach. First, regarding the number of seeding nodes, either $k = 1$ or $k = 2$ can be used. Next, regarding the maximum number of neighbors $n$ to be included in each sequential sample, $n = 50$ is recommended and we advise against going beyond 50 due to the high computational cost in the evaluation of the composite likelihood function at each iteration. In practice, one can also use a smaller $n$ to reduce the computational cost. In addition, for networks where the maximum number of connections for each network member is less than 50 (such as the small world networks in our study), no sample taken will reach size 50 and therefore a smaller $n$ should be used. Third, we recommend using the dynamic stopping rule as described in Section 3.2, that is, stopping the procedure when the absolute difference between the maximum and the minimum estimates from the last 20 consecutive estimates does not exceed 0.01. If the maximum number of sequential samples needs to be set as an alternative stopping rule, then, $B = 200$ can be used in general based on the simulation finding that the proposed procedure generally converges well before 200 iterations (average = 77.91). For networks other than the WS networks, a smaller $B = 100$ can be used. Finally, we recommend leveraging parallel computing to run the SEQ-MCLE multiple number (e.g., 30) of times in parallel to take the mean estimate of $\rho$. This also allows for the approximation of confidence intervals of the estimates.

## 5 Comparison with benchmark approaches

In this section, we compare the proposed SEQ-MCLE approach with the following benchmark methods: the population-based MLE, and the sample-based MLE. For the sample-based MLE, the snowball sampling method is used to obtain the sample based on recommendations from Chen et al. (2013). Note that the population-based MLE is only computationally feasible when the network is of small or moderate size. When the network size is large, only the sample-based MLE and the proposed SEQ-MCLE approach are feasible.

For a thorough comparison, we take an example from each network type outlined in Table 1 and generate 30 sets of response data for each of the five magnitudes of $\rho \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ based on the spatial error model (2). Then, for each dataset, we obtain model parameter estimates using the following three methods: (1) the proposed SEQ-MCLE with $k$, $n$ and $B$ chosen according to the guidelines discussed in Section 4.2; (2) the sample-based MLE with a snowball sample of size $Bn$; and (3) the population-based MLE (if computationally feasible).

### 5.1 On simulated networks

We start the comparison using PL1 as an example of the simulated power-law networks. The network characteristics of this power-law network are summarized in Table 1. The true parameter values used for the spatial error model are $\boldsymbol{\beta} = (1, 1)'$, $\sigma = 2$, and $\rho \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. Figure 1 shows the boxplots of the social intercorrelation parameter estimates over the 30 replications for each estimation

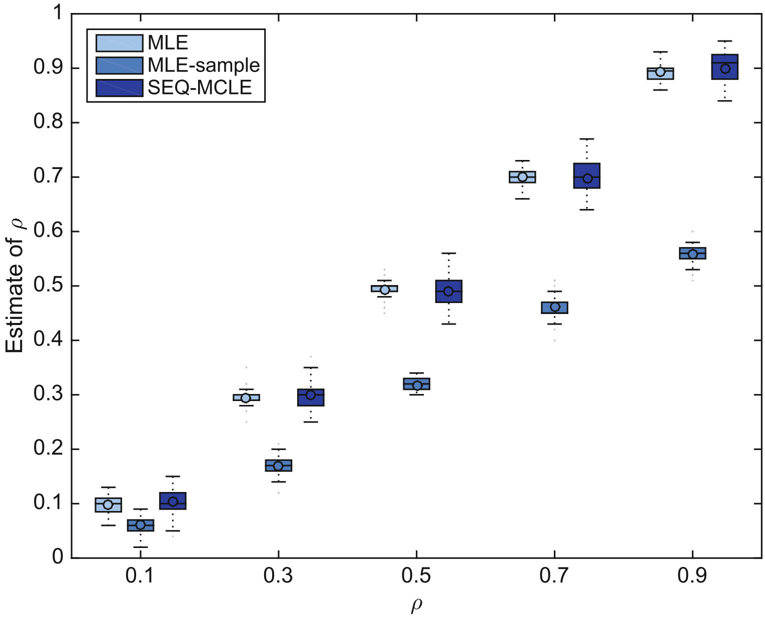

**Fig. 1** Boxplots of $\hat{\rho}$ for different estimation methods and under different parameter settings

method and each true value of $\rho$. The circles and the lines inside the boxes represent the means and the medians of the $\rho$ estimates. The mean and standard deviation (in parentheses) of the computation time for each method are reported in Table 3.

The boxplots of $\hat{\rho}$ clearly show an underestimation bias from the sample-based MLE. This is consistent with findings in Chen et al. (2013). The underestimation bias becomes more prominent when the magnitude of social intercorrelation increases from low to high. In contrast, our proposed SEQ-MCLE approach with $k = 2, n = 50$ and $B = 100$ can recover the true values of the social intercorrelation parameter for this power-law network. So does the population-based MLE, but it is computationally intensive. As shown in Table 3, the mean CPU time (in seconds) taken by the population-based MLE is around 30 to 50 times of that taken by the proposed approach.

Results on the other two types of simulated networks are consistent with those from the power-law network, as shown in Appendix D for the small world network example WS1L and the power-cluster network example PC1L, both with network size $N = 50000$. Note that the population-based MLE is no longer feasible computationally and thus the comparison is only between the sample-based MLE and the proposed SEQ-MCLE approach.

In regards to the other parameters in the spatial error model, we find that the proposed SEQ-MCLE approach can recover the true values of $\boldsymbol{\beta}$ and $\sigma$, just like the population-based MLE when it is computationally feasible. In contrast, the estimates obtained from the sample-based MLE start to deviate from the true values when the magnitude of the social intercorrelation increases from low to high, as shown in Appendix E. This makes intuitive sense because the estimates of $\boldsymbol{\beta}$ and $\sigma$ are dependent on the $\rho$ estimate, as shown in Eqs. 6 and 7. Thus, as the bias of the $\rho$ estimate from the sample-based MLE gets larger when the magnitude of the social intercorrelation increases, the biases in the estimates of $\boldsymbol{\beta}$ and $\sigma$ become more pominent.

The key message from the comparative study based on the three types of simulated networks is the following: there are significant advantages of using the proposed SEQ-MCLE approach over benchmark estimation methods—it is computationally efficient and at the same time achieves high estimation accuracy. This makes the proposed SEQ-MCLE approach a good choice for the analysis of large social networks, as we will demonstrate next using real networks of large size.

**Table 3** Mean and standard deviation of the computation time (in seconds) under five choices of $\rho$

| Method | $\rho = 0.1$ | $\rho = 0.3$ | $\rho = 0.5$ | $\rho = 0.7$ | $\rho = 0.9$ |
|---|---|---|---|---|---|
| MLE | 9271 | 9019 | 9281 | 9368 | 9209 |
| | (232) | (226) | (417) | (502) | (433) |
| MLE-sample | 1233 | 1243 | 1206 | 1202 | 1196 |
| | (49) | (29) | (34) | (27) | (38) |
| SEQ-MCLE | 191 | 188 | 187 | 189 | 261 |
| | (12) | (16) | (11) | (10) | (22) |

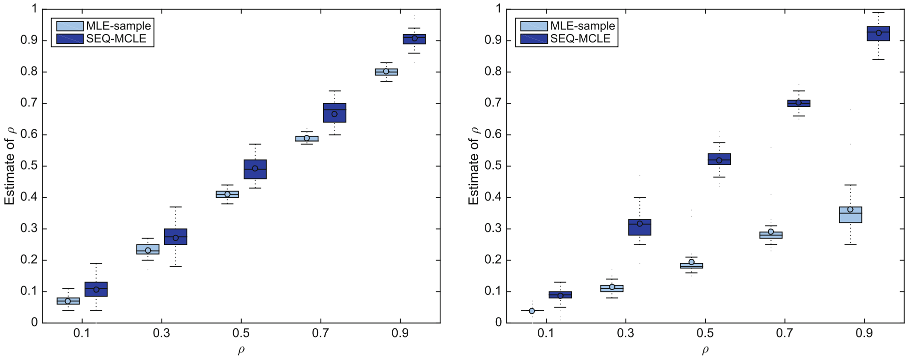

**Fig. 2** Boxplots of $\hat{\rho}$ for the two real networks—Colla (left) and Flickr (right)

## 5.2 On real networks

We report in Fig. 2 and Table 4 the results from the comparative study based on the two real networks. As described in Table 1, the Colla network is a scientific collaboration network with size $N = 23,133$ and 93,497 edges, and the Flickr network is an image sharing network with network size $N = 105,938$ and over 2 million edges. The population-based MLE is not feasible computationally and we focus on the comparison between the proposed SEQ-MCLE approach and the sample-based MLE. For the proposed SEQ-MCLE approach, we follow the guidelines discussed in Section 4.2 and use $k = 2$, $n = 50$, $B = 100$ for the Colla network, and $k = 2$, $n = 50$, $B = 200$ for the Flickr network. Correspondingly, we take snowball samples of size $Bn$ for the sample-based MLE under comparison.

The results on these real networks clearly demonstrate the benefit of using the proposed SEQ-MCLE approach for the social intercorrelation estimation. It recovers the true social intercorrelation parameter within a relatively short amount of time. In contrast, the sample-based MLE significantly underestimates the strength of the social intercorrelation. Consistent with the findings on the simulated networks, the underestimation bias is intensified when the true social intercorrelation is high. For

**Table 4** Mean and standard deviation of the computation time for Colla (top) and Flickr (bottom)

| Method | $\rho = 0.1$ | $\rho = 0.3$ | $\rho = 0.5$ | $\rho = 0.7$ | $\rho = 0.9$ |
|---|---|---|---|---|---|
| MLE-sample | 1191 | 1151 | 1187 | 1146 | 1188 |
| | (53) | (19) | (27) | (31) | (28) |
| SEQ-MCLE | 176 | 162 | 359 | 368 | 258 |
| | (52) | (33) | (86) | (106) | (85) |
| | | | | | |
| MLE-sample | 9030 | 8726 | 8721 | 8793 | 8745 |
| | (536) | (922) | (938) | (809) | (908) |
| SEQ-MCLE | 1434 | 2407 | 1637 | 1310 | 1587 |
| | (475) | (974) | (464) | (441) | (351) |

example, when the true social intercorrelation $\rho$ is 0.9, the estimate produced by the sample-based MLE has a mean of 0.3627 for the Flickr network, which is significantly lower than the true value. Such a bias can easily result in economic harm to managers using the information to make marketing decisions, which we will discuss in more detail in the next section.

### 5.3 Marketing implications

The proposed SEQ-MCLE approach provides a valuable tool to marketing researchers and practitioners, with important managerial implications. It enables managers to make better business decisions that are based on accurate estimates of social intercorrelations, such as market promotions (Bronnenberg and Mahajan 2001), brand management (Bronnenberg and Sismeiro 2002), and allocation of advertising spending (Aravindakshan et al. 2012). On the other hand, biased social intercorrelation estimates can easily result in wrong decisions with material consequences (Chen et al. 2013).

To see the important role that an accurate estimate of the social intercorrelation plays, note that the estimates of the regression coefficients $\boldsymbol{\beta}$ and the variance parameter $\sigma$ all depend on the estimate of the social intercorrelation parameter $\rho$. This can be seen clearly in Eqs. 6 and 7 for the MLE estimation, and in Eqs. 14 and 15 for the MCLE estimation. Consequently, any prediction $\hat{\mathbf{y}}$ given a new $\mathbf{X}$ is also dependent on the social intercorrelation estimate $\hat{\rho}$. Therefore, accurate learning of social intercorrelation is critical to managers in making business decisions based on these estimates or predictions. For example, if a pricing decision is to be made based on the estimate of $\beta_{price}$, then a biased $\rho$ estimate would lead to a biased estimate of the price coefficient, resulting in suboptimal pricing decisions. Similarly, if an investment decision is to be made based on the predicted responses $\hat{\mathbf{y}}$, then a biased $\rho$ estimate would lead to inaccurate predictions that would steer the investment decision to the wrong way.

For further illustration, take the "PL1" network in our simulation study with $\rho = 0.8$ and let it represent the network of stores available to a brand. For simplicity, let $x$ represent the magnitude of brand promotion and the brand sales $y$ follow a spatial model with a single predictor variable $x$ and parameters $\boldsymbol{\beta} = (0, 0.5)'$, $\sigma = 2$. Suppose the brand manager is trying to decide if it is a good investment to increase the magnitude of its promotion by 1 unit at the top 100 stores with the most number of connections. Based on the model parameter estimates obtained from previous sales responses, the return on investment (ROI) can be predicted as the increment in brand sales minus the cost ($100 \times 1 = 100$), then divided by the cost. An accurate estimate of social intercorrelation from the proposed approach helps managers make better predictions on brand sales increase and subsequently the ROI. In contrast, an underestimated social intercorrelation from the sample-based MLE can result in wrong predictions which have material consequences. This can be seen in Fig. 3 where the predicted ROI is plotted against the bias in the estimate of $\rho$. As shown in the plot, while the ROI based on the true value of $\rho$ is positive and approximately 1.75, it is predicted to be negative when the bias in $\hat{\rho}$ is larger than 0.2. The wrong negative prediction would clearly lead managers to the wrong decision and result in financial losses.

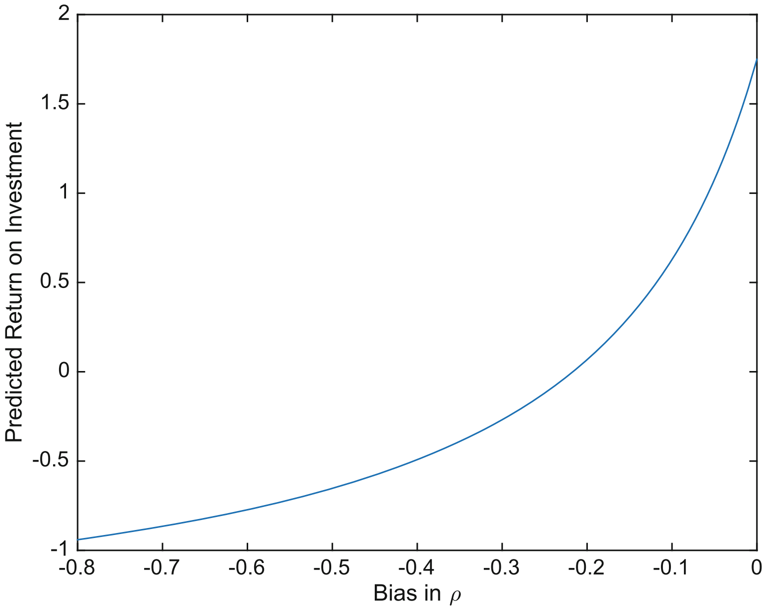

**Fig. 3** The effect of the bias in $\hat{\rho}$ on the predicted return on investment

## 6 Extensions of the proposed approach

So far we have assumed a known population network or the analyst has the firm's permission to access the entire network (e.g., Trusov et al. 2010; Katona et al. 2011). However, it is not always practical to have such knowledge. In this section, we discuss how the proposed approach can be applied when the structure of the population network is unknown. We also discuss how the proposed approach can be extended to an alternative form of the spatial model—the spatial lag model.

### 6.1 When the structure of the population network is unknown

In the context of an unknown population network, we can first take a sample network ("approximation sample") that approximates the structure of the population network and then apply the proposed SEQ-MCLE approach on the approximation sample. The performance of the proposed SEQ-MCLE approach thus heavily depends on how good the approximation sample is in preserving the structure of the population network.

To obtain a good approximation sample that works well with the proposed SEQ-MCLE approach, we make the following three recommendations: First, use the multiple-seeds snowball sampling method to take the approximation sample because it better preserves the population network structure (Chen et al. 2013). Second, to work well with the SEQ-MCLE estimation, the number of nodes in all but the last two stages of the approximation sample needs to be at least $Bn$, such that the maximum $B$ sequential samples $\{\mathbf{s}_1, \ldots, \mathbf{s}_B\}$ of the SEQ-MCLE approach will be taken from all nodes except those in the last two stages of the approximation sample. Third, to control computational cost, one may need to set an upper bound for the number of nodes in all but the last two stages of the approximation sample. This is necessary especially when the average degree of the network is so large that the number of nodes in the last two stages may explode if no upper bound is set.

To better understand the second recommendation above, note that the evaluation of the term $\mathbf{Q}_{\mathbf{s}\mathbf{s}^c}(\mathbf{y}_{\mathbf{s}^c} - \mathbf{x}_{\mathbf{s}^c}\boldsymbol{\beta})$ in the conditional distribution (11) only depends on a subset of $\mathbf{s}^c$ such that the corresponding elements in $\mathbf{Q}_{\mathbf{s}\mathbf{s}^c}$ are nonzero. That is, it only depends on the subset of nodes in $\mathbf{s}^c$ that are either directly connected to the nodes in the sequential sample $\mathbf{s}$ or indirectly connected through common neighbors. Hence, if all $B$ sequential samples $\{\mathbf{s}_1, \ldots, \mathbf{s}_B\}$ of the SEQ-MCLE approach are restricted to the nodes in the $m - 2$ stages of the approximation sample with $m$ stages, then we can approximate the conditional distribution well because we have access to all nodes that are directly connected to the sequential samples and a good portion of the nodes that are indirectly connected.

We give an illustration example using the data on the Flickr network in Section 4 for the case where $\rho = 0.1$. Suppose the structure of this large network is unknown. To take the approximation sample which works well with the SEQ-MCLE approach with the setting $k = 2$, $n = 10$ and $B = 100$, we follow the recommendation provided above. Once we reach the stage in snowball sampling where the total number of nodes exceeds $Bn = 1000$, we keep sampling for two additional stages before stopping to obtain the approximation sample. To avoid explosion of the number of nodes, we also set the upper bound to be 1500 for the number of nodes in all but the last two stages of the approximation sample. As a result, an approximation sample is obtained with five seeding nodes, five stages of snowball sampling and a total of 35153 nodes. We then apply the SEQ-MCLE approach on the approximation sample and report in Fig. 4 the time-series plots of the average estimates of $\rho$ over 30 replications of the proposed approach (blue solid line), as well as the 90% and 10% quantiles of the estimates (blue dashed line). The plots demonstrate that when the population network structure is unknown, the proposed SEQ-MCLE approach can work with a good approximation sample in recovering the true social intercorrelation parameter (red line).

### 6.2 Extension to the spatial lag model

An alternative form of the spatial model used in marketing is the spatial lag model that captures the lagged effects of the dependent variable (see review by Bradlow et al. 2005), with the following expression

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{16}$$

where $\mathbf{y}$ and $\boldsymbol{\epsilon}$ are $N \times 1$ vectors, $\mathbf{X}$ is an $N \times p$ matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. The expressions of the likelihood function and corresponding maximum likelihood estimates for the spatial lag model can be found in Appendix A.

The proposed SEQ-MCLE approach can be similarly applied to the spatial lag model. The only difference is that the conditional distribution of the

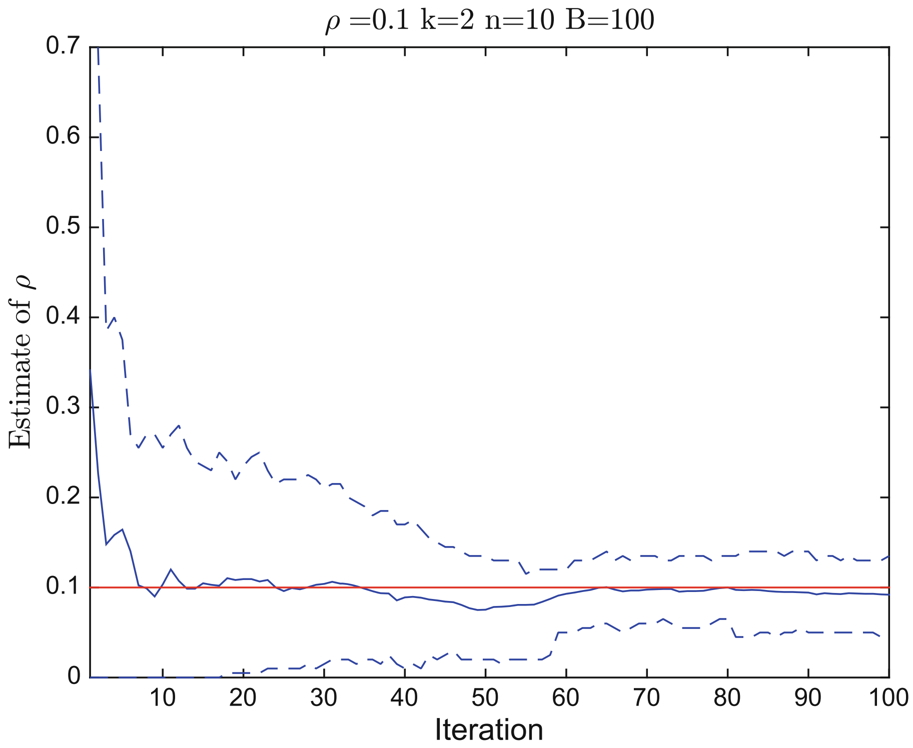

**Fig. 4** Estimates of $\rho$ When Structure of the Population Network is Unknown

observations in the sample $\mathbf{y}_\mathbf{s}$ given the rest of the observations in the population $\mathbf{y}_{\mathbf{s}^\mathbf{c}}$ becomes

$$\mathbf{y}_\mathbf{s}|\mathbf{y}_{\mathbf{s}^c} \sim N(\tilde{\mathbf{X}}_\mathbf{s}\boldsymbol{\beta} - \mathbf{Q}_{\mathbf{ss}}^{-1}\mathbf{Q}_{\mathbf{ss}^c}(\mathbf{y}_{\mathbf{s}^c} - \tilde{\mathbf{X}}_{\mathbf{s}^c}\boldsymbol{\beta}), \sigma^2\mathbf{Q}_{\mathbf{ss}}^{-1}), \tag{17}$$

where $\tilde{\mathbf{X}} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}$. Details of the derivation can be found in Appendix B. For computational efficiency, we approximate $(\mathbf{I} - \rho\mathbf{W})^{-1}$ by $\mathbf{I} + \rho\mathbf{W}$ or $\mathbf{I} + \rho\mathbf{W} + \rho^2\mathbf{W}^2$ (Petersen and Pedersen 2008). According to Theorem 4.20 at Page 55 of Stewart (1998), such approximations work well given that $|\rho| < 1$.

To examine the performance of the proposed approach on the spatial lag model, we conduct a simulation study using the same simulation design as described in Section 4.1 except that the spatial error model is replaced with the spatial lag model. We find similar simulation results. The proposed approach works well in general across all types of networks in recovering the true spatial lag effect $\rho$. The time series plots of the estimates are provided in the supplemental online Appendix.

## 7 Discussion and conclusion

In this paper, we have proposed SEQ-MCLE, a sequential sampling enhanced composite likelihood approach for efficient estimation of social intercorrelations in large-scale networks. In the proposed SEQ-MCLE approach, small samples are sequentially taken from a given known population network. At each stage of the sequential procedure, a composite likelihood function is formed based on the samples obtained so far where each component of the likelihood function is the conditional density of one sample given the rest of the population. A set of parameter estimates is then obtained by maximizing the composite likelihood function. The parameter estimates are gradually improved until the sequential procedure converges or stops according to a pre-specified stopping rule.

Through comparative studies based on both simulated networks and real networks, we have demonstrated the superior performance of the proposed approach over benchmark estimation methods. In comparison to population-based MLE that becomes computationally prohibitive when the network size gets large, the proposed approach scales well to large network data. In comparison to sample-based MLE that relies on the information from the sample only and thus produces biased estimates of social intercorrelations, the proposed approach effectively uses information from the population without compromising computational efficiency.

When the structure of the population network is unknown, the proposed SEQ-MCLE approach can be extended to work well with a carefully selected sample of the network that approximates the structure of the network, as discussed in Section 6. We have also shown in Section 6 the extension of the SEQ-MCLE approach to the spatial lag model for the estimation of the overall lagged effect on the dependent variable.

We have focused on the scenario where the interest is in the learning of the overall social intercorrelation, or the effect of the overall/global network topology on the social interactions among network members. If there are reasons to believe that a large network consists of multiple communities with different magnitudes of social intercorrelations, then the proposed SEQ-MCLE approach can be applied separately to obtain a social intercorrelation estimate for each community of the network. This is a straightforward extension if the communities are known in advance, such as the communities formed by geographic locations. In circumstances where the communities are latent, community-detection algorithms can be used first to group the nodes of the network into different communities. Then the SEQ-MCLE approach can be applied to obtain a separate social intercorrelation estimate for each community.

There are several limitations of our research that call for further investigation. First, all samples are currently given the same weight in the composite likelihood in the proposed approach. To further improve computational efficiency, it is worth considering allocating more weights to the more influential samples through mechanisms such as sample reweighting (Bradlow and Zaslavsky 1997).

Second, we have focused on a static model that does not take into account of the time variation. For evolving networks with evolving social intercorrelations, the SEQ-MCLE approach can be applied at different time points to gain understandings of how the social intercorrelations change over time. The social intercorrelation estimate from the previous time point can be used to narrow the range of the grid search and speed up computation at the new time point. However, a big drawback is that temporal dependence is not accounted for in the underlying model. To address this issue, the SEQ-MCLE approach needs to be modified for the spatio-temporal model that accounts for both spatial and temporal dependence. This creates considerable computational complexity and we leave it for future research.

Finally, the independent variable $x$ in the spatial model investigated in our study is assumed to be exogenous to the network structure. When $x$ is dependent of the network structure (e.g., when $x$ for a network member is defined to be the number of connections the member has), the representativeness of $x$ should also be considered in each sequential sample of the proposed approach. Without the consideration of the representativeness of $x$, there may not be enough variation for the efficient estimation of the regression coefficient $\boldsymbol{\beta}$. For example, if all nodes in a sample have an equal number of connections, then there is not any variation in $x$ and it would be problematic to estimate $\boldsymbol{\beta}$. To ensure the representativeness of $x$, methodologies from the experimental design literature (see Atkinson et al. 2007 for a review) can be used to help with the optimal selection of nodes for each sequential sample in the proposed approach.

**Acknowledgments** This research was supported in part by the Vilas Life Cycle Professorship Grant from University of Wisconsin-Madison and NSF Grant 1564376.

## Appendix A: Maximum likelihood estimates of the spatial lag model

The maximum likelihood estimates of the model parameters $(\rho, \boldsymbol{\beta}, \sigma^2)$ in the spatial lag model (16) can be obtained by maximizing the log likelihood function

$$l(\rho, \boldsymbol{\beta}, \sigma^2) = -\frac{N}{2}\log 2\pi\sigma^2 + \log|\det(\mathbf{I} - \rho\mathbf{W})| - \frac{1}{2\sigma^2}\|(\mathbf{I} - \rho\mathbf{W})\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2. \quad (18)$$

Given $\rho$, the maximum likelihood estimates of $\boldsymbol{\beta}$ and $\sigma^2$ can be expressed in closed-forms as $\hat{\boldsymbol{\beta}}(\rho)$ and $\hat{\sigma}^2(\rho)$ such that

$$\hat{\boldsymbol{\beta}}(\rho) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{y}}, \quad (19)$$

$$\hat{\sigma}^2(\rho) = \frac{1}{N}\|\tilde{\mathbf{y}} - \mathbf{X}\hat{\boldsymbol{\beta}}(\rho)\|_2^2, \quad (20)$$

where $\tilde{\mathbf{y}} = (\mathbf{I} - \rho\mathbf{W})\mathbf{y}$. Substituting $\boldsymbol{\beta}$ and $\sigma^2$ in the log likelihood function (18) with $\hat{\boldsymbol{\beta}}(\rho)$ and $\hat{\sigma}^2(\rho)$, the maximum likelihood estimate of $\rho$ can be obtained by maximizing

$$l(\rho, \hat{\boldsymbol{\beta}}(\rho), \hat{\sigma}^2(\rho)) = -\frac{N}{2}\log 2\pi\hat{\sigma}^2(\rho) + \log|\det(\mathbf{I} - \rho\mathbf{W})| - \frac{N}{2}. \quad (21)$$

## Appendix B: SEQ-MCLE of the spatial lag model

To find the parameter estimates that maximize the composite likelihood function (13), note that when $t$ samples have been collected, the estimate on the social intercorrelation parameter $\hat{\rho}^{(t)}$ based on the composite likelihood is the one that maximizes:

$$\prod_{i=1}^{t} L(\rho, \boldsymbol{\beta}_c(\rho), \sigma_c^2(\rho), \mathbf{y}_{\mathbf{s}_i}|\mathbf{y}_{\mathbf{s}_i^c}), \quad (22)$$

where $\boldsymbol{\beta}_c(\rho)$ and $\sigma_c^2(\rho)$ are the closed-form expressions of the estimates to $\boldsymbol{\beta}$ and $\sigma^2$ which maximize the composite likelihood function $\prod_{i=1}^{t} L(\boldsymbol{\theta}, \mathbf{y}_{\mathbf{s}_i}|\mathbf{y}_{\mathbf{s}_i^c})$ given $\rho$. After $\hat{\rho}^{(t)}$ is obtained, the estimates to $\boldsymbol{\beta}$ and $\sigma^2$ are given by $\boldsymbol{\beta}_c(\hat{\rho}^{(t)})$ and $\sigma_c^2(\hat{\rho}^{(t)})$.

For the spatial lag model, we have

$$\boldsymbol{\beta}_c(\rho) = \left(\sum_{i=1}^{t} \tilde{\tilde{\mathbf{X}}}'_{\mathbf{s}_i} \mathbf{Q}_{\mathbf{s}_i\mathbf{s}_i} \tilde{\tilde{\mathbf{X}}}_{\mathbf{s}_i}\right)^{-1} \left(\sum_{i=1}^{t} \tilde{\tilde{\mathbf{X}}}'_{\mathbf{s}_i} \mathbf{Q}_{\mathbf{s}_i\mathbf{s}_i} \tilde{\mathbf{y}}_{\mathbf{s}_i}\right), \tag{23}$$

$$\sigma_c^2(\rho) = \frac{1}{\sum_{i=1}^{t} |\mathbf{s}_i|} \sum_{i=1}^{t} (\tilde{\mathbf{y}}_{\mathbf{s}_i} - \tilde{\tilde{\mathbf{X}}}_{\mathbf{s}_i} \boldsymbol{\beta}_c(\rho))' \mathbf{Q}_{\mathbf{s}_i\mathbf{s}_i} (\tilde{\mathbf{y}}_{\mathbf{s}_i} - \tilde{\tilde{\mathbf{X}}}_{\mathbf{s}_i} \boldsymbol{\beta}_c(\rho)), \tag{24}$$

where $|\mathbf{s}_i|$ is the size of sample $\mathbf{s}_i$, $\tilde{\mathbf{y}}_{\mathbf{s}_i} = \mathbf{y}_{\mathbf{s}_i} + \mathbf{Q}^{-1}_{\mathbf{s}_i\mathbf{s}_i} \mathbf{Q}_{\mathbf{s}_i\mathbf{s}_i^c} \mathbf{y}_{\mathbf{s}_i^c}$, $\tilde{\tilde{\mathbf{X}}}_{\mathbf{s}_i} = \tilde{\mathbf{X}}_{\mathbf{s}_i} + \mathbf{Q}^{-1}_{\mathbf{s}_i\mathbf{s}_i} \mathbf{Q}_{\mathbf{s}_i\mathbf{s}_i^c} \tilde{\mathbf{X}}_{\mathbf{s}_i^c}$ and $\tilde{\mathbf{X}} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}$.

## Appendix C: Proof of the consistency of the SEQ-MCLE estimates

The SEQ-MCLE estimates are obtained by solving

$$S(\mathbf{y}, \mathbf{s}, \boldsymbol{\theta}) = \sum_{i=1}^{B} \frac{\partial \log L(\boldsymbol{\theta}, \mathbf{y}_{\mathbf{s}_i}|\mathbf{y}_{\mathbf{s}_i^c})}{\partial \boldsymbol{\theta}} = 0, \tag{25}$$

where $S(\mathbf{y}, \mathbf{s}, \boldsymbol{\theta})$ is the estimating function, and $\mathbf{s} = (\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_B)$, with $\mathbf{s}_i$ being the $i$th sample in the proposed approach. To prove the consistency of the SEQ-MCLE estimates, we need to prove that this estimating function is unbiased such that $E_{\boldsymbol{\theta}}[S(\mathbf{y}, \mathbf{s}, \boldsymbol{\theta})] = 0$. This is because that according to Desmond (1997), Bera et al. (2006), the unbiasedness of the estimating function $S(\mathbf{y}, \mathbf{s}, \boldsymbol{\theta})$ implies the consistency of the estimates obtained by solving the estimating function. Note that the estimating function $S(\mathbf{y}, \mathbf{s}, \boldsymbol{\theta})$ is a $(p+2)$-dimensional vector and now let us focus on one component of $S(\mathbf{y}, \mathbf{s}, \boldsymbol{\theta})$, for example, the component corresponding to $\rho$. Given a known population network and under standard regularity conditions such that the integrals and derivatives are exchangeable, let $f_{\boldsymbol{\theta}}$ represent the p.d.f. or conditional p.d.f. of the corresponding variables, let $\vec{\mathbf{s}} = (\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_B)$, then we have

$$\begin{aligned}
& E_{\boldsymbol{\theta}}\left[\sum_{i=1}^{B} \frac{\partial \log L(\boldsymbol{\theta}, \mathbf{y}_{\mathbf{s}_i}|\mathbf{y}_{\mathbf{s}_i^c})}{\partial \rho}\right] \\
&= E_{\boldsymbol{\theta}}\left[\sum_{i=1}^{B} \frac{\partial \log f_{\boldsymbol{\theta}}(\mathbf{y}_{\mathbf{s}_i}|\mathbf{y}_{\mathbf{s}_i^c}, \mathbf{s}_i)}{\partial \rho}\right] \\
&= \sum_{i=1}^{B} \int\int \frac{\partial \log f_{\boldsymbol{\theta}}(\mathbf{y}_{\mathbf{s}_i}|\mathbf{y}_{\mathbf{s}_i^c}, \mathbf{s_i})}{\partial \rho} f_{\boldsymbol{\theta}}(\mathbf{y}, \vec{\mathbf{s}}) d\mathbf{y} d\vec{\mathbf{s}} \\
&= \sum_{i=1}^{B} \int\int \frac{\partial f_{\boldsymbol{\theta}}(\mathbf{y}_{\mathbf{s}_i}|\mathbf{y}_{\mathbf{s}_i^c}, \mathbf{s_i})}{\partial \rho} \frac{f_{\boldsymbol{\theta}}(\mathbf{y}, \vec{\mathbf{s}})}{f_{\boldsymbol{\theta}}(\mathbf{y}_{\mathbf{s}_i}|\mathbf{y}_{\mathbf{s}_i^c}, \mathbf{s_i})} d\mathbf{y} d\vec{\mathbf{s}}
\end{aligned}$$

which can be expanded as

$$\int\int \frac{\partial f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_1}|\mathbf{y}_{\mathbf{s}_1^c}, \mathbf{s}_1)}{\partial\rho} \frac{f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_1}|\mathbf{y}_{\mathbf{s}_1^c}, \mathbf{s}_1) f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_1^c}, \mathbf{s}_1) f_{\boldsymbol\theta}(\mathbf{s}_2|\mathbf{y}, \mathbf{s}_1) f_{\boldsymbol\theta}(\mathbf{s}_3|\mathbf{y}, \mathbf{s}_1, \mathbf{s}_2)\ldots f_{\boldsymbol\theta}(\mathbf{s}_B|\mathbf{y}, \mathbf{s}_1,\ldots, \mathbf{s}_{B-1})}{f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_1}|\mathbf{y}_{\mathbf{s}_1^c}, \mathbf{s}_1)} d\mathbf{y}d\vec{\mathbf{s}} +$$
$$\int\int \frac{\partial f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_2}|\mathbf{y}_{\mathbf{s}_2^c}, \mathbf{s}_2)}{\partial\rho} \frac{f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_2}|\mathbf{y}_{\mathbf{s}_2^c}, \mathbf{s}_1, \mathbf{s}_2) f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_2^c}, \mathbf{s}_1, \mathbf{s}_2) f_{\boldsymbol\theta}(\mathbf{s}_3|\mathbf{y}, \mathbf{s}_1, \mathbf{s}_2)\ldots f_{\boldsymbol\theta}(\mathbf{s}_B|\mathbf{y}, \mathbf{s}_1, \ldots, \mathbf{s}_{B-1})}{f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_2}|\mathbf{y}_{\mathbf{s}_2^c}, \mathbf{s}_2)} d\mathbf{y}d\vec{\mathbf{s}} +$$
$$\int\int \frac{\partial f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_3}|\mathbf{y}_{\mathbf{s}_3^c}, \mathbf{s}_3)}{\partial\rho} \frac{f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_3}|\mathbf{y}_{\mathbf{s}_3^c}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3) f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_3^c}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3)\ldots f_{\boldsymbol\theta}(\mathbf{s}_B|\mathbf{y}, \mathbf{s}_1,\ldots, \mathbf{s}_{B-1})}{f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_3}|\mathbf{y}_{\mathbf{s}_3^c}, \mathbf{s}_3)} d\mathbf{y}d\vec{\mathbf{s}} +$$
$$\ldots$$
$$\int\int \frac{\partial f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_B}|\mathbf{y}_{\mathbf{s}_B^c}, \mathbf{s}_B)}{\partial\rho} \frac{f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_B}|\mathbf{y}_{\mathbf{s}_B^c}, \mathbf{s}_1, \mathbf{s}_2,\ldots, \mathbf{s}_B) f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_B^c}, \mathbf{s}_1, \mathbf{s}_2,\ldots, \mathbf{s}_B)}{f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_B}|\mathbf{y}_{\mathbf{s}_B^c}, \mathbf{s}_B)} d\mathbf{y}d\vec{\mathbf{s}}$$

Note that

$$f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_2}|\mathbf{y}_{\mathbf{s}_2^c}, \mathbf{s}_1, \mathbf{s}_2) = f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_2}|\mathbf{y}_{\mathbf{s}_2^c}, \mathbf{s}_2)$$
$$\ldots$$
$$f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_B}|\mathbf{y}_{\mathbf{s}_B^c}, \mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_B) = f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_B}|\mathbf{y}_{\mathbf{s}_B^c}, \mathbf{s}_B)$$

and that for $i = 1, \ldots, B - 1$,

$$f_{\boldsymbol\theta}(\mathbf{s}_{i+1}|\mathbf{y}, \mathbf{s}_1, \ldots, \mathbf{s}_i) = f_{\boldsymbol\theta}(\mathbf{s}_{i+1}|\mathbf{y}_{\mathbf{s}_i}, \mathbf{y}_{\mathbf{s}_i^c}, \mathbf{s}_1, \ldots, \mathbf{s}_i)$$

Because the seeding nodes $\mathbf{j}_{i+1}$ for the sample $\mathbf{s}_{i+1}$ are obtained deterministically given $\mathbf{y}_{\mathbf{s}_i}, \mathbf{y}_{\mathbf{s}_i^c}, \mathbf{s}_1, \ldots, \mathbf{s}_i$. Thus,

$$f_{\boldsymbol\theta}(\mathbf{s}_{i+1}|\mathbf{y}_{\mathbf{s}_i}, \mathbf{y}_{\mathbf{s}_i^c}, \mathbf{s}_1, \ldots, \mathbf{s}_i) = \mathbb{P}_{\boldsymbol\theta}(\mathbf{S}_{i+1} = \mathbf{s}_{i+1}|\mathbf{j}_{i+1})$$

Let $N_{i+1}$ denote the number of neighbors of $\mathbf{j}_{i+1}$ that have not been included in the previous samples. If it is less than the recommended sample size $n$ (30 or 50) in the proposed SEQ-MCLE approach, then $\mathbb{P}_{\boldsymbol\theta}(\mathbf{S}_{i+1} = \mathbf{s}_{i+1}|\mathbf{j}_{i+1}) = 1$ because all neighbors are deterministally included in the sample. Otherwise, $n$ neighbors out of the total $N_{i+1}$ are randomly selected and the above probability equals $\frac{1}{\binom{N_{i+1}}{n}}$. Hence, $f_{\boldsymbol\theta}(\mathbf{s}_{i+1}|\mathbf{y}, \mathbf{s}_1, \ldots, \mathbf{s}_i)$ is a constant and

$$\int f_{\boldsymbol\theta}(\mathbf{s}_{i+1}|\mathbf{y}, \mathbf{s}_1, \ldots, \mathbf{s}_i)d\mathbf{s}_{i+1} = 1.$$

The above expansion therefore simplifies to

$$\sum_{i=1}^{B}\int\int\int \frac{\partial f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_i}|\mathbf{y}_{\mathbf{s}_i^c}, \mathbf{s_i})}{\partial\rho} f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_i^c}, \mathbf{s_1}, \ldots, \mathbf{s_i})d\mathbf{y}_{\mathbf{s}_i}d\mathbf{y}_{\mathbf{s}_i^c}d\mathbf{s}_1\ldots d\mathbf{s}_i$$
$$= \sum_{i=1}^{B}\int\int\left[\int \frac{\partial f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_i}|\mathbf{y}_{\mathbf{s}_i^c}, \mathbf{s_i})}{\partial\rho} d\mathbf{y}_{\mathbf{s}_i}\right] f_{\boldsymbol\theta}(\mathbf{y}_{\mathbf{s}_i^c}, \mathbf{s_1}, \ldots, \mathbf{s_i})d\mathbf{y}_{\mathbf{s}_i^c}d\mathbf{s}_1\ldots d\mathbf{s}_i$$

$$= \sum_{i=1}^{B} \int \int \left[ \frac{\partial}{\partial \rho} \int f_{\boldsymbol{\theta}}(\mathbf{y}_{\mathbf{s}_i} | \mathbf{y}_{\mathbf{s}_i^c}, \mathbf{s_i}) d\mathbf{y}_{\mathbf{s}_i} \right] f_{\boldsymbol{\theta}}(\mathbf{y}_{\mathbf{s}_i^c}, \mathbf{s_1}, \ldots, \mathbf{s_i}) d\mathbf{y}_{\mathbf{s}_i^c} d\mathbf{s}_1 \ldots d\mathbf{s}_i$$

$$= 0, \quad \text{because} \int f_{\boldsymbol{\theta}}(\mathbf{y}_{\mathbf{s}_i} | \mathbf{y}_{\mathbf{s}_i^c}, \mathbf{s_i}) d\mathbf{y}_{\mathbf{s}_i} = 1.$$

Similarly, one can show the above fact holds for all the other components of $S(\mathbf{y}, \mathbf{s}, \boldsymbol{\theta})$ proving

$$E_{\boldsymbol{\theta}}[S(\mathbf{y}, \mathbf{s}, \boldsymbol{\theta})] = 0.$$

## Appendix D: Comparative study on "WS1L" and "PC1L"

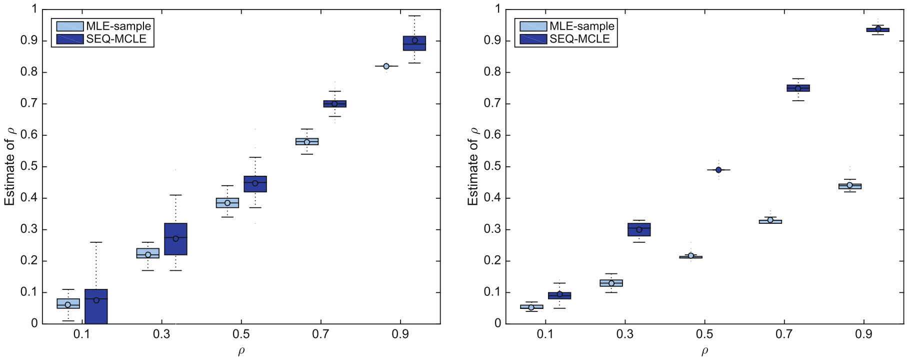

**Fig. 5** Boxplots of $\hat{\rho}$ for different estimation methods under "WS1L" (left) and "PC1L" (right)

**Table 5** Mean and standard deviation of the computation time for "WS1L" (top) and "PC1L" (bottom)

| Method | $\rho = 0.1$ | $\rho = 0.3$ | $\rho = 0.5$ | $\rho = 0.7$ | $\rho = 0.9$ |
|---|---|---|---|---|---|
| MLE-sample | 6123 | 4832 | 4636 | 5906 | 5510 |
| | (776) | (183) | (57) | (640) | (658) |
| SEQ-MCLE | 1089 | 1315 | 1107 | 1277 | 953 |
| | (446) | (339) | (235) | (262) | (268) |
| MLE-sample | 3746 | 3934 | 3822 | 3687 | 4040 |
| | (69) | (108) | (126) | (43) | (48) |
| SEQ-MCLE | 320 | 188 | 315 | 223 | 227 |
| | (41) | (183) | (59) | (9) | (321) |

## Appendix E: The estimation of $\beta_1$ and $\sigma$ in the simulation study

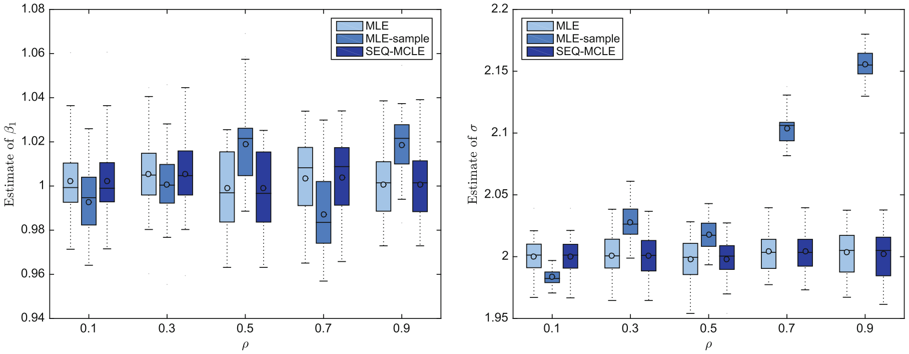

**Fig. 6** Boxplots of $\hat{\beta}_1$ and $\hat{\sigma}$ with $\rho = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ under "PL1" for the spatial error model

## References

Anselin, L. (1988). *Spatial econometrics: methods and models*. Dorddrecht: Kluwer Academic Publishers.

Aravindakshan, A.W., Peters, K., Naik, P.A. (2012). Spatiotemporal allocation of advertising budgets. *Journal of Marketing Research*, *49*, 1–14.

Atkinson, A., Donev, A., Tobias, R. (2007). *Optimum experimental designs, with SAS* (Vol. 34). Oxford: Oxford University Press.

Barabási, A.L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.

Barabási, A.L., & Bonabeau, E. (2003). Scale-free networks. *Scientific American*, *288*(5), 50–59.

Barry, R.P., & Pace, R. (1999). Monte Carlo estimates of the log determinant of large sparse matrices. *Linear Algebra and its Applications*, *289*(1–3), 41–54.

Bera, A.K., Bilias, Y., Simlai, P. (2006). Estimating functions and equations: an essay on historical developments with applications to econometrics. *Palgrave Handbook of Econometrics*, *1*, 427–476.

Besag, J. (1975). Statistical analysis of non-lattice data. *Statistician*, *24*(3), 179–195.

Bradlow, E.T., & Zaslavsky, A.M. (1997). Case influence analysis in bayesian inference. *Journal of Computational and Graphical Statistics*, *6*(3), 314–331.

Bradlow, E.T., Bronnenberg, B., Russell, G., Arora, N., Bell, D.R., Duvvuri, S.D., Hofstede, F.T., Sismeiro, C., Thomadsen, R., Yang, S. (2005). Spatial models in marketing. *Marketing Letters*, *16*(3/4), 267–278.

Bronnenberg, B.J., & Mahajan, V. (2001). Unobserved retailer behavior in multimarket data: joint spatial dependence in marketing shares and promotion variables. *Marketing Science*, *20*(3), 284–299.

Bronnenberg, B.J., & Sismeiro, C. (2002). Using multimarket data to predict brand performance in markets for which no or poor data exist. *Journal of Marketing Research*, *39*, 1–17.

Chen, X., Chen, Y., Xiao, P. (2013). The impact of sampling and network topology on the estimation of social intercorrelations. *Journal of Marketing Research*, *50*(1), 95–110.

Cohn, D., Ghahramani, Z., Jordan, M. (1996). Active learning with statistical models. *Journal of Artificial Intelligence Research*, *4*, 129–145.

Cressie, N.A.C. (1993). *Statistics for spatial sata*. New York: Wiley.

Desmond, A.F. (1997). Optimal estimating functions, quasi-likelihood and statistical modeling. *Journal of Statistical Planning and Inference*, *60*(1), 77–104.

Ebbes, P., Huang, Z., Rangaswamy, A. (2016). Sampling designs for recovering local and global characteristics of social networks. *International Journal of Research in Marketing*, *33*(3), 578–599.

Frenzen, J.K., & Davis, H.L. (1990). Purchasing behavior in embedded markets. *Journal of Consumer Research*, *12*, 1–12.

Godambe, V. (1960). An optimal property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, *31*(4), 1208–1211.

Goodman, L.A. (1961). Snowball sampling. *Annals of Mathematical Statistics*, *32*(1), 148–170.

Hartmann, W.R. (2010). Demand estimation with social interactions and the implications for targeted marketing. *Marketing Science*, *29*(4), 585–601.

Hartmann, W.R., Manchanda, P., Nair, H., Bothner, M., Dodds, P., Godes, D., Hosanagar, K., Tucker, C. (2008). Modeling social interactions: identification, empirical methods, and policy implications. *Marketing Letters*, *19*(3), 287–304.

Henry, P.C. (2005). Social class, market situation, and consumers' metaphors of (dis)empowerment. *Journal of Consumer Research*, *31*, 766–778.

Holme, P., & Kim, B.J. (2002). Growing scale-free networks with tunable clustering. *Physical Review*, *65*(2), 1–4.

Katona, Z., & Sarvary, M. (2007). Network formation and the structure of the commercial world wide web. *Marketing Science*, *27*(5), 764–778.

Katona, Z., Zubcsek, P., Sarvary, M. (2011). Network effects and personal influences: the diffusion of an online social network. *Journal of Marketing Research*, *48*(3), 425–443.

Kelejian, H., & Prucha, I. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, *157*, 53–67.

Krapivsky, P.L., & Redner, S. (2001). Organization of growing random networks. *Physical Review*, *63*(6), 1–14.

LeSage, J.P., & Pace, R.K. (2007). A matrix exponential spatial specification. *Journal of Econometrics*, *140*(1), 190–214.

Liang, G., & Yu, B. (2003). Maximum pseudo likelihood estimation in network tomography. *IEEE Transactions on Signal Processing*, *51*, 2043–2053.

Lindsay, B. (1988). Composite likelihood methods. *Contemporary Mathematics*, *80*, 220–239.

Nair, H.S., Manchanda, P., Bhatia, T. (2010). Asymmetric social interactions in physician prescription behavior: the role of opinion leaders. *Journal of Marketing Research*, *47*, 883–895.

Nam, S., Manchanda, P., Chintagunta, P. (2010). The effect of signal quality and contiguous word of mouth on customer acquisition for a video-on-demand service. *Marketing Science*, *29*, 690–700, 779, 781.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, *70*(349), 120–126.

Pace, R.K., & Barry, R. (1997). Quick computation of regressions with a spatially autoregressive dependent variable. *Geographical Analysis*, *29*(3), 232–247.

Pace, R.K., & LeSage, J. (2004). Chebyshev approximation of log-determinants of spatial weight matrices. *Computational Statistics & Data Analysis*, *45*(2), 179–196.

Pace, R.K., & Zou, D. (2000). Closed-form maximum likelihood estimates of nearest neighbor spatial dependence. *Geographical Analysis*, *32*(1), 154–172.

Petersen, K.B., & Pedersen, M.S. (2008). The matrix cookbook. *Technical University of Denmark*, *7*, 15.

Renard, D., Molenberghs, G., Geys, H. (2004). A pairwise likelihood approach to estimation in multilevel probit models. *Computational Statistics and Data Analysis*, *44*, 649–667.

Salganik, M.J., & Heckathorn, D.D. (2004). Sampling and estimation in hidden populations using respondent-driven sampling. *Sociological Methodology*, *34*(1), 193–239.

Saramaki, J., & Kaski, K. (2004). Scale-free networks generated by random walkers. *Physica A: Statistical Mechanics and its Applications*, *341*, 80–86.

Schein, A.I., & Ungar, L.H. (2007). Active learning for logistic regression: an evaluation. *Machine Learning*, *68*, 235–265.

Settles, B. (2010). Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin, Madison.

Smirnov, O., & Anselin, L. (2001). Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach. *Computational Statistics & Data Analysis*, *35*(3), 301–319.

Stein, M.L., Chi, Z., Welty, L.J. (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society, Series B*, *66*(2), 275–296.

Stewart, G.W. (1998). *Matrix Algorithms* (Vol. 1). Philadelphia: SIAM.

Tepper, K. (1994). The role of labeling processes in elderly consumers' responses to age segmentation cues. *Journal of Consumer Research*, *20*, 503–519.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 267–288.

Trusov, M., Bodapati, A., Bucklin, R.E. (2010). Determining influential users in internet social networks. *Journal of Marketing Research*, *47*(4), 643–658.

Varin, C., Reid, N., Firth, D. (2011). An overview of the composite likelihood methods. *Statistica Sinica*, *21*, 5–42.

Van Den Bulte, C., & Wuyts, S. (2007). *Social networks and marketing*. Cambridge: Marketing Science Institute.

Vecchia, A.V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society. Series B*, *50*(2), 297–312.

Wang, J., Aribarg, A., Atchade, Y.F. (2013). Modeling choice interdependence in a social network. *Marketing Science*, *32*(6), 977–997.

Watts, D.J., & Strogatz, S.H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, *393*(6684), 440–442.

Xu, X., & Reid, N. (2011). On the robustness of maximum composite likelihood estimate. *Journal of Statistical Planning and Inference*, *141*, 3047–3054.

Yang, S., & Allenby, G.M. (2003). Modeling interdependent consumer preferences. *Journal of Marketing Research*, *40*, 282–294.

Yang, S., Narayan, V., Assael, H. (2006). Estimating the interdependence of television program viewership between spouses: a bayesian simultaneous equation model. *Marketing Science*, *25*(4), 336–349.

Zhou, J., Tu, Y., Chen, Y., Wang, H. (2017). Estimating spatial autocorrelation with sampled network data. *Journal of Business & Economic Statistics*, *35*(1), 130–138.