Loglinear model selection and human mobility

Dobra, A.; Mohammadi, A.

Publication date
2017
Document Version
Submitted manuscript

LOGLINEAR MODEL SELECTION AND HUMAN MOBILITY

ADRIAN DOBRA AND ABDOLREZA MOHAMMADI

Abstract. Methods for selecting loglinear models were among Steve Fienberg's research interests since the start of his long and fruitful career. After we dwell upon the string of papers focusing on loglinear models that can be partly attributed to Steve's contributions and influential ideas, we develop a new algorithm for selecting graphical loglinear models that is suitable for analyzing hyper-sparse contingency tables. We show how multi-way contingency tables can be used to represent patterns of human mobility. We analyze a dataset of geolocated tweets from South Africa that comprises 46 million latitude/longitude locations of 476,601 Twitter users that is summarized as a contingency table with 214 variables.
KEYWORDS: contingency tables, model selection, human mobility, graphical models, Bayesian structural learning, birth-death processes, pseudo-likelihood

Contents

1. Introduction 1
2. Research on human mobility 3
3. Modeling human mobility 3
4. Description of the geolocated Twitter data 4
5. Bayesian structural learning in graphical loglinear models 6
5.1. Bayesian structural learning via birth-death processes 6
5.2. Posterior estimation via sampling in continuous time 10
5.3. Birth and death rates with the marginal pseudo-likelihood 10
5.4. Speeding up the BDMCMC algorithm 11
6. Simulation Study 12
7. Analysis of the Geolocated Twitter Data 13
8. Conclusions 23
Acknowledgment 24
References 27

1. Introduction

Steve Fienberg was one of the founders of modern multivariate categorical data analysis. In two of the books he wrote early in his career [9, 31] he laid out key notation, definitions, modeling techniques, and also open research directions for building approaches for analyzing contingency tables. More than forty years ago, he argued that interactions of various orders among categorical variables are of great interest – a fact that is now recognized in the literature from several fields (e.g., biology, social sciences, public health, transportation research). Hierarchical loglinear models that represent log expected cell counts as sums of main effects of variables cross-classified in a table, and interactions of two, three or more of these variables are well suited to capture complex multivariate patterns of dependencies. The selection of the interaction structure in hierarchical loglinear models was a problem Steve discussed in considerable length in [9, Chapter 9], [31, Chapter 4], and also in several papers he subsequently published later on in his career.

Date: November 8, 2017.

[30] laid out one of the first strategies for hierarchical loglinear model determination which is based on partitioning the Pearson or the likelihood-ratio goodness-of-fit statistics into several additive parts. Steve's approach starts with a hierarchy of models, and a significance level. Interactions are sequentially added or deleted based on a series of tests that correspond with the partitioned components of the most complex models. The model search stops when the difference between consecutive models is significant. Steve properly recognized that a good model building strategy must walk the fine line between goodness-of-fit and parsimony, that is, including more interactions to obtain a better fit of the data, and leaving fewer interactions in the model to create simpler representations of the association structure. However, this early method for loglinear model selection can compare only models that are nested (i.e., a simpler model is obtained from a more complex one by deleting interactions), and can be successfully used for datasets that involve no more than 5 variables.

Due in part to Steve's early contributions and ideas, several approaches to selection of loglinear models have started to emerge [28, 1, 88], but these methods turned out to be quite ineffective even for contingency tables with 7 variables. One bottleneck is due to the exponential increase in the number of possible hierarchical loglinear models: while there are 7580 models with 5 variables, there are about 5.6×10^{22} models with 8 variables [16]. Moreover, contingency tables that involve a large number of variables are sparse and their non-zero counts are imbalanced. That is, almost all the counts in large tables are zero; most of their positive counts are small (1, 2 or 3), and there are always a few counts that are quite large. Sparsity and imbalance give rise to severe difficulties when performing model selection due to the invalidation of the asymptotic approximations to the null distribution of the generalized likelihood-ratio test statistic, or the non-existence of the maximum likelihood estimates [32, 33].

The Bayesian paradigm avoids some of these issues through the specification of prior distributions for model parameters [15]. [16] represents a key contribution that proposed a Markov chain Monte Carlo (MCMC) algorithm to identify loglinear models with high posterior probability. Other notable papers develop various stochastic search schemes for discrete data [49, 50, 51, 78, 17, 22]. These methods are known to work well for datasets with no more than 8 variables. Another approach for Bayesian model selection in contingency tables is called copula Gaussian graphical models [20], and it has successfully been used to analyze a 16-dimensional table. More recently, ultra-sparse high-dimensional contingency tables have been analyzed using probabilistic tensor factorizations induced through a Dirichlet process (DP) mixture model of product multinomial distributions [26, 12, 8, 45]. These papers present simulation studies and real-world data examples that involve up to 50 categorical variables.

Penalized likelihood methods for categorical data have focused on Markov random fields for binary variables [40, 69]. [86] show that higher-order interactions and variables with three or more categories can be modeled by introducing additional binary variables in the model specification. Such claims have never been tested on known examples; from a theoretical perspective, there is no proof that the extension of the work of [40] or [69] to general multi-way tables preserves the hierarchical structure of loglinear parameters, or yields consistent parameter and model estimates. The group lasso estimator for loglinear models [63], despite having desirable theoretical properties, does not provide guarantees that the hierarchical structure of interaction terms is preserved.

In this paper we introduce a Bayesian framework for loglinear model determination that is suitable for the analysis of a contingency table with 214 variables. Our method determines graphical loglinear models that are a special type of hierarchical loglinear models. Our key application comes from human mobility. In this context, multivariate categorical data capture the movement of individuals across multiple geographical areas. In Section 2 we discuss the relevance of massive unsolicited geolocated data for human mobility research, and in Section 3 we explain the role of loglinear models in modeling human movement. In Section 4 we describe our collection process of a geolocated Twitter dataset from South Africa; these data are subsequently transformed in the 214 dimensional contingency table we analyze in Section 7. Our modeling framework is presented in Section 5. In Section 6 we provide information

about the efficiency of our proposed method in a simulation study. In Section 8 we give some concluding comments.

2. Research on human mobility

Human mobility, or movement over short or long distances for short or long periods of time, is an important yet under-studied phenomenon in the social and demographic sciences. Migration processes represent a special case of human mobility that involve movements over longer periods of time and over longer distances. The impact of migration on human well-being, macro-social, political, and economic organization is a hot topic in the current literature [24, 27, 39, 54, 55, 56, 57, 75, 76, 80, 81, 82, 85, 89]. Similar advances in understanding human mobility have been hindered by difficulties in recording and measuring how humans move on a minute and detailed scale. A notable exception is the relatively rich literature focusing on urban mobility and transportation studies. But much of this literature relies on travel surveys which are expensive to collect, have small sample sizes and limited spatial and temporal scales, are updated infrequently, and suffer from recall bias [11, 77, 91]. Until recently, studies of mobility could not benefit from large scale data to widely address how differentials in mobility influence other outcomes. This is quite problematic given that mobility is likely a fundamental factor in behavior and macro-level social change, with potential associations with key issues that face human societies today, including spread of infectious diseases, responses to armed conflict and natural disasters, health behaviors and outcomes, economic, social, and political well-being, and migration.

Massive unsolicited geolocated data from mobile phones have recently become available for the study of human mobility. Such data are continuously collected by social media websites, search engines and wireless-service providers [5]. Every time a person makes a voice call, sends a text message, goes online or makes posts through a social media service from their mobile phone, a record is generated with information about the time and day, duration and type of communication, as well as positional information. This could be the exact latitude and longitude of the mobile phone, or an identifier of the cellular tower that handled the request. The approximate spatiotemporal trajectory of a mobile phone and its user can be reconstructed by linking the records associated with that phone. This exciting new type of data holds immense promise for studying human behavior with precision and accuracy on a vast scale never before possible with surveys or other data collection techniques [79, 23, 90].

User communications and check-ins through social media platforms such as Twitter generate publicly-available world-wide databases of human activity that can be readily accessed online free of charge. Recent evidence suggests that Twitter is a reliable source for examining human mobility patterns whose quality is comparable at the ecological level with mobile phone call records [44]. The cultural role of Twitter which serves a dual role as both a microblog and a social network, is evidenced by the Library of Congress' decision to store a permanent, daily updated archive of the site from its first tweet. Social media offers location sharing services whose growing popularity generate digital traces that can be located in space and time. Each day, Twitter records 7 million tweets with explicit geolocation (latitude and longitude) information from mobile devices with GPS sensors [64] that represent about 1.6% of the total number of tweets [47]. The geographic information from geolocated tweets (geotweets) reveals the locations of human settlements and transportation networks [47]. As the number of smartphone users continues to rise around the world, especially in low income countries, the potential of geolocated social media data to improve our knowledge of human geography will constantly grow. These are the data we collect and analyze in this paper – see Sections 4 and 7.

3. Modeling human mobility

The majority of the literature on human mobility is concerned with Lèvy flights models and with Markov process models. Let us assume that traveling patterns are observed with respect to p distinct areas or locations $\{1, 2, \ldots, p\}$. Denote by N_{ij} the number of individuals that

traveled from location i to location j in a given time interval, and by P_{ij} the probability that a random individual will travel to location j given that they are currently at location i. A class of stochastic process models called Lèvy flights [10] is one of the most popular way of modeling human mobility, or to model its limits [36]. This model represents the probability of traveling a distance d as a power law: $\mathsf{P}(d) \propto d^{-(1+\beta)}$, where $\beta < 2$ is a diffusion parameter. The Lèvy flight model says that traveling a shorter distance is more likely than traveling longer distances, but long-distance travel can still occur even if it is rare. While this assumption is reasonable, the model implies that P_{ij} depends exclusively on d_{ij} – the distance between locations i and j. This represents a serious limitation since it implies that traveling to destinations that are located at the same distance from an origin is equally likely. A more recent contribution [38] builds on multiplicative factor models from social network analysis [41] to improve the Lèvy flights model which lacks the ability to quantify the desirability of certain travel locations. They propose a model in which P_{ij} depends of a function $f(d_{ij}, \tau)$ of distance d_{ij} and of location-specific latent factors $\mathbf{u}_i \in \mathbb{R}^q$ and $\mathbf{v}_j \in \mathbb{R}^q$: $\mathsf{P}_{ij} \propto \exp\left(f(d_{ij}, \tau) + \mathbf{u}_i^T \mathbf{v}_j\right)$, where $\mathbf{u}_i^T \mathbf{v}_j$ represent the affinity of locations i and j. Inference for this latent factor model is performed based on its log-likelihood that is proportional to $-\sum_{i,j} N_{ij} \log P_{ij}$.

Both the Lèvy flights models [10] and the multiplicative latent factor models [38] are based on the crude assumption that human travel can be seen as a Markov process in which the probability of traveling to a location depends only on the origin of the trip's segment, and does not depend on previous locations visited. However, individuals are likely to travel repeatedly across multiple locations in a given period of time. Markov process models break mobility trajectories that involve multiple locations into pairs of consecutive locations, and, by doing so, loose key dependencies that are induced by multiple locations being visited by the same individuals in the reference time frame.

Loglinear models also have a long tradition in the human mobility literature, specifically, to estimate flows of migration by origin, destination, age, sex and other categorical sociodemographic variables such as economic activity group [70, 74, 71]. Migration flows are represented as origin-destination migration flow tables. These are square tables in which the rows and columns correspond with places, regions, aggregation of places or countries of interest. The (i, j) cell contains a count of the number of individuals that left from region i and moved to region j over the course of a specified time frame. The inclusion of other categorical variables lead to higher-dimensional migration flow tables. Modeling these tables involves spatial interaction loglinear models of the form [70]:

$$\log(\lambda_{ijk}) = \log(\alpha_i) + \log(\beta_j) + \log(m_{ijk}),$$

where λ_{ijk} is the expected migration flow from origin i to destination j for a combination of levels k of one, two or more additional categorical variables, and m_{ijk} is auxiliary information on the migration flow. The characteristics of the origin i and the destination j are represented through the parameters α_i and β_j. However, migration flow tables cannot capture the movement of those individuals that live in more than three regions during the time frame of observation. An example individual that left from region 1 to move to region 2, then moved again to region 3, would contribute with a count of 1 in the $(1, 2)$ and $(2, 3)$ cells of the resulting migration flow table. But, the link between these two counts will be lost. For this reason, loglinear models that estimate migration flows suffer from the same shortcoming as Markov process models.

4. Description of the geolocated Twitter data

In this article we analyze a large-scale database of geolocated tweets from South Africa. This sub-Saharan country has been selected due to its high rates of internal and external migration caused by violent internal conflicts, war, political and economical instability, poverty, racial discrimination. Statistics South Africa reports that, in October 2016, 3.5 million travelers passed through South Africa's ports of entry. They were made up of $925,796$ South African residents and 2.6 million foreign travelers. In this country, human mobility is known to be one of the major contributors to the spread of infectious diseases (HIV, tuberculosis, malaria) [79, 19].

Our geotweets database was put together in a two step process. First, geolocated tweets posted in South Africa between September 2011 and September 2016 have been obtained directly from Twitter through GNIP, a reseller of social data owned by Twitter, as part of a no-cost collaborative research agreement between the University of Washington and Twitter. A geotweet is classified to have been posted inside a country based on a country code field derived by GNIP from the latitude and longitude of the tweet. Second, we used the Twitter REST APIs [84] to obtain geolocated tweets of the 476,601 users whose geotweets have been captured in the first step. The REST APIs allow access to up to 3,200 most recent geotweets in each user's timeline irrespective of the time when they have been posted, or the location they have been posted from. For this purpose, we used a customized version of the `smappR` R package [73]. The second data collection step took place continuously between January and December 2016. During this period, the most recent geotweets of each of the 476,601 users have been retrieved at least twice per month.

The total number of unique geotweets acquired in both steps is 46,210,370. The actual tweets have been discarded after we extracted tuples of the form `<user key, time of the posting, latitude, longitude, ...>` from the rich content of each tweet. To assure privacy protection, each Twitter user is identified by a randomly generated key which replaces their Twitter identifier. Additional filtering steps were performed to eliminate any non-human activity (e.g., Twitter robots) or any geotweets with coding errors. We emphasize that this database comprises only public information which can be viewed online, and replicated using the APIs provided by Twitter or downloaded directly from a third party provider of social media data such as GNIP.

For each of the 476,601 users, we determined their country of residence. We estimated the amount of time a user spent in a country they visited as the cumulative periods of time between consecutive geotweets posted in that country. A user's country of residence was defined as the country with the largest amount of time spent among all the countries this user tweeted from. Our method for identifying the users' country of residence has certain limitations. First, it is possible that a user could choose to post geotweets only when they are away from their country of residence. Second, it is also possible that our two step process of collecting geotweets might have missed relevant time intervals in which a user tweeted from their country of residence. However, after carefully examining the spatial patterns of geotweets with respect to the estimated countries of residence, we are confident that our method of determination worked fine for a large percentage of users. Based on this information, we classified 41,049 (8.62%) of the 476,601 users as visitors of South Africa, and the rest as locals, that is, individuals that most likely see South Africa as their home country.

We subsequently mapped the geotweets into the 213 district municipalities of South Africa – see Figure 1. This allowed us to determine, for each user, the municipalities they were present and absent during the five years data collection time frame. Here we assume that absence from a municipality is implied by the user not posting any geotweets within its boundaries. These presence/absence patterns together with the Local (yes/no) variable define a 214 dimensional binary contingency table. This table is hyper-sparse: only 55015 cells contain positive counts (the logarithm of the percentage of non-zero counts is -132.813). Among the 55015 non-zero counts, there are 46175 (83.93%) counts of 1, 3439 (6.25%) counts of 2, 1411 (2.56%) counts of 3, 747 (1.36%) counts of 4, and 476 (0.87%) counts of 5. The top five largest counts are 58929, 42781, 28731, 28197 and 22313, and represent the number of users that were locals to South Africa and posted geotweets only from one of following five metropolitan municipalities: Johannesburg (JHB, Gauteng), Cape Town (CPT, Western Cape), Tshwane (TSH, Gauteng), eThekwini (ETH, KwaZulu-Natal), and Ekurhuleni (EKU, Gauteng), respectively. The sixth largest count is 9568, and represents the number of users that were locals to South Africa, and posted geotweets from two metropolitan municipalities, Johannesburg (JHB) and Ekurhuleni (EKU). The seventh largest count count is 8464, and represents the number of users that were visitors (non-locals) to South Africa, and posted geotweets only from Johannesburg (JHB). In the next section we present our framework for determining the multivariate patterns of interactions among these 214 binary variables.

FIGURE 1. Administrative divisions of South Africa: nine provinces divided into 52 metropolitan and district municipalities (dashed lines) that are further divided into 213 local municipalities (dashed lines). The motorways, trunk, and primary roads (source: [66]) of the country are also shown.

5. Bayesian structural learning in graphical loglinear models

An undirected interaction graph $G = (V, E)$ ($V = \{1, ..., p\}$ are vertices, and $E \subset V \times V$ are edges) is defined for a hierarchical loglinear model $\mathcal{H}$ that involves p categorical variables $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ as follows. A vertex $i \in V$ of G corresponds with variable X_i. An edge $e = (i, j)$ appears in G if and only if the variables X_i and X_j appear together in an interaction term of $\mathcal{H}$. Model $\mathcal{H}$ is graphical if the subsets of V that are the vertices of the complete subgraphs of G that are maximal with respect to inclusion, are also maximal interaction terms in $\mathcal{H}$ [46]. In this case, the absence of an edge between vertices i and j in G means that X_i and X_j are conditional independent given the remaining variables $X_{V\setminus\{i,j\}}$. For this reason, the interaction graph G of a graphical loglinear model is called a conditional independence graph. This graph also has a predictive interpretation. Denote by $\mathsf{nbd}_G(i) = \{j \in V : (i, j) \in E\}$ the neighbors of vertex i in G. Then X_i is conditionally independent of $X_{V\setminus(\mathsf{nbd}_G(i)\cup\{i\})}$ given $X_{\mathsf{nbd}_G(i)}$ which implies that, given G, a mean squared optimal prediction of X_i can be made from the neighboring variables $X_{\mathsf{nbd}_G(i)}$.

We focus on the structural learning problem [43, 25] which aims to estimate the structure of G (i.e., which edges are present or absent in E) from the available data $\mathbf{x} = (x^{(1)}, ..., x^{(n)})$. In a Bayesian framework, we explore the posterior distribution of G conditional on the data $\mathbf{x}$, i.e.

$$\mathsf{P}(G \mid \mathbf{x}) = \frac{\mathsf{P}(G)\mathsf{P}(\mathbf{x} \mid G)}{\sum_{G\in\mathcal{G}_p} \mathsf{P}(G)\mathsf{P}(\mathbf{x} \mid G)}, \tag{1}$$

where $\mathsf{P}(G)$ is a prior distribution on the space $\mathcal{G}_p$ of undirected graphs with p vertices, and $\mathsf{P}(\mathbf{x} \mid G)$ is the marginal likelihood of the data conditional on G [43]. Identifying the graphs with the largest posterior probability (1) is a complex problem because the number of undirected graphs $2^{\binom{p}{2}}$ in $\mathcal{G}_p$ becomes large very fast as p increases. For example, for $p = 20$, the number of undirected graphs in $\mathcal{G}_p$ exceeds 10^{70}. In this paper we introduce a computationally efficient search algorithm that takes advantage of parallelizable local computations at the vertex level that moves fast towards regions with high posterior probabilities (1).

5.1. **Bayesian structural learning via birth-death processes.** To efficiently explore the graph space $\mathcal{G}_p$, [61] developed the birth-death Markov chain Monte Carlo (BDMCMC) algorithm. This is a trans-dimensional MCMC algorithm, and represents an alternative to the well

known reversible jump MCMC algorithm [37]. The version of BDMCMC presented in [61] was developed specifically for Gaussian graphical models. In this section we give a general formulation for sampling from any distributions on a space of graphs $\mathcal{G}_p$.

The BDMCMC algorithm is based on a continuous time birth-death Markov process [68]. Its underlying sampling scheme traverses $\mathcal{G}_p$ by adding and removing edges corresponding to the birth and death events. Given that the process is at state $G = (V, E)$, we define the birth and death events as independent Poisson processes as follows:

Birth event – each edge $e \in \overline{E}$ where $\overline{E} = \{e \in V \times V : e \notin E\}$, is born independently of other edges that do not belong to G as a Poisson process with rate $B_e(G)$. If the birth of edge e occurs, the process jumps to $G^{+e} = (V, E \cup \{e\})$ which is a graph with one edge more than G.

Death event – each edge $e \in E$ dies independently of other edges that belong to G as a Poisson process with rate $D_e(G)$. If the death of edge e occurs, the process jumps to $G^{-e} = (V, E \backslash \{e\})$ which is a graph with one edge less than G.

This birth-death Markov process is a jump process with intensity $\alpha(G) = \sum_{e \in \overline{E}} B_e(G) + \sum_{e \in E} D_e(G)$. Its waiting time to the next jump follows an exponential distribution with expectation $1/\alpha(G)$. The birth and death probabilities are

$$\mathsf{P}(\text{birth of edge } e) \propto B_e(G), \quad \text{for } e \in \overline{E}, \tag{2}$$

$$\mathsf{P}(\text{death of edge } e) \propto D_e(G), \quad \text{for } e \in E. \tag{3}$$

The following theorem provides sufficient conditions on the birth and death rates to guarantee that the corresponding process on $\mathcal{G}_p$ has stationary distribution (1).

Theorem 5.1. *The birth-death process defined by the birth and death probabilities (2) and (3) has the stationary distribution* $\mathsf{P}(G \mid \mathbf{x})$ *given in (1), if the following detailed balance condition is satisfied:*

$$B_e(G)\mathsf{P}(G \mid \mathbf{x}) = D_e(G^{+e})\mathsf{P}(G^{+e} \mid \mathbf{x}), \tag{4}$$

where $e \in \overline{E}$, $G = (V, E)$, *and* $G^{+e} = (V, E \cup \{e\})$.

Proof. We take advantage of the theory on general classes of Markov birth-death processes from [68, Section 7 and 8]. This class of Markov jump processes evolve in jumps which occur a finite number of times in any finite time interval. These jumps are of two types: (i) *birth* in which a single point is added, and the process jumps to a state that contains the additional point; and (ii) *death* in which one of the points in the current state is deleted, and the process jumps to a state with one less point. [68] shows that the process converges to a unique stationary distribution provided that the detailed balance conditions hold.

To define the balance conditions for our process, assume that at a given time, the process is in a graph state $G = (V, E)$ with $\theta_G \in \Theta_G$ as a vector of parameters. The process is characterized by the *birth rates* $B_e(G, \theta_G)$ for each $e \in \overline{E}$, the *death rates* $D_e(G, \theta_G)$ for each $e \in E$, and the birth and death *transition kernels* $K^{(G)}_{B_e}(\theta_G; \cdot)$ and $K^{(G)}_{D_e}(\theta_G; \cdot)$. Birth and death events occur as independent Poisson processes with rates $B_e(G, \theta_G)$ and $D_e(G, \theta_G)$ respectively. Given the birth of $e \in \overline{E}$ occurs, the probability that the following jump leads to a point in $\mathsf{F} \in \Theta_{G^{+e}}$ is

$$K^{(G)}_{B_e}(\theta_G; \mathsf{F}) = \frac{B_e(G, \theta_G)}{B(G, \theta_G)} \int_{\theta_e : \theta_G \cup \theta_e \in \mathsf{F}} b_e(\theta_e; \theta_G)\, \mathrm{d}\theta_e,$$

in which $B(G, \theta_G) = \sum_{e \in \overline{E}} B_e(G, \theta_G)$. Similarly, given the death of $e \in E$ occurs, the probability that the following jump leads to a point in $\mathsf{F} \in \Theta_{G^{-e}}$ is

$$K^{(G)}_{D_e}(\theta_G; \mathsf{F}) = \frac{D_e(G, \theta_G)}{D(G, \theta_G)} \mathsf{I}(\theta_{G^{-e}} \in \mathsf{F}), \tag{5}$$

in which $D(G, \theta_G) = \sum_{e \in E} D_e(G, \theta_G)$.

For this birth-death process, $\mathsf{P}(G, \theta_G \mid \mathbf{x})$ satisfies detailed balance conditions if

$$\int_{\mathsf{F}} B(G, \theta_G)\mathsf{P}(G, \theta_G \mid \mathbf{x})\, \mathrm{d}\theta_G = \tag{6}$$
$$\sum_{e \in \overline{E}} \int_{\Theta_{G^{+e}}} D(G^{+e}, \theta_{G^{+e}}) K_{D_e}^{(G^{+e})}(\theta_{G^{+e}}; \mathsf{F})\mathsf{P}(G^{+e}, \theta_{G^{+e}} \mid \mathbf{x})\, \mathrm{d}\theta_{G^{+e}},$$

and

$$\int_{\mathsf{F}} D(G, \theta_G)\mathsf{P}(G, \theta_G \mid \mathbf{x})\, \mathrm{d}\theta_G = \tag{7}$$
$$\sum_{e \in E} \int_{\Theta_{G^{-e}}} B(G^{-e}, \theta_{G^{-e}}) K_{B_e}^{(G^{-e})}(\theta_{G^{-e}}; \mathsf{F})\mathsf{P}(G^{-e}, \theta_{G^{-e}} \mid \mathbf{x})\, \mathrm{d}\theta_{G^{-e}},$$

where $\mathsf{F} \subset \Theta_G$.

We check the first part of the detailed balance conditions (6) as follows

$$\begin{aligned}
LHS &= \int_{\mathsf{F}} B(G, \theta_G)\mathsf{P}(G, \theta_G \mid \mathbf{x})\, \mathrm{d}\theta_G \\
&= \int_{\Theta_G} \mathsf{I}(\theta_G \in \mathsf{F}) B(G, \theta_G)\mathsf{P}(G, \theta_G \mid \mathbf{x})\, \mathrm{d}\theta_G \\
&= \int_{\Theta_G} \mathsf{I}(\theta_G \in \mathsf{F}) \sum_{e \in \overline{E}} B_e(G, \theta_G)\mathsf{P}(G, \theta_G \mid \mathbf{x})\, \mathrm{d}\theta_G \\
&= \sum_{e \in \overline{E}} \int_{\Theta_G} \mathsf{I}(\theta_G \in \mathsf{F}) B_e(G, \theta_G)\mathsf{P}(G, \theta_G \mid \mathbf{x})\, \mathrm{d}\theta_G \\
&= \sum_{e \in \overline{E}} \int_{\Theta_G} \mathsf{I}(\theta_G \in \mathsf{F}) B_e(G, \theta_G)\mathsf{P}(G, \theta_G \mid \mathbf{x}) \left[\int_{\Theta_e} b_e(\theta_e; \theta_G)\, \mathrm{d}\theta_e \right] \mathrm{d}\theta_G \\
&\qquad\qquad [b_e \text{ must integrate to } 1] \\
&= \sum_{e \in \overline{E}} \int_{\Theta_G} \int_{\Theta_e} \mathsf{I}(\theta_G \in \mathsf{F}) B_e(G, \theta_G)\mathsf{P}(G, \theta_G \mid \mathbf{x}) b_e(\theta_e; \theta_G)\, \mathrm{d}\theta_e\, \mathrm{d}\theta_G.
\end{aligned}$$

$$\begin{aligned}
RHS &= \sum_{e \in \overline{E}} \int_{\Theta_{G^{+e}}} D(G^{+e}, \theta_{G^{+e}}) K_{D_e}^{(G^{+e})}(\theta_{G^{+e}}; \mathsf{F})\mathsf{P}(G^{+e}, \theta_{G^{+e}} \mid \mathbf{x})\, \mathrm{d}\theta_{G^{+e}} \\
&\qquad\qquad [\text{equation (5) }] \\
&= \sum_{e \in \overline{E}} \int_{\Theta_{G^{+e}}} \mathsf{I}(\theta_G \in \mathsf{F}) D_e(G^{+e}, \theta_{G^{+e}})\mathsf{P}(G^{+e}, \theta_{G^{+e}} \mid \mathbf{x})\, \mathrm{d}\theta_{G^{+e}}.
\end{aligned}$$

Therefore we have LHS=RHS provided that

$$B_e(G, \theta_G)\mathsf{P}(G, \theta_G \mid \mathbf{x}) b_e(\theta_e; \theta_G) = D_e(G^{+e}, \theta_{G^{+e}})\mathsf{P}(G^{+e}, \theta_{G^{+e}} \mid \mathbf{x}).$$

Now, by integrating over $\theta_{G^{+e}} = \theta_G \cup \theta_e$ and knowing that the function $b_e(\theta_e; \theta_G)$ must integrate to 1 over Θ_e, we have

$$B_e(G)\mathsf{P}(G \mid \mathbf{x}) = D_e(G^{+e})\mathsf{P}(G^{+e} \mid \mathbf{x}),$$

which is the expression (4) in Theorem 5.1. In a similar way, it can be shown that the remaining part of the detailed balance conditions (7) also hold. □

Based on Theorem 5.1, we define the birth and death rates of the BDMCMC algorithm as a function of the ratio of the corresponding posterior probabilities to optimize the convergence

speed:

$$B_e(G) = \min\left\{\frac{\mathsf{P}(G^{+e} \mid \mathbf{x})}{\mathsf{P}(G \mid \mathbf{x})}, 1\right\}, \quad \text{for each} \quad e \in \overline{E},$$

$$D_e(G) = \min\left\{\frac{\mathsf{P}(G^{-e} \mid \mathbf{x})}{\mathsf{P}(G \mid \mathbf{x})}, 1\right\}, \quad \text{for each} \quad e \in E.$$

We show the birth and death rates as follows

$$R_e(G) = \min\left\{\frac{\mathsf{P}(G^* \mid \mathbf{x})}{\mathsf{P}(G \mid \mathbf{x})}, 1\right\}, \quad \text{for each} \quad e \in \{E \cup \overline{E}\}, \tag{8}$$

where for the birth of edge e we take $G^* = (V, E \cup \{e\})$, and for the death of edge e we take $G^* = (V, E \backslash \{e\})$.

Algorithm 1 provides the pseudo-code for the BDMCMC algorithm which samples from the posterior distribution (1) on $\mathcal{G}_p$ by using the above birth-death mechanism. In Section 5.3 we explain how to efficiently compute the ratio of posterior probabilities in the birth and death rates (8) for multivariate discrete data by using the marginal pseudo-likelihood approach [67].

Algorithm 1 . BDMCMC algorithm for undirected graphical models

Input: A graph $G = (V, E)$ with p nodes and data $\mathbf{x}$
  for N iterations **do**
    for all the possible edges in parallel **do**
      Calculate the birth and death rates in (8),
    end for
    Calculate the waiting time for G by $W(G) = \frac{1}{\sum_{e \in \overline{E}} B_e(G) + \sum_{e \in E} D_e(G)}$
    Update G based on birth/death probabilities in (2) and (3)
  end for
Output: Samples from the posterior distribution (1).

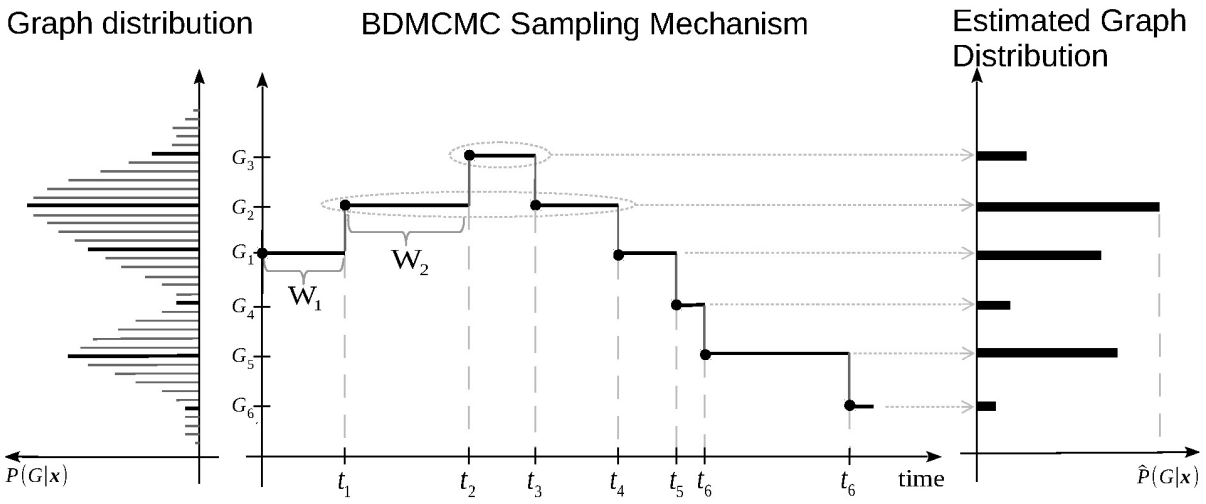


FIGURE 2. The left and right panels show the true and estimated posterior distribution (1) on the space the graphs. The middle panel shows an example output from an application of Algorithm 1 where $\{W_1, W_2, ...\}$ denote waiting times, and $\{t_1, t_2, ...\}$ denote jumping times.

5.2. **Posterior estimation via sampling in continuous time.** Figure 1 illustrates how the output of Algorithm 1 can be used to estimate posterior quantities of interest. The output consists of a set of sampled graphs, a set of waiting times $\{W_1, W_2, ...\}$, and a set of jumping times $\{t_1, t_2, ...\}$. Based on the Rao-Blackwellized estimator [13], the estimated posterior probability of each sampled graph is proportional to the expectation of length of the holding time in that graph which is estimated as the sum of the waiting times in that graph. The posterior inclusion probability of an edge $e \in V \times V$ is estimated by

$$\widehat{\mathsf{P}}(\text{edge } e \mid \mathbf{x}) = \frac{\sum_{t=1}^{N} \mathsf{I}(e \in G^{(t)}) W(G^{(t)})}{\sum_{t=1}^{N} W(G^{(t)})}, \tag{9}$$

where N denotes the number of iterations, $\mathsf{I}(e \in G^{(t)})$ denotes an indicator function: $\mathsf{I}(e \in G^{(t)}) = 1$ if $e \in G^{(t)}$, and 0 otherwise.

5.3. **Birth and death rates with the marginal pseudo-likelihood.** We assume that the observed random variables $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ are categorical, with each variable X_i taking values in a discrete set $\mathcal{X}_i = \{1, 2, ..., r_i\}$. The determination of the birth and death rates (8) involves the marginal likelihood conditional on a graph $G \in \mathcal{G}_p$:

$$\mathsf{P}(\mathbf{x} \mid G) = \int_{\Theta_G} \mathsf{P}(\mathbf{x} \mid \theta_G, G) \mathsf{P}(\theta_G \mid G) \, \mathrm{d}\theta_G, \tag{10}$$

where $\theta_G \in \Theta_G$ are the parameters of a multivariate model associated with G, $\mathsf{P}(\theta_G \mid G)$ is prior for θ_G, and $\mathsf{P}(\mathbf{x} \mid \theta_G, G)$ is the full likelihood function. However, the exact calculation of the marginal likelihood $\mathsf{P}(\mathbf{x} \mid G)$ is possible only for decomposable graphs G which represent a small fraction of the graphs in $\mathcal{G}_p$ [53]. Numerical approximations for the marginal likelihood for arbitrary undirected graphs have been developed [22], but their application is computationally expensive even for datasets that involve $p < 20$ variables. This high computational effort renders them inapplicable for the Twitter mobility data described in Section 4 with $p = 214$ observed variables.

A computationally cheaper alternative comes from approximating the full likelihood $\mathsf{P}(\mathbf{x} \mid \theta_G, G)$ with the pseudo-likelihood [6, 7] which is the product of the full conditionals of the random variables $\mathbf{X}$ given their neighbors in G:

$$\mathsf{P}_{pl}(\mathbf{x} \mid \theta_G^{pl}, G) = \prod_{d=1}^{n} \prod_{i=1}^{p} \mathsf{P}(X_i = x_i^{(d)} \mid \mathbf{X}_{\mathsf{nbd}_G(i)} = \mathbf{x}_{\mathsf{nbd}_G(i)}^{(d)}, \theta_{i,G}^{pl}). \tag{11}$$

We denote $\mathcal{X}_{\mathsf{nbd}_G(i)} = \times_{j \in \mathsf{nbd}_G(i)} \mathcal{X}_j$, $\theta_{i,+l} = \{\theta_{i,kl} : k \in \mathcal{X}_i\}$, and $\theta_{i,G}^{pl} = \{\theta_{i,+l} : l \in \mathcal{X}_{\mathsf{nbd}_G(i)}\}$. In (11), $\theta_G^{pl} = \times_{i=1}^{p} \theta_{i,G}^{pl} \in \Theta_G^{pl}$ are the set of parameters of the full conditionals

$$\mathsf{P}(X_i = k \mid \mathbf{X}_{\mathsf{nbd}_G(i)} = l) = \theta_{i,kl}, \quad \text{for } i = 1, ..., p,$$

where $k \in \mathcal{X}_i$, $l \in \mathcal{X}_{\mathsf{nbd}_G(i)}$. Thus, the pseudo-likelihood (11) can be written as:

$$\mathsf{P}_{pl}(\mathbf{x} \mid \theta_G^{pl}, G) = \prod_{i=1}^{p} \prod_{k \in \mathcal{X}_i} \prod_{l \in \mathcal{X}_{\mathsf{nbd}_G(i)}} \theta_{i,kl}^{n_{i,kl}}, \tag{12}$$

where $n_{i,kl}$ represents the number of samples $x^{(d)}$, $d = 1, 2, \ldots, n$, such that $x_i^{(d)} = k$ and $x_{\mathsf{nbd}_G(i)}^{(d)} = l$.

For computational convenience, we assume that the set of parameters $\theta_{i,G}^{pl}$ and $\theta_{i',G}^{pl}$ associated with the full conditionals of X_i and $X_{i'}$, $i \neq i'$ are independent. This assumption is certainly not consistent with the assumption that the full conditionals are derived from the same full joint distribution of $\mathbf{X}$. Nevertheless, the approximation of the full likelihood with the pseudo-likelihood (10) is based on the same premise [6, 7]. We also assume that, within the same full conditional associated with the variable X_i, the parameters $\theta_{i,+l}$ and $\theta_{i,+l'}$ associated with the

different levels l and l' of the variables $\mathbf{X}_{\mathsf{nbd}_G(i)}$ are independent [67]. We impose a prior for θ_G^{pl} that factorizes according to these two assumptions:

$$\mathsf{P}(\theta_G^{pl}) = \prod_{i=1}^{p} \mathsf{P}(\theta_{i,G}) = \prod_{i=1}^{p} \prod_{l \in \mathcal{X}_{\mathsf{nbd}_G(i)}} \mathsf{P}(\theta_{i,+l}). \tag{13}$$

Furthermore, we impose a Dirichlet prior on the conditional probabilities of X_i at level l of $\mathbf{X}_{\mathsf{nbd}_G(i)}$:

$$\theta_{i,+l} \sim \mathsf{Dir}(\alpha_{i,1l}, \ldots, \alpha_{i,r_i l}). \tag{14}$$

From (12), (13), and (14), it follows that the marginal pseudo-likelihood is [67]:

$$\mathsf{P}_{pl}(\mathbf{x} \mid G) = \prod_{i=1}^{p} \mathsf{P}(\mathbf{x}_i \mid \mathbf{x}_{\mathsf{nbd}_G(i)}), \tag{15}$$

with

$$\mathsf{P}(\mathbf{x}_i \mid \mathbf{x}_{\mathsf{nbd}_G(i)}) = \prod_{l \in \mathcal{X}_{\mathsf{nbd}_G(i)}} \frac{\Gamma(\alpha_{i,+l})}{\Gamma(\alpha_{i,+l} + n_{i,+l})} \prod_{k \in \mathcal{X}_i} \frac{\Gamma(\alpha_{i,kl} + n_{i,kl})}{\Gamma(\alpha_{i,kl})}, \tag{16}$$

where $\alpha_{i,+l} = \sum_{k \in \mathcal{X}_i} \alpha_{i,kl}$ and $n_{i,+l} = \sum_{k \in \mathcal{X}_i} n_{i,kl}$.

A prior on the space of graphs $\mathcal{G}_p$ that encourages sparsity by penalizing for the inclusion of additional edges in the graph $G = (V, E)$ is [43]:

$$\mathsf{P}(G) \propto \left(\frac{\beta}{1-\beta}\right)^{|E|} = \left(\prod_{i=1}^{p} \left(\frac{\beta}{1-\beta}\right)^{|\mathsf{nbd}_G(i)|}\right)^{1/2}, \tag{17}$$

where $\beta \in (0,1)$ is set to a small value, e.g. $\beta = 1/\binom{p}{2}$. While other priors on $\mathcal{G}_p$ are available [21], the prior (17) can be decomposed as the product of independent priors for the p full conditionals given G such that the probability of inclusion of a vertex in each of these conditionals is equal with β as shown in (17).

The marginal posterior distribution on $\mathcal{G}_p$ based on the marginal pseudo-likelihood (15) and the prior on $\mathcal{G}_p$ (17) is

$$\mathsf{P}_{pl}(G \mid \mathbf{x}) \propto \mathsf{P}_{pl}(\mathbf{x} \mid G)\mathsf{P}(G) = \prod_{i=1}^{p} \mathsf{P}(\mathbf{x}_i \mid \mathbf{x}_{\mathsf{nbd}_G(i)}) \left(\frac{\beta}{1-\beta}\right)^{\frac{|\mathsf{nbd}_G(i)|}{2}}. \tag{18}$$

The birth and death rates in (8) based on the marginal pseudo-likelihood for an edge $e = (i,j) \in V \times V$ are calculated from

$$\widehat{R}_e(G) = \min\left\{ \frac{\mathsf{P}(\mathbf{x}_i \mid \mathbf{x}_{\mathsf{nbd}_{G^*}(i)})}{\mathsf{P}(\mathbf{x}_i \mid \mathbf{x}_{\mathsf{nbd}_G(i)})} \frac{\mathsf{P}(\mathbf{x}_j \mid \mathbf{x}_{\mathsf{nbd}_{G^*}(j)})}{\mathsf{P}(\mathbf{x}_j \mid \mathbf{x}_{\mathsf{nbd}_G(j)})} \left(\frac{\beta}{1-\beta}\right)^{\delta}, 1 \right\},$$

where for the birth of edge e we take $G^* = (V, E \cup \{e\})$, $\delta = 1$, and for the death of edge e we take $G^* = (V, E \backslash \{e\})$, $\delta = -1$.

5.4. **Speeding up the BDMCMC algorithm.** The key bottleneck of the BDMCMC algorithm is the computation at every iteration of the birth and death rates (8) for all the $p(p-1)/2$ possible edges. Fortunately, the rates associated with one edge can be calculated independently of the rates associated with the other edges, and can be performed in parallel which represents a first key computational improvement. We implemented parallel computations of the birth and death rates in the current version of the `R` package `BDgraph` [62] using `OpenMP` [65]. Most code in this package is written in `C++` and interfaced in `R`.

A second key computational improvement is possible when the marginal likelihood is replaced with the marginal pseudo-likelihood as detailed in Section 5.3. Since at each step of the BDMCMC algorithm one edge $e = (i,j)$ is selected for addition or removal, only the marginal likelihood (16) of the full conditionals of the two vertices i and j will change. Thus, we need to recalculate the $(p-1)+(p-1)-1 = 2p-3$ rates that correspond with these two vertices. The

remaining rates will stay the same. As such, at each iteration we update $2p-3$ rates instead of $p(p-1)/2$ rates. This represents a huge computational saving especially for graphs with many vertices. For example, for the Twitter mobility data we analyze in Section 7, we look at graphs with $p = 214$ vertices. Instead of computing 22791 rates at each step of the BDMCMC algorithm, we only need to determine 422 rates which means that a single edge update can be done approximately 54 times faster.

A third key computational improvement comes from allowing multiple edge updates at each iteration. The vast majority of the MCMC and stochastic search algorithms that have been developed in the Bayesian graphical models literature are based on adding or removing one edge at each iteration [43, 48, 72, 48, 87, 60, 61, 58, 14]. These single edge updates are in part responsible for making these structural learning algorithms quite slow for datasets that comprise a larger number of variables p. Multiple birth-death sampling approaches have been used to address image processing problems that aim to detect a configuration of objects from a digital image, and have been found to outperform the convergence speed of competing reversible jump MCMC algorithms [18, 29, 35, 34].

By following this idea, it is possible to transform Algorithm 1 into a multiple birth-death MCMC algorithm based on a multiple birth-death process. At each iteration, after computing and ranking the birth and deaths rates (8), we select not one but a fixed number $N_0 \geqslant 2$ of edges to be added or removed from the graph. By doing so, N_0 edges are updated at no computational cost compared to a single edge update. Through multiple edge updates which we have also implemented in the `R` package `BDgraph` [62], the BDMCMC algorithm can quickly move to regions with high posterior probability in the graph space $\mathcal{G}_p$. The ability to move towards high posterior probability graphs in a small number of iterations is especially important in applications in which the ratio between the number of samples available and the number of variables is small. However, performing multiple edge updates at each iteration of the BDMCMC algorithm does not have any theoretical guarantees related to sampling from the correct target posterior distribution (1). Multiple edge updates in can be performed for a reduced number of iterations to identify several graphs that have higher posterior probabilities compared to the empty graph, the full graph or a random graph sampled from $\mathcal{G}_p$. These graphs can be subsequently used as starting points for Algorithm 1 with single edge updates.

6. Simulation Study

We investigate the performance of the BDMCMC algorithm in recovering the graph structure from categorical data by comparing it to the hill-climbing (HC) algorithm proposed by [67]. While the BDMCMC algorithm samples from the marginal posterior distribution (18), the HC algorithm solves the optimization problem $\max\{\mathsf{P}_{pl}(G \mid \mathbf{x}) : G \in \mathcal{G}_p\}$ using a method that involves two phases.

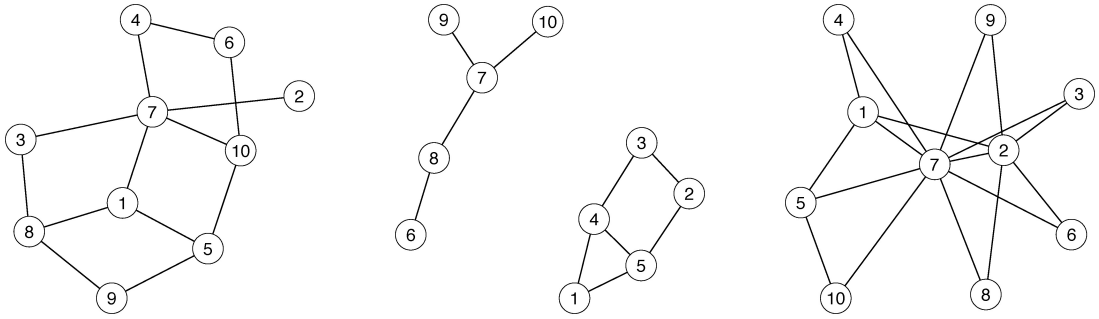
We consider three types of graphs (see Figure 3):

1. **Random:** A graph in which edges were randomly generated from the prior (17) with $\beta = 0.4$.
2. **Cluster:** A graph with two clusters (connected components) each with $p = 5$ vertices. The edges in both clusters were randomly generated from the prior (17) with $\beta = 0.6$.
3. **Scale-free:** A graph sampled from a power-law degree distribution with the Barabási-Albert algorithm [2].

We also consider graphs with $p = 20$ vertices that have two connected components with 10 vertices and the same edge structure of type "Random", "Cluster", or "Scale-free". We simulated binary contingency tables with $p \in \{10, 20\}$ variables that comprise $n \in \{200, 500, 1000\}$ samples from random graphs of these three types. We repeated the simulation experiment that involves the generation of 18 contingency tables 50 times. We performed all computations with the `R` package `BDgraph` [62, 59]. For each contingency table we generated, we run the BDMCMC and the HC algorithms using a uniform prior on the graph space $\mathcal{G}_p$ (the equivalent of setting

$\beta = 0.5$ in (17)) starting from the empty graph. The BDMCMC algorithm was run for $100,000$ iterations. The first $60,000$ iterations were discarded as burn-in.

We estimated the structure of the true graph based on model averaging [52] of the graphs sampled by the BDMCMC algorithm. We calculate the posterior inclusion probabilities of edges (9), and determine the median graph whose edges have posterior inclusion probabilities greater than 0.5. The structure of the true graph was estimated with the HC algorithm based on the "and" and the "or" criteria in the first phase of the algorithm [67].



(a) Random (b) Cluster (c) Scale-free

FIGURE 3. Example graphs with $p = 10$ vertices used in the simulation study from Section 6.

We evaluate the performance of the two algorithms in recovering the structure of the true graphs using the F_1-score measure [3],

$$F_1\text{-score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}, \tag{19}$$

and the Structure Hamming distance (SHD) [83],

$$\text{`SHD} = \text{FP} + \text{FN}, \tag{20}$$

where TP, TN, FP and FN are the number of true positives, true negatives, false positives and false negatives, respectively. The values of the F_1-score range between 0 and 1, and the values of the SHD are positive. A better performance in recovering the true graph is associated with larger values of the F_1-score, and with smaller values of the SHD.

The results are summarized in Table 1. For most simulation experiments, the BDMCMC algorithm has an advantage over the HC algorithm especially for the F_1-score. ROC curves showing the performance of the BDMCMC algorithm are presented in Figure 4.

7. Analysis of the Geolocated Twitter Data

We come back to the $p = 214$ dimensional binary contingency table constructed from geotweets that was described in Section 4. We use the BDMCMC algorithm to sample graphs from the marginal posterior distribution (18) on $\mathcal{G}_{214}$. We employ the prior (17) with $\beta = 1/\binom{214}{2} = 4.388 \times 10^{-5}$. Under this prior, the expected number of edges is 1, thus sparser graphs receive larger prior probabilities compared to denser graphs. We performed all computations on a cluster with 7 compute nodes, each with 48 Intel Xeon 2.6 GHz cores with a Linux operating system.

First, we want to gain some understanding of the ability of the BDMCMC algorithm to move towards graphs with large posterior probabilities in $\mathcal{G}_{214}$. To this end, we sample 20 graphs $\{G_i\}_{i=1}^{20}$ having increasing number of edges: G_i has a number of edges randomly sampled from $(200(i-1), 200i)$. The resulting set of graphs ranges from most sparse (G_1) to most dense (G_{20}). Starting from each graph G_i, $i = 1, \ldots, 20$, we ran the BDMCMC algorithm for $10,000$

TABLE 1. Results from the simulation study from Section 6. Binary tables with $p \in \{10, 20\}$ variables and $n \in \{200, 500, 1000\}$ samples were generated from graphical models defined by three types of graphs: "Random", "Cluster", and "Scale-free". This table reports means and standard deviations of the F_1-score (19) and SHD (20) measures for the accuracy in recovering the structure of the true graph across 50 replicate binary tables generated for every combination of p, n and graph type. The best performing models in terms of the F_1-score and SHD are shown in boldface.

		F_1-score			SHD		
p	n	BDMCMC	HC(or)	HC(and)	BDMCMC	HC(or)	HC(and)
				Random			
	200	**0.7**(0.12)	0.68(0.11)	0.57(0.11)	**8.2**(2.9)	8.7(2.9)	10.6(3)
10	500	**0.8**(0.1)	0.78(0.1)	0.7(0.1)	**5.8**(2.5)	6.2(2.6)	8(2.6)
	1000	**0.87**(0.08)	0.86(0.08)	0.8(0.09)	**3.9**(2.2)	4(2.1)	5.6(2.2)
	200	**0.7**(0.08)	0.69(0.07)	0.58(0.09)	17.5(4.6)	**17.3**(3.9)	21.5(5.7)
20	500	**0.8**(0.08)	0.79(0.08)	0.7(0.09)	**11.9**(4.2)	12.5(4.5)	16.6(4.4)
	1000	**0.85**(0.07)	**0.85**(0.07)	0.78(0.08)	**8.9**(4)	9.3(3.8)	12.6(4)
				Cluster			
	200	**0.76**(0.13)	0.75(0.13)	0.66(0.14)	**4.5**(2)	4.6(1.8)	5.8(2)
10	500	**0.86**(0.1)	0.83(0.13)	0.75(0.15)	**2.7**(1.7)	3.3(2.1)	4.6(2.4)
	1000	**0.91**(0.09)	0.9(0.09)	0.87(0.09)	**1.7**(1)	1.8(1.1)	2.4(1.3)
	200	0.69(0.13)	**0.71**(0.11)	0.63(0.14)	14.8(10.2)	**11.7**(4.2)	13.5(4.9)
20	500	**0.86**(0.08)	0.84(0.08)	0.76(0.11)	**5.9**(2.8)	6.7(3.2)	9.5(4.1)
	1000	**0.93**(0.06)	0.92(0.06)	0.87(0.08)	**3.3**(2.7)	3.6(2.7)	5.9(2.9)
				Scale-free			
	200	**0.67**(0.12)	0.66(0.1)	0.56(0.1)	**8.5**(2.6)	**8.5**(2.1)	10(1.9)
10	500	**0.73**(0.11)	**0.73**(0.11)	0.62(0.12)	**6.9**(2.4)	**6.9**(2.2)	8.8(2.2)
	1000	0.8(0.07)	**0.81**(0.08)	0.7(0.1)	5.3(1.7)	**5.2**(1.9)	7.4(1.9)
	200	**0.63**(0.13)	**0.63**(0.13)	0.53(0.09)	21.3(12.7)	**19.1**(5.6)	21.4(3.7)
20	500	**0.74**(0.08)	**0.74**(0.09)	0.63(0.08)	14(3.6)	**13.8**(3.8)	17.7(3.2)
	1000	**0.78**(0.09)	**0.78**(0.09)	0.7(0.09)	11.8(4.3)	**11.4**(4.3)	14.7(3.9)

iterations. Figures 5 and 6 show the sum of the estimated posterior edge inclusion probabilities and the number of edges included in the sampled graphs against iteration number. After $7,000$ iterations in each of the 20 runs, the BDMCMC algorithm seems to have reached the same neighborhood of graphs. Thus, although the number of graphs in $\mathcal{G}_{214}$ is extremely large ($\approx 10^{6861}$), the BDMCMC algorithm seems to be very efficient in identifying graphs with high posterior probability.

Next, we ran the BDMCMC algorithm for $400,000$ iterations using parallel calculations of the birth and death rates from a starting graph sampled from the prior (17) on $\mathcal{G}_{214}$. The first $200,000$ iterations were discarded as burn-in. Figures 7 and 8 show the BDMCMC algorithm seems to have reached convergence in less than $10,000$ iterations.

We estimate the posterior inclusion probabilities (9) of the $\binom{214}{2} = 22,791$ edges. We show a heatmap of the matrix of the estimated posterior edge inclusion probabilities in Figure 9. Most of the estimated posterior edge inclusion probabilities are zero: $21,138$ (92.78%). A number of 12, 5 and 7 edges have estimated posterior inclusion probabilities in $(0, 0.5)$, $[0.5, 0.9)$ and $[0.9, 0.1)$, respectively. The remaining 1522 (6.65%) have estimated posterior inclusion probabilities equal

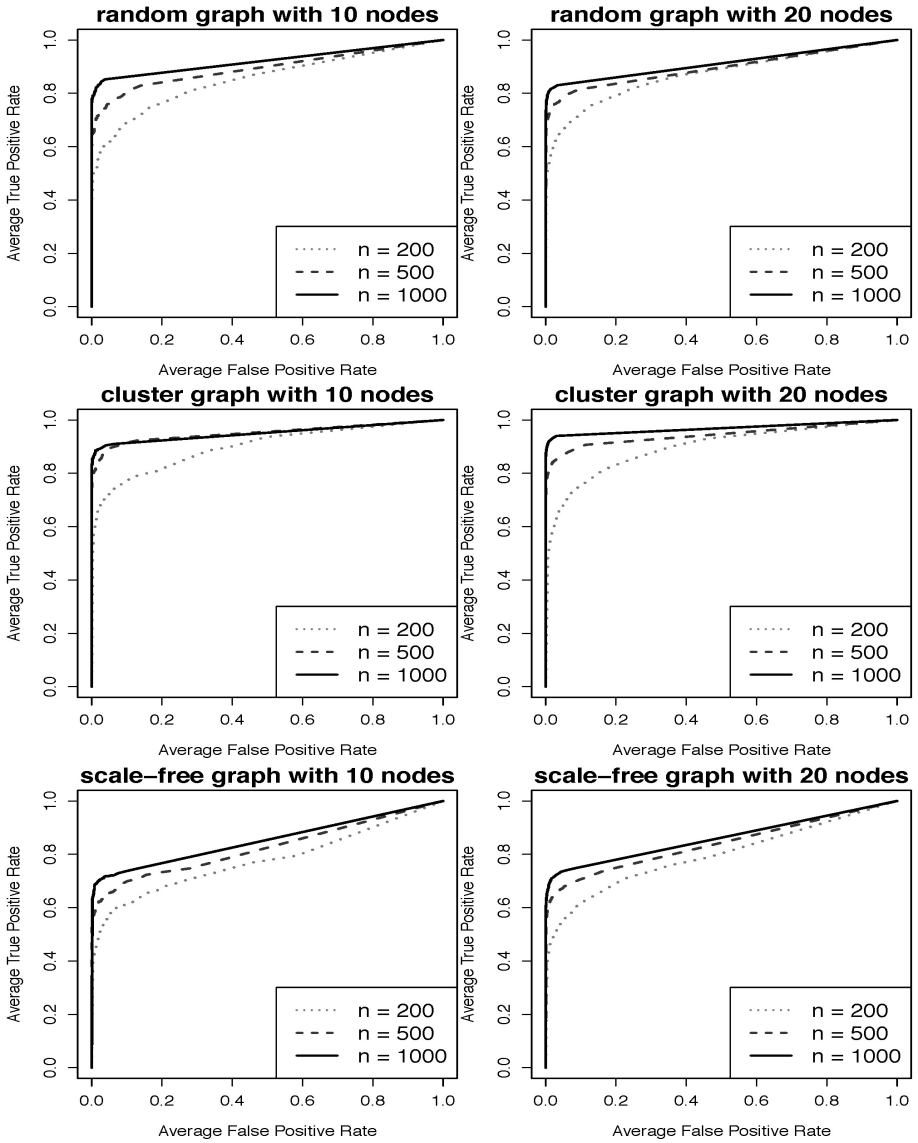


FIGURE 4. ROC curves showing the performance of the BDMCMC algorithm in recovering the structure of the true graph in the simulation study from Section 6. Binary tables with $p \in \{10, 20\}$ variables and $n \in \{200, 500, 1000\}$ samples were simulated from graphs of three types ("Random", "Cluster", and "Scale-free").

to 1. We use the median graph which includes the 1534 edges with estimated posterior inclusion probabilities greater than 0.5 as our estimate of the conditional independence graph. Henceforth we refer to this graph as the South Africa (SA) Twitter graph.

A rendering of the SA Twitter graph is presented in Figure 10. The 213 district municipalities of South Africa are denoted by their identifiers – see Table 5. This table provides the identifier, complete name, province to which it belongs, area, population size, and density for each municipality. The 214-th vertex of this graph is associated with the Local (yes/no) variable. We explore the SA Twitter graph using four centrality measures [42] that capture the extent to which a vertex is connected to other vertices, and occupies a central position in the structure of the network: (i) *degree* counts the number of edges that originate from a given vertex; (ii) *closeness* measures how close is a vertex from each one of the other vertices; (iii) *betweenness* finds vertices that connect other vertices (i.e., belong to the shortest paths connecting pairs of vertices); and (iv) *page rank* defines more central vertices based on a voting process which allocates votes to a vertex based on other connected vertices, and it is determined through an

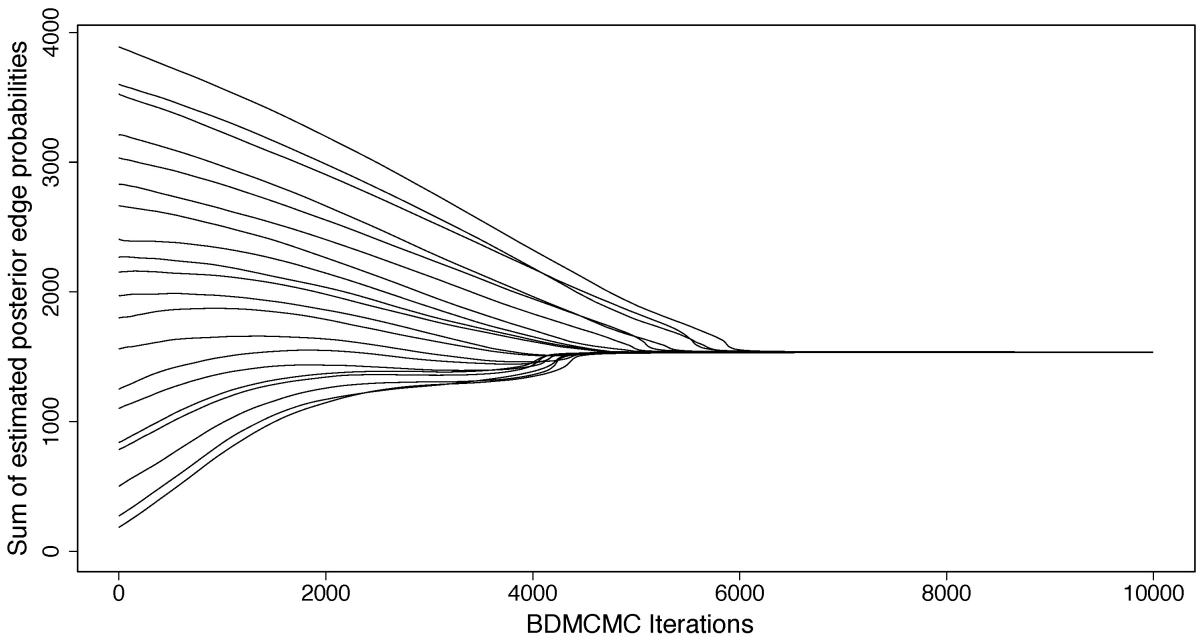


FIGURE 5. Convergence plot of the BDMCMC algorithm for the Twitter data introduced in Section 4 and analyzed in Section 7. The plot shows the sum of the estimated posterior edge inclusion probabilities (y axis) against iteration number (x axis) from 20 runs of the BDMCMC algorithm starting from 20 different graphs sampled from graph space $\mathcal{G}_{214}$ with different number of edges.

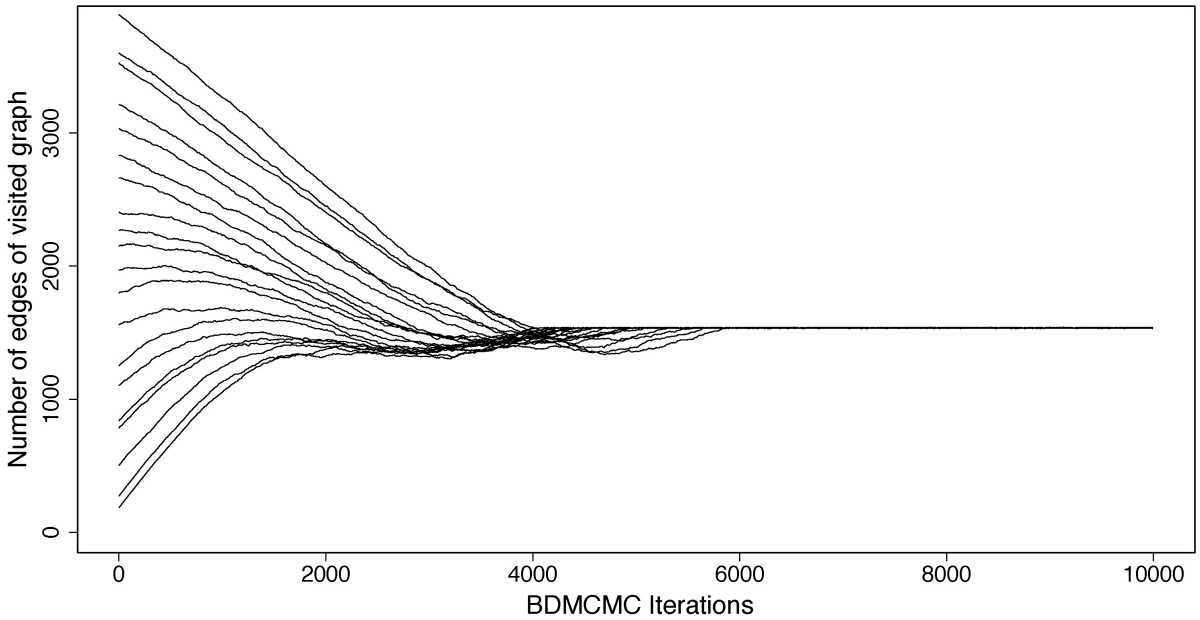


FIGURE 6. Convergence plot of the BDMCMC algorithm for the Twitter data introduced in Section 4 and analyzed in Section 7. The plot shows the number of edges included in the sampled graphs (y axis) against iteration number (x axis) from 20 runs of the BDMCMC algorithm starting from 20 different graphs sampled from graph space $\mathcal{G}_{214}$ with different number of edges.

iterative algorithm. Barplots of the largest 10 values of each of the four centrality measures are presented in Figures 11, 12, 13, and 14.

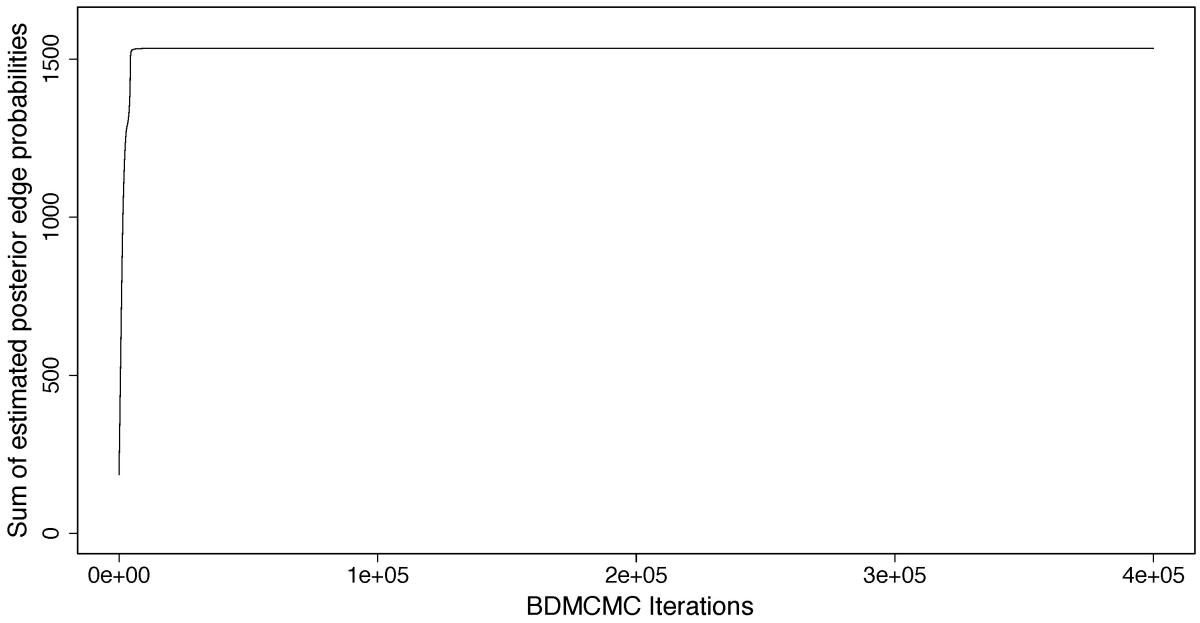


FIGURE 7. Convergence plot for a longer run of 400,000 iterations of the BDMCMC algorithm for the Twitter data introduced in Section 4 and analyzed in Section 7. The plot shows the sum of the estimated posterior edge inclusion probabilities (y axis) against iteration number (x axis).

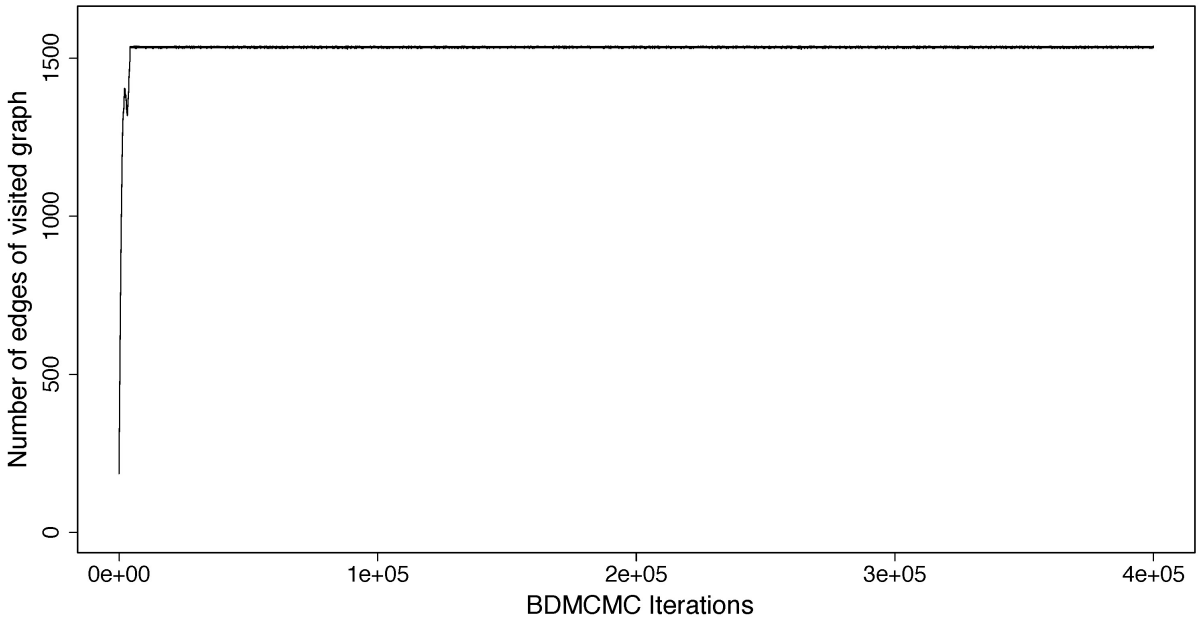


FIGURE 8. Convergence plot for a longer run of $400,000$ iterations of the BDMCMC algorithm for the Twitter data introduced in Section 4 and analyzed in Section 7. The plot shows the number of edges included in the sampled graphs (y axis) against iteration number (x axis).

For each of the four centrality measures, their top five largest values correspond with the following municipalities: Johannesburg (JHB, Gauteng), Ekurhuleni (EKU, Gauteng), Tshwane (TSH, Gauteng), eThekwini (ETH, KwaZulu-Natal), and Cape Town (CPT, Western Cape).

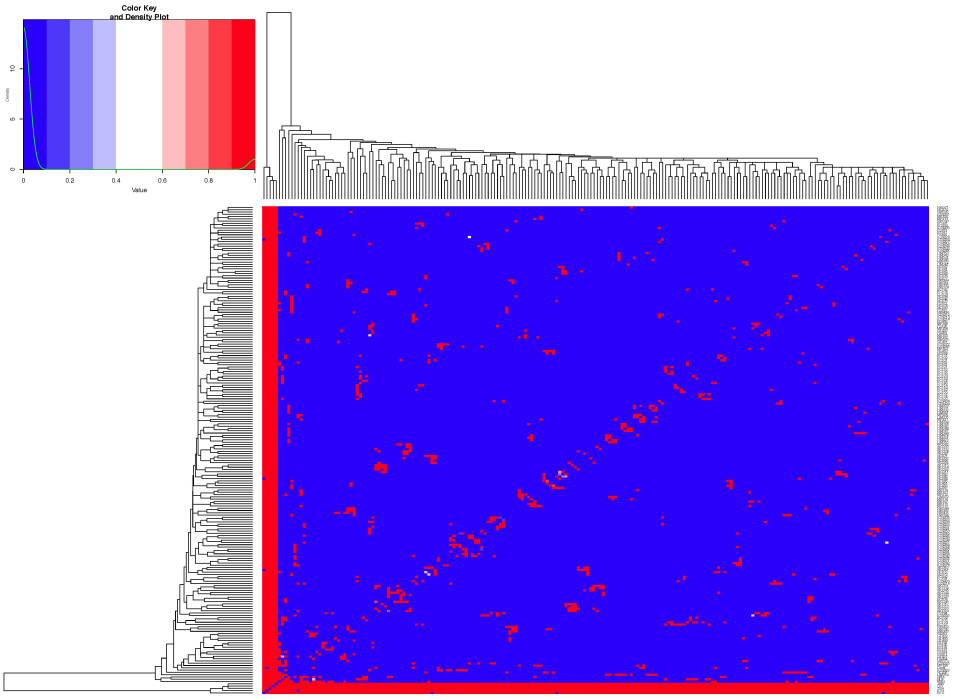


FIGURE 9. Heatmap of the 214×214 matrix of posterior inclusion probability of edges for the Twitter data introduced in Section 4 and analyzed in Section 7. The rows are shown in increasing order with respect to the row sums from top to bottom. The columns are shown in the same order from left to right. Red marks edges with posterior inclusion probability very close to 1, while blue marks edges with posterior inclusion probability close to zero. The names of the variables involved are shown in the right column.

The next five largest values also correspond with the same five municipalities for all four measures: Mangaung (MAN, Free State), Nelson Mandela Bay (NMA, Eastern Cape), Polokwane (LIM354, Limpopo), Buffalo City (BUF, Eastern Cape), and Sol Plaatjie (NC091, Northern Cape). We remark that the values of centrality measures for JHB, EKU, TSH, ETH and CPT are significantly larger than the values for MAN, NMA, LIM354, BUF and NC091. For example, the degrees for the first group are 213, 210, 213, 212, and 213, while the degrees for the second group are 25, 24, 21, 20, and 19.

As such, MAN, NMA, LIM354, BUF and NC091 are the five key hubs of the SA Twitter graph. Their geographical location is mapped in Figure 15, while Table 2 gives summary information about them. Three hubs (JHB, TSH, EKU) are located in the Johannesburg/Pretoria area which represents the region of South Africa in which more than 11 million people reside (2015 South African National Census) either permanently, or temporarily to find employment in factories or gold mines. The other two hubs are located around the cities of Cape Town and Durban which, together with Johannesburg and Pretoria, are among the largest South African cities. A great number of local and international travelers visit these five hubs for shorter or longer periods of times. Based on the predictive interpretation of the SA Twitter graph, the presence or absence of a Twitter user from one of the five hubs is predictive of the presence or absence of this user from almost all the other district municipalities. Furthermore, the presence or absence of an user from almost all the municipalities that are not hubs is predictive of their presence or absence from each of the hubs.

The vertex associated with the variable Local is not central in the structure of the SA Twitter graph. Its degree is 12, and the other centrality measures are also significantly smaller compared

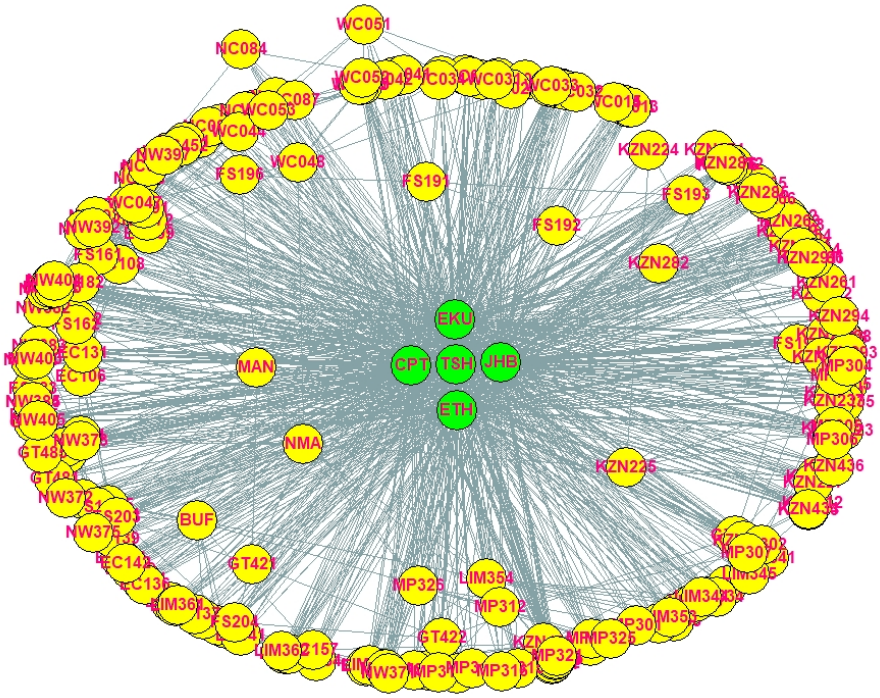


FIGURE 10. Rendering of the SA Twitter graph. The five hubs are shown in green, and the remaining vertices are shown in yellow.

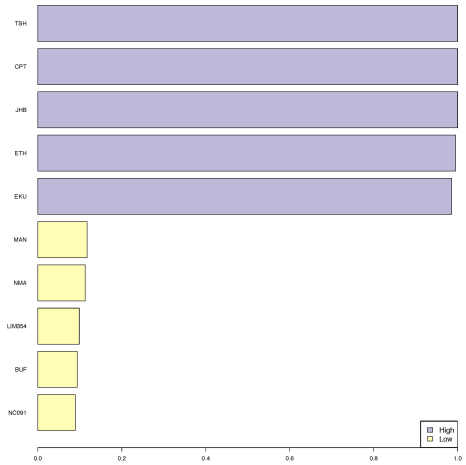


FIGURE 11. Barplot showing the top 10 municipalities with the largest degree in the SA Twitter graph. The five hubs are shown in purple.

to those of the five hubs. The 12 district municipalities that are connected with an edge with the Local vertex are mapped in Figure 17; the corresponding vertices in the SA Twitter graph are shown in Figure 16. In addition to four of the hubs (JHB, EKU, TSH, CPT), the presence or absence patterns of a Twitter user from the following municipalities are predictive of whether this user is local to South Africa: FS162 (Kopanong, Free State), KZN216 and KZN276 (Ray Nkonyeni and Big Five Hlabisa, KwaZulu-Natal), LIM335 (Maruleng, Limpopo), MP325 (Bushbuckridge, Mpumalanga), NW375 (Moses Kotane, North West), GT421 (Emfuleni, Gauteng),

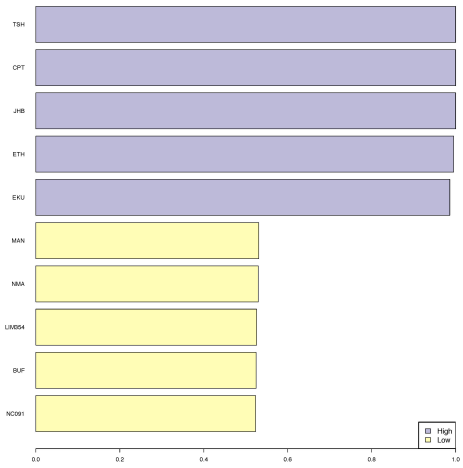


FIGURE 12. Barplot showing the top 10 municipalities with the largest closeness in the SA Twitter graph. The five hubs are shown in purple.

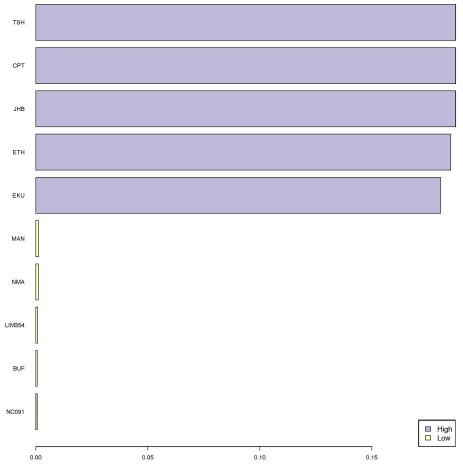


FIGURE 13. Barplot showing the top 10 municipalities with the largest betweenness in the SA Twitter graph. The five hubs are shown in purple.

TABLE 2. Summary geographic and demographic information about the five hub municipalities in the SA Twitter graph. Population data extracted from the 2016 Community Survey, Statistics South Africa. Retrieved from `https://interactive2.statssa.gov.za/webapi`.

Id.	Municipality name	Province	Area (km^2)	Population	Density
TSH	City of Tshwane	Gauteng	6,298	3,275,152	520
ETH	eThekwini	KwaZulu-Natal	2,556	3,702,231	1,448.50
CPT	City of Cape Town	Western Cape	2,446	4,005,016	1,637.60
JHB	City of Johannesburg	Gauteng	1,645	4,949,347	3,008.80
EKU	Ekurhuleni	Gauteng	1,975	3,379,104	1,710.60

and WC024 (Stellenbosch, Western Cape). It is quite interesting to examine the spatial distribution of these 12 municipalities: CPT and WC024 are adjacent municipalities around Cape

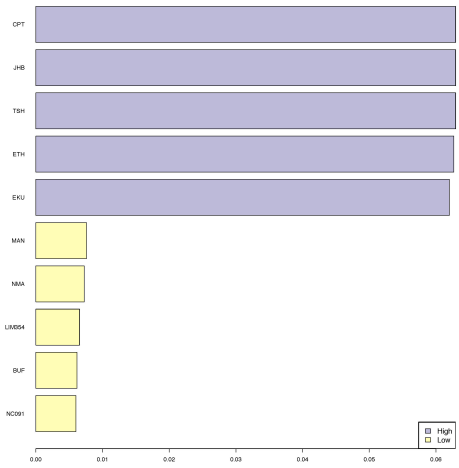


FIGURE 14. Barplot showing the top 10 municipalities with the largest page rank in the SA Twitter graph. The five hubs are shown in purple.

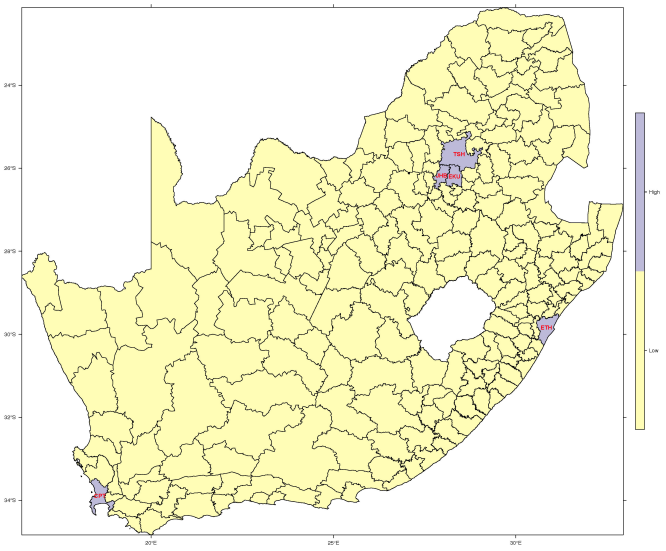


FIGURE 15. Map of South Africa showing five municipalities (purple) with the largest centrality measures in the SA Twitter graph. These are the five hubs of this graph. The same municipalities score highest with respect to degree, closeness, betweenness and page rank. The remaining 208 municipalities are shown in yellow.

Town; TSH, JHB and EKU define a spatially contiguous region in the Johannerburg/Pretoria area; while LIM335 and MP325 are adjacent municipalities at the border between South Africa and Mozambique. In the KwaZulu-Natal province, the municipalities KZN216 and KZN276 that are located to the south and to the north of the city of Durban are among the neighbors of Local, but the ETH municipality in which Durban is located is not (quite surprisingly) among the neighbors of Local. The FS162 municipality is located south of Bloemfontein – a major city in South Africa known for its mining industry. The NW375 municipality is located north west of the Johannesburg/Pretoria area, and it comprises Sun City and a major national park – both key touristic destinations.

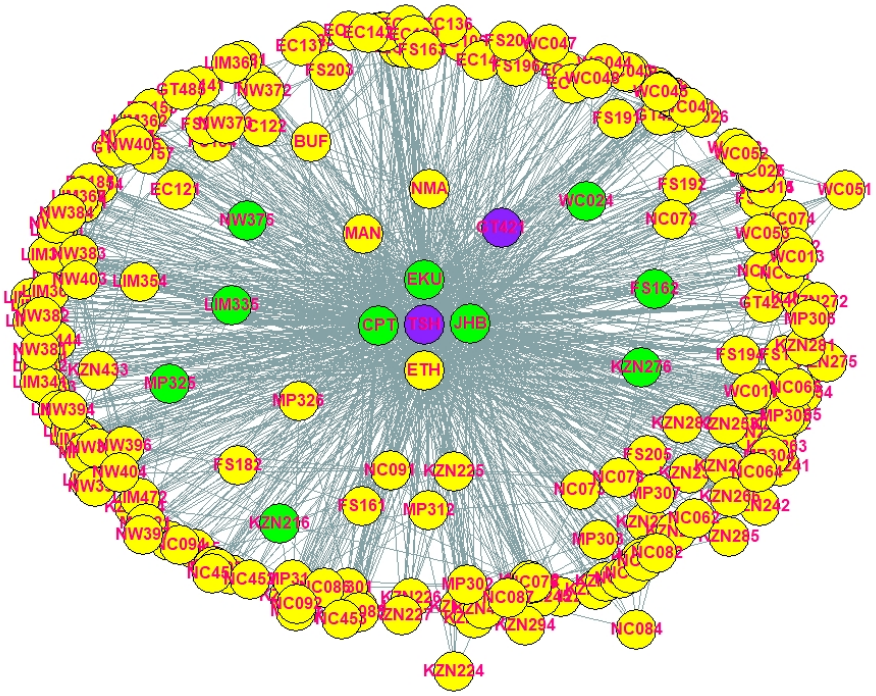


FIGURE 16. Rendering of the SA Twitter graph that shows the 12 district municipalities that are linked by an edge with the vertex associated with the Local variable. The two municipalities with an aOR greater than 1 are shown in purple, the 10 municipalities that have an aOR smaller than 1 are shown in green, while the rest of the municipalities are shown in yellow.

TABLE 3. Summary geographic and demographic information about the 10 municipalities linked by an edge with the Local (yes/no) variable in the SA Twitter graph that have an aOR smaller than 1. Population data extracted from the 2016 Community Survey, Statistics South Africa. Retrieved from `https://interactive2.statssa.gov.za/webapi`.

Id.	Municipality name	Province	Area (km^2)	Population	Density	aOR (95% CI)
FS162	Kopanong	Free State	15,645	49,999	3.2	0.259 (0.237,0.284)
KZN216	Ray Nkonyeni	KwaZulu-Natal	1,487	348,533	234.4	0.731 (0.6769839,0.790)
LIM335	Maruleng	Limpopo	3,563	99,946	28.1	0.251 (0.225,0.280)
NW375	Moses Kotane	North West	5,726	243,648	42.5	0.692 (0.640,0.748)
KZN276	Big Five Hlabisa	KwaZulu-Natal	3,466	116,622	33.6	0.482 (0.419,0.555)
WC024	Stellenbosch	Western Cape	831	173,197	208.4	0.760 (0.726,0.796)
MP325	Bushbuckridge	Mpumalanga	10,248	546,215	53.3	0.413 (0.380,0.448)
CPT	City of Cape Town	Western Cape	2,446	4,005,016	1,637.60	0.354 (0.346,0.362)
JHB	City of Johannesburg	Gauteng	1,645	4,949,347	3,008.80	0.920 (0.898,0.942)
EKU	Ekurhuleni	Gauteng	1,975	3,379,104	1,710.60	0.804 (0.781,0.827)

TABLE 4. Summary geographic and demographic information about the 2 municipalities linked by an edge with the Local (yes/no) variable in the SA Twitter graph that have an aOR greater than 1. Population data extracted from the 2016 Community Survey, Statistics South Africa. Retrieved from `https://interactive2.statssa.gov.za/webapi`.

Id.	Municipality name	Province	Area (km^2)	Population	Density	aOR (95% CI)
TSH	City of Tshwane	Gauteng	6,298	3,275,152	520	2.347 (2.260,2.437)
GT421	Emfuleni	Gauteng	966	733,445	759.3	5.414 (4.714,6.218)

We determine the effect of the presence and absence patterns of Twitter users from these 12 municipalities on the odds of being local to South Africa by fitting a logistic regression model

for the Local variable with 12 explanatory variables associated with these municipalities. The estimated adjusted odds ratios are given in Tables 3 and 4. A number of 10 municipalities have adjusted odds ratios significantly smaller than 1 at significance level $\alpha = 0.05$. Given the same presence and absence pattern in the remaining 11 municipalities, a Twitter user that posted geotweets from one of these municipalities has smaller odds of being local to South Africa compared to another Twitter user that did not post geotweets from that municipality. However, the TSH and GT421 municipalities located to the north and to the south of the Johannesburg/Pretoria area have estimated adjusted odds ratios significantly greater than 1 at significance level $\alpha = 0.05$. Given the same presence and absence pattern in the remaining 11 municipalities, the odds of being local to South Africa of a Twitter user that was present in GT421 (TSH) are 5.414 (2.347) times larger than the odds of being local to South Africa of another Twitter user that was absent from GT421 (TSH). It is known that a considerable number of Mozambicans come to work in the mines in the Johannesburg/Pretoria area for extended periods of time [4]. Their residences might be located in the GT421 and TSH municipalities where they could exceed the number of South African Twitter users.

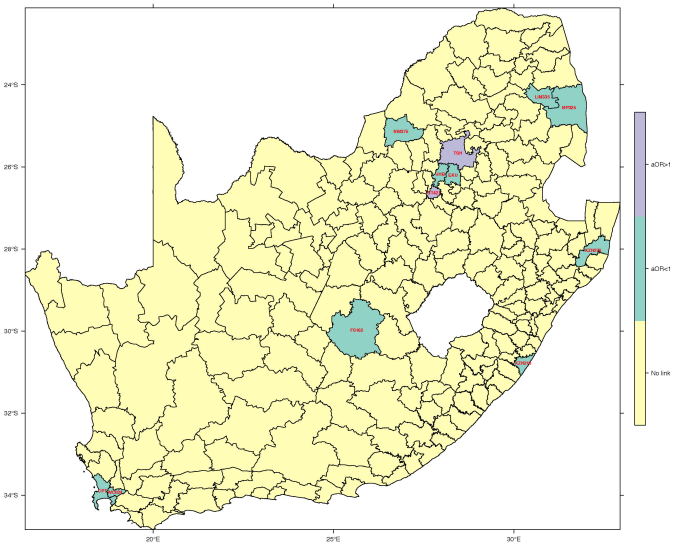


FIGURE 17. Map of South Africa showing the 12 municipalities that are linked by an edge with the vertex associated with the Local variable in the SA Twitter graph. Two of these municipalities (purple) have an adjusted odds ratio (aOR) greater than 1, while 10 of them (green) have an aOR smaller than 1. The remaining 208 municipalities are shown in yellow.

8. Conclusions

This paper makes several contributions. First, it generalizes the birth-death Markov chain Monte Carlo (BDMCMC) algorithm introduced by [61] in the context of Gaussian graphical models to general undirected graphical models. Second, based on marginal pseudo-likelihood for categorical data of [67], we show how to efficiently calculate the birth and death rates for the BDMCMC algorithm for arbitrary undirected graphs. Third, we use our methodology to analyze a 214-dimensional contingency table that captures the mobility patterns of Twitter users in South Africa. This is a dataset we collected at the University of Washington which has never been analyzed before.

The hill-climbing (HC) algorithm [67] determines graphs with high posterior probability using a greedy hill-climbing optimization algorithm. For this reason, the HC algorithm will inevitably end up in a local maximum. Which local maximum the HC algorithm will find depends on the choice of starting graph. The results of the simulation study from Section 6 were obtained by starting the HC algorithm from empty graphs. Since the true graphs were sparse, the HC algorithm recorded a good performance that was comparable with the performance of the BDMCMC algorithm. However, if we would have started the HC algorithm from random graphs that contained a larger number of edges, the HC algorithm might have been at a disadvantage. As we illustrated in Section 7, starting the BDMCMC algorithm from sparser or denser graphs led to the identification of the same neighborhood of graphs with high posterior probabilities. The BDMCMC algorithm has a key advantage over the HC algorithm in terms of its ability to visit graphs with lower posterior probability in order to escape local optima, and move towards other graphs with larger posterior probabilities.

Our applied results give an understanding of the movements of $476,601$ individuals that used geolocated tweets in South Africa between 2011 and 2016. It is true that the movements of this specific group of people cannot be considered to be representative of major flows of movement of South Africans, or of the visitors of this country. And, due to the selected locations Twitter users choose to post their tweets from, it is possible that even the travel trajectories of these individuals could be only partially captured. However, to the best of our knowledge, there is no other study on human mobility that involves a larger number of individuals in South Africa, and comprises a larger number of recorded locations (> 46 millions). While our findings must be interpreted with care from a sociodemographic perspective, the methodology we introduce in this article can be successfully applied to modeling patterns of repeated across regions movement that span entire countries, and comprise a large number of individuals. Our modeling approach is based on multi-way tables that cross-classify presence/absence patterns from regions of interest, together with other relevant categorical factors. Our framework goes beyond methods that focus exclusively on modeling flows of migration between origin and destination areas.

We showed that our version of the BDMCMC algorithm can efficiently determine conditional independence graphs for 214 categorical variables. As far as we know, this is the largest categorical dataset analyzed so far with loglinear models. These developments would not have been possible without Steve Fienberg's visionary life long work which led to the birth of a research community that spans several disciplines (social sciences, health and medical sciences, computer science, and statistics) and will without doubt generate fundamental scientific knowledge for many generations to come.

Acknowledgment

This work was supported in part by the National Science Foundation Grant DMS/MPS-1737746 to University of Washington. The authors thank Johan Pensar for providing some of the code used in the simulation study and Sven Baars for his suggestions related to parallel coding in `C++`. We also would like to thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster.

Table 5. Summary geographic and demographic information 213 South African municipalities. Population data extracted from the 2016 Community Survey, Statistics South Africa. Retrieved from `https://interactive2.statssa.gov.za/webapi`.

Id.	Municipality name	Province	Area (km^2)	Population	Density
EC101	Dr Beyers Naude	Eastern Cape	28,653	82,197	2.9
EC104	Makana	Eastern Cape	4,376	82,060	18.8
EC105	Ndlambe	Eastern Cape	1,841	63,180	34.3
EC121	Mbhashe	Eastern Cape	3,303	277,250	84
EC122	Mnquma	Eastern Cape	3,137	246,813	78.7
EC131	Inxuba Yethemba	Eastern Cape	11,663	70,493	6
EC137	Engcobo	Eastern Cape	2,484	162,014	65.2
EC141	Elundini	Eastern Cape	5,019	144,929	28.9

Table 5 – continued from previous page

Id.	Municipality name	Province	Area (km^2)	Population	Density
EC142	Senqu	Eastern Cape	7,329	140,720	19.2
EC153	Ngquza Hill	Eastern Cape	2,477	303,379	122.5
EC154	Port St Johns	Eastern Cape	1,291	166,779	129.2
EC155	Nyandeni	Eastern Cape	2,474	309,702	125.2
EC156	Mhlontlo	Eastern Cape	2,880	189,176	65.7
EC157	King Sabata Dalindyebo	Eastern Cape	3,019	488,349	161.8
EC441	Matatiele	Eastern Cape	4,352	219,447	50.4
FS161	Letsemeng	Free State	9,828	40,044	4.1
FS162	Kopanong	Free State	15,645	49,999	3.2
FS163	Mohokare	Free State	8,776	35,840	4.1
FS181	Masilonyana	Free State	6,618	62,770	9.5
FS182	Tokologo	Free State	9,326	29,149	3.1
FS183	Tswelopele	Free State	6,524	47,373	7.3
FS184	Matjhabeng	Free State	5,690	429,113	75.4
FS185	Nala	Free State	4,129	78,515	19
FS191	Setsoto	Free State	5,431	117,362	21.6
FS192	Dihlabeng	Free State	4,868	140,044	28.8
FS193	Nketoana	Free State	5,611	64,893	11.6
FS194	Maluti a Phofung	Free State	4,338	353,452	81.5
FS195	Phumelela	Free State	8,196	50,054	6.1
FS196	Mantsopa	Free State	4,291	53,525	12.5
FS201	Moqhaka	Free State	7,925	154,732	19.5
FS203	Ngwathe	Free State	7,055	118,907	16.9
FS204	Metsimaholo	Free State	1,717	163,564	95.3
FS205	Mafube	Free State	3,971	57,574	14.5
GT481	Mogale City	Gauteng	1,342	383,864	286
GT485	Rand West City	Gauteng	1,115	265,887	238.5
GT484	Merafong City	Gauteng	1,631	188,843	115.8
TSH	City of Tshwane	Gauteng	6,298	3,275,152	520
KZN212	Umdoni	KwaZulu-Natal	994	144,551	145.5
KZN214	uMuziwabantu	KwaZulu-Natal	1,089	108,576	99.7
KZN216	Ray Nkonyeni	KwaZulu-Natal	1,487	348,533	234.4
KZN238	Alfred Duma	KwaZulu-Natal	3,764	356,274	94.6
KZN237	Inkosi Langalibalele	KwaZulu-Natal	3,399	215,182	63.3
KZN235	Okhahlamba	KwaZulu-Natal	3,971	135,132	34
KZN241	Endumeni	KwaZulu-Natal	1,610	76,639	47.6
KZN242	Nqutu	KwaZulu-Natal	1,962	171,325	87.3
KZN252	Newcastle	KwaZulu-Natal	1,856	389,117	209.7
KZN253	Emadlangeni	KwaZulu-Natal	3,539	36,869	10.4
KZN254	Dannhauser	KwaZulu-Natal	1,707	105,341	61.7
KZN261	eDumbe	KwaZulu-Natal	1,943	89,614	46.1
KZN262	uPhongolo	KwaZulu-Natal	3,110	141,247	45.4
KZN263	Abaqulusi	KwaZulu-Natal	4,314	243,795	56.5
KZN265	Nongoma	KwaZulu-Natal	2,182	211,892	97.1
KZN266	Ulundi	KwaZulu-Natal	3,250	205,762	63.3
KZN271	Umhlabuyalingana	KwaZulu-Natal	4,977	172,077	34.6
KZN272	Jozini	KwaZulu-Natal	3,442	198,215	57.6
KZN275	Mtubatuba	KwaZulu-Natal	1,970	202,176	102.6
KZN284	uMlalazi	KwaZulu-Natal	2,214	223,140	100.8
KZN286	Nkandla	KwaZulu-Natal	1,828	114,284	62.5
KZN291	Mandeni	KwaZulu-Natal	545	147,808	271
KZN292	KwaDukuza	KwaZulu-Natal	735	276,719	376.5
KZN293	Ndwedwe	KwaZulu-Natal	1,093	143,117	131
KZN294	Maphumulo	KwaZulu-Natal	896	89,969	100.4
KZN436	Dr Nkosazana Dlamini Zuma	KwaZulu-Natal	3,602	118,480	32.9
KZN434	Ubuhlebezwe	KwaZulu-Natal	1,669	118,346	70.9
LIM331	Greater Giyani	Limpopo	4,172	256,127	61.4
LIM332	Greater Letaba	Limpopo	1,891	218,030	115.3
LIM333	Greater Tzaneen	Limpopo	2,897	416,146	143.7
LIM334	Ba-Phalaborwa	Limpopo	7,489	168,937	22.6
LIM335	Maruleng	Limpopo	3,563	99,946	28.1
LIM355	Lepele-Nkumpi	Limpopo	3,484	235,380	67.6
LIM361	Thabazimbi	Limpopo	11,190	96,232	8.6
LIM362	Lephalale	Limpopo	13,794	140,240	10.2
LIM366	Bela-Bela	Limpopo	3,406	76,296	22.4
LIM367	Mogalakwena	Limpopo	6,156	325,291	52.8
LIM471	Ephraim Mogale	Limpopo	2,011	127,168	63.2
LIM472	Elias Motsoaledi	Limpopo	3,713	268,256	72.2
LIM473	Makhuduthamaga	Limpopo	2,110	284,435	134.8
MP301	Chief Albert Luthuli	Mpumalanga	5,559	187,629	33.7
MP302	Msukaligwa	Mpumalanga	6,016	164,608	27.4
MP303	Mkhondo	Mpumalanga	4,882	189,036	38.7
MP304	Dr Pixley Ka Isaka Seme	Mpumalanga	5,227	85,395	16.3
MP305	Lekwa	Mpumalanga	4,557	123,419	27.1
MP306	Dipaleseng	Mpumalanga	2,645	45,232	17.1
MP307	Govan Mbeki	Mpumalanga	2,955	340,091	115.1
MP311	Victor Khanye	Mpumalanga	1,568	84,151	53.7
MP312	Emalahleni	Mpumalanga	2,678	455,228	170
MP313	Steve Tshwete	Mpumalanga	3,976	278,749	70.1
MP314	Emakhazeni	Mpumalanga	4,736	48,149	10.2
MP315	Thembisile	Mpumalanga	2,384	333,331	139.8
MP316	Dr JS Moroka	Mpumalanga	1,416	246,016	173.7
MP321	Thaba Chweu	Mpumalanga	5,719	101,895	17.8
MP324	Nkomazi	Mpumalanga	4,787	410,907	85.8

Table 5 – continued from previous page

Id.	Municipality name	Province	Area (km^2)	Population	Density
NW371	Moretele	North West	1,498	191,306	127.7
NW372	Local Municipality of Madibeng	North West	3,720	536,110	144.1
NW373	Rustenburg	North West	3,416	626,522	183.4
NW374	Kgetlengrivier	North West	3,973	59,562	15
NW375	Moses Kotane	North West	5,726	243,648	42.5
NW381	Ratlou	North West	4,884	106,108	21.7
NW382	Tswaing	North West	5,875	129,052	22
NW383	Mafikeng	North West	3,646	314,394	86.2
NW384	Ditsobotla	North West	6,387	181,865	28.5
NW385	Ramotshere Moiloa	North West	7,323	157,690	21.5
NW392	Naledi	North West	7,030	68,803	9.8
NW393	Mamusa	North West	3,614	64,000	17.7
NW394	Greater Taung	North West	5,639	167,827	29.8
NW396	Lekwa-Teemane	North West	3,654	56,025	15.3
NW397	Kagisano/Molopo	North West	23,827	102,703	4.3
NW403	City of Matlosana	North West	3,602	417,282	115.8
NW404	Maquassi Hills	North West	4,671	82,012	17.6
NC061	Richtersveld	Northern Cape	9,608	12,487	1.3
NC062	Nama Khoi	Northern Cape	17,990	46,512	2.6
NC064	Kamiesberg	Northern Cape	14,208	9,605	0.7
NC065	Hantam	Northern Cape	39,085	21,540	0.6
NC066	Karoo Hoogland	Northern Cape	30,230	13,009	0.4
NC067	Khi-Ma	Northern Cape	15,715	12,333	0.8
NC071	Ubuntu	Northern Cape	20,393	19,471	1
NC072	Umsobomvu	Northern Cape	6,813	30,883	4.5
NC073	Emthanjeni	Northern Cape	13,472	45,404	3.4
NC074	Kareeberg	Northern Cape	17,701	12,772	0.7
NC075	Renosterberg	Northern Cape	5,529	11,818	2.1
NC076	Thembelihle	Northern Cape	8,023	16,230	2
NC077	Siyathemba	Northern Cape	14,727	23,075	1.6
NC078	Siyancuma	Northern Cape	16,753	35,941	2.1
NC082	Kai !Garib	Northern Cape	26,377	68,929	2.6
NC084	!Kheis	Northern Cape	11,107	16,566	1.5
NC085	Tsantsabane	Northern Cape	18,290	39,345	2.2
NC086	Kgatelopele	Northern Cape	2,478	20,691	8.3
NC091	Sol Plaatjie	Northern Cape	3,145	255,041	81.1
NC092	Dikgatlong	Northern Cape	7,316	48,473	6.6
NC453	Gamagara	Northern Cape	2,648	53,656	20.3
WC011	Matzikama	Western Cape	12,981	71,045	5.5
WC012	Cederberg	Western Cape	8,007	52,949	6.6
WC013	Bergrivier	Western Cape	4,407	67,474	15.3
WC022	Witzenberg	Western Cape	10,753	130,548	12.1
WC023	Drakenstein	Western Cape	1,538	280,195	182.2
WC025	Breede Valley	Western Cape	3,834	176,578	46.1
WC026	Langeberg	Western Cape	4,518	105,483	23.3
WC033	Cape Agulhas	Western Cape	3,471	36,000	10.4
WC034	Swellendam	Western Cape	3,835	40,211	10.5
WC041	Kannaland	Western Cape	4,765	24,168	5.1
WC042	Hessequa	Western Cape	5,733	54,237	9.5
WC043	Mossel Bay	Western Cape	2,001	94,135	47
WC044	George	Western Cape	5,191	208,237	40.1
WC045	Oudtshoorn	Western Cape	3,540	97,509	27.5
WC047	Bitou	Western Cape	992	59,157	59.6
WC048	Knysna	Western Cape	1,109	73,835	66.6
WC051	Laingsburg	Western Cape	8,784	8,895	1
WC052	Prince Albert	Western Cape	8,153	14,272	1.8
WC053	Beaufort West	Western Cape	21,917	51,080	2.3
NC451	Joe Morolong	Northern Cape	20,180	84,201	4.2
NC452	Ga-Segonyana	Northern Cape	4,495	104,408	23.2
KZN213	Umzumbe	KwaZulu-Natal	1,221	151,676	124.2
KZN276	Big Five Hlabisa	KwaZulu-Natal	3,466	116,622	33.6
KZN227	Richmond	KwaZulu-Natal	1,231	71,322	57.9
KZN433	Greater Kokstad	KwaZulu-Natal	2,680	76,753	28.6
KZN435	Umzimkhulu	KwaZulu-Natal	2,436	197,286	81
NC093	Magareng	Northern Cape	1,546	24,059	15.6
NC094	Phokwane	Northern Cape	828	60,168	72.7
WC024	Stellenbosch	Western Cape	831	173,197	208.4
WC031	Theewaterskloof	Western Cape	3,259	117,167	36
EC442	Umzimvubu	Eastern Cape	2,579	199,620	77.4
EC444	Ntabankulu	Eastern Cape	1,385	128,848	93.1
EC443	Mbizana	Eastern Cape	2,415	319,948	132.5
EC123	Great Kei	Eastern Cape	1,700	31,692	18.6
EC124	Amahlathi	Eastern Cape	4,505	101,826	22.6
KZN221	uMshwathi	KwaZulu-Natal	1,866	111,645	59.8
KZN244	Msinga	KwaZulu-Natal	2,375	184,494	77.7
ETH	eThekwini	KwaZulu-Natal	2,556	3,702,231	1,448.50
KZN226	Mkhambathini	KwaZulu-Natal	868	57,075	65.7
KZN225	The Msunduzi	KwaZulu-Natal	751	679,039	904.1
KZN222	uMngeni	KwaZulu-Natal	1,520	109,867	72.3
KZN224	Impendle	KwaZulu-Natal	1,610	29,526	18.3
KZN281	Mfolozi	KwaZulu-Natal	1,300	144,363	111.1
KZN282	uMhlathuze	KwaZulu-Natal	1,233	410,465	332.8
KZN285	Mthonjaneni	KwaZulu-Natal	1,639	78,883	48.1
EC106	Sundays River Valley	Eastern Cape	5,995	59,793	10

Table 5 – continued from previous page

Id.	Municipality name	Province	Area (km^2)	Population	Density
EC108	Kouga	Eastern Cape	2,670	112,941	42.3
EC109	Kou-Kamma	Eastern Cape	3,642	43,688	12
NMA	Nelson Mandela Bay	Eastern Cape	1,957	1,263,051	645.4
BUF	Buffalo City	Eastern Cape	2,750	834,997	303.6
EC126	Ngqushwa	Eastern Cape	2,115	63,694	30.1
MP325	Bushbuckridge	Mpumalanga	10,248	546,215	53.3
EC135	Intsika Yethu	Eastern Cape	2,873	152,159	53
EC136	Emalahleni	Eastern Cape	3,484	124,532	35.7
EC138	Sakhisizwe	Eastern Cape	2,318	63,846	27.5
WC014	Saldanha Bay	Western Cape	2,015	111,173	55.2
WC015	Swartland	Western Cape	3,707	133,762	36.1
WC032	Overstrand	Western Cape	1,675	93,407	55.8
CPT	City of Cape Town	Western Cape	2,446	4,005,016	1,637.60
LIM351	Blouberg	Limpopo	9,540	172,601	18.1
LIM353	Molemole	Limpopo	3,628	125,327	34.5
LIM354	Polokwane	Limpopo	5,054	797,127	157.7
LIM368	Modimolle/Mookgophong	Limpopo	10,367	107,699	10.4
LIM476	Greater Tubatse/Fetakgomo	Limpopo	5,693	489,902	86
LIM341	Musina	Limpopo	10,347	132,009	12.8
LIM343	Thulamela	Limpopo	2,642	497,237	188.2
LIM344	Makhado	Limpopo	7,605	416,728	54.8
LIM345	New	Limpopo	5,003	347,974	69.6
NC087	Dawid Kruiper	Northern Cape	44,231	107,161	2.4
MP326	Mbombela	Mpumalanga	7,141	695,913	97.4
NW405	Ventersdorp/Tlokwe	North West	6,398	243,527	38.1
MAN	Mangaung	Free State	9,886	787,803	79.7
EC145	Walter Sisulu	Eastern Cape	13,269	87,263	6.6
EC139	Enoch Mgijima	Eastern Cape	13,584	267,011	19.7
EC129	Raymond Mhlaba	Eastern Cape	6,357	159,515	25.1
KZN245	Umvoti	KwaZulu-Natal	2,705	122,423	45.3
KZN223	Mpofana	KwaZulu-Natal	1,757	37,391	21.3
GT423	Lesedi	Gauteng	1,484	112,472	75.8
GT422	Midvaal	Gauteng	1,722	111,612	64.8
GT421	Emfuleni	Gauteng	966	733,445	759.3
EC102	Blue Crane Route	Eastern Cape	11,068	36,063	3.3
JHB	City of Johannesburg	Gauteng	1,645	4,949,347	3,008.80
EKU	Ekurhuleni	Gauteng	1,975	3,379,104	1,710.60

References

[1] A. Agresti. *Categorical Data Analysis*. John Wiley & Sons, New York, 1990.

[2] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[3] Pierre Baldi, Søren Brunak, Yves Chauvin, Claus AF Andersen, and Henrik Nielsen. Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*, 16(5):412–424, 2000.

[4] Cynthia Semá Baltazar, Roberta Horth, Celso Inguane, Isabel Sathane, Freide César, Helena Ricardo, Carlos Botão, Ângelo Augusto, Laura Cooley, Beverly Cummings, Henry F. Raymond, and Peter W. Young. HIV prevalence and risk behaviors among Mozambicans working in South African mines. *AIDS and Behavior*, 19:59–67, 2015.

[5] R. Becker, R. Cáceres, K. Hanson, S. Isaacman, J. M. Loh, M. Martonosi, J. Rowland, S. Urbanek, A. Varshavsky, and C. Volinsky. Human mobility characterization from cellular network data. *Communications of the ACM*, 56:74–82, 2013.

[6] J. Besag. Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 24:179–195, 1975.

[7] J. Besag. Efficiency of pseudolikelihood estimation for simple gaussian fields. *Biometrika*, 64:616–618, 1977.

[8] A. Bhattacharya and D. Dunson. Simplex factor models for multivariate unordered categorical data. *Journal of the American Statistical Association*, 107:362–377, 2012.

[9] Y. M. M. Bishop, S. E. Fienberg, and P. W. Holland. *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, MA, 1975. Reprinted 2007, Springer-Verlag, New York.

[10] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, jan 2006.

[11] F. Calabrese, M. Diao, G. D. Lorenzo, J. Ferreira Jr., and C. Ratti. Understanding individual mobility patterns from urban sensing data: a mobile phone trace example. *Transportation Research Part C: Emerging Technologies*, 26:301–313, 2013.

[12] A. Canale and D. Dunson. Bayesian kernel mixtures of counts. *Journal of the American Statistical Association*, 106:1528–1539, 2011.

[13] O. Cappé, C.P. Robert, and T. Rydén. Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(3):679–700, 2003.

[14] Yuan Cheng, Alex Lenkoski, et al. Hierarchical Gaussian graphical models: Beyond reversible jump. *Electronic Journal of Statistics*, 6:2309–2331, 2012.

[15] M. Clyde and E. I. George. Model uncertainty. *Statistical Science*, 19:81–94, 2004.
[16] P. Dellaportas and J. J. Forster. Markov chain Monte Carlo model determination for hierarchical and graphical log-linear models. *Biometrika*, 86:615–633, 1999.
[17] P. Dellaportas and C. Tarantola. Model determination for categorical data with factor level merging. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67:269–283, 2005.
[18] Xavier Descombes, Robert Minlos, and Elena Zhizhina. Object extraction using a stochastic birth-and-death dynamics in continuum. *Journal of Mathematical Imaging and Vision*, 33(3):347–359, 2009.
[19] A. Dobra, T. Baernighausen, A. Vandormael, and F. Tanser. Space-time migration patterns and risk of HIV acquisition in rural South Africa. *AIDS*, 31:37–145, 2017.
[20] A. Dobra and A. Lenkoski. Copula Gaussian graphical models and their application to modeling functional disability data. *Annals of Applied Statistics*, 5:969–993, 2011.
[21] A. Dobra, A. Lenkoski, and A. Rodriguez. Bayesian inference for general Gaussian graphical models with application to multivariate lattice data. *Journal of the American Statistical Association*, 106:1418–1433, 2011.
[22] A. Dobra and H. Massam. The mode oriented stochastic search (MOSS) algorithm for log-linear models with conjugate priors. *Statistical Methodology*, 7:240–253, 2010.
[23] A. Dobra, N. E. Williams, and N. Eagle. Spatiotemporal detection of unusual human population behavior using mobile phone data. *PLOS ONE*, 10:1–20, 2015.
[24] K. Donato. Current trends and patterns of female migration: Evidence from Mexico. *International Migration Review*, 27:748–771, 1993.
[25] M. Drton and M. H. Maathuis. Structure learning in graphical modeling. *The Annual Review of Statistics and Its Application*, 4:365–393, 2017.
[26] D. B. Dunson and C. Xing. Nonparametric Bayes modeling of multivariate categorical data. *Journal of the American Statistical Association*, 104:1042 – 1051, 2009.
[27] J. Durand, W. Kandel, E. A. Parrado, and D. S. Massey. International migration and development in Mexican communities. *Demography*, 33:249–264, 1996.
[28] D. E. Edwards and T. Havranek. A fast procedure for model search in multidimensional contingency tables. *Biometrika*, 72:339–351, 1985.
[29] Ahmed Gamal Eldin, Xavier Descombes, Guillaume Charpiat, and Josiane Zerubia. Multiple birth and cut algorithm for multiple object detection. *Journal of Multimedia Processing and Technologies*, 2012.
[30] S. E. Fienberg. The analysis of multidimensional contingency tables. *Ecology*, 51:419–433, 1970.
[31] S. E. Fienberg. *The Analysis of Cross-Classified Categorical Data*. MIT Press, Cambridge, MA, 1980. Reprinted 2007, Springer-Verlag, New York.
[32] S. E. Fienberg and A. Rinaldo. Three centuries of categorical data analysis: log-linear models and maximum likelihood estimation. *Journal of Statistical Planning and Inference*, 137:3430–3445, 2007.
[33] S. E. Fienberg and A. Rinaldo. Maximum likelihood estimation in log-linear models. *Annals of Statistics*, 40:996–1023, 2012.
[34] Ahmed Gamal-Eldin, Xavier Descombes, Guillaume Charpiat, and Josiane Zerubia. A fast multiple birth and cut algorithm using belief propagation. In *2011 18th IEEE International Conference on Image Processing*, pages 2813–2816. IEEE, 2011.
[35] Ahmed Gamal-Eldin, Xavier Descombes, and Josiane Zerubia. Multiple birth and cut algorithm for point process optimization. In *Signal-Image Technology and Internet-Based Systems (SITIS), 2010 Sixth International Conference on*, pages 35–42. IEEE, 2010.
[36] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, jun 2008.
[37] P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995.
[38] M. Guerzhoy and A. Hertzmann. Learning latent factor models of travel data for travel prediction and analysis. In M. Sokolova and P. van Beek, editors, *Advances in Artificial Intelligence: 27th Canadian Conference on Artificial Intelligence, Canadian AI 2014, Montréal, QC, Canada, May 6-9, 2014. Proceedings*, pages 131–142. Springer International Publishing, Cham, 2014.
[39] J. R. Harris and M. P. Todaro. Migration, unemployment and development: A two-sector analysis. *American Economic Review*, 60:126–142, 1970.
[40] H. Höefling and R. Tibshirani. Estimation of sparse binary pairwise markov networks with pseudo-likelihoods. *Journal of Machine Learning Research*, 10:883–906, 2009.
[41] P. D. Hoff. Multiplicative latent factor models for description andprediction of social networks. *Computational and Mathematical Organization Theory*, 15(4):261, 2008.
[42] K. Imai. *Quantitative Social Science: An Introduction*. Princeton University Press, 2017.
[43] B. Jones, C. Carvalho, A. Dobra, C. Hans, C. Carter, and M. West. Experiments in stochastic computation for high-dimensional graphical models. *Statistical Science*, 20:388–400, 2005.
[44] R. Jurdak, K. Zhao, J. Liu, M. AbouJaoude, M. Cameron, and D. Newth. Understanding human mobility from Twitter. *PLOS ONE*, 10(7):1–16, 07 2015.

[45] T. Kunihama and D. B. Dunson. Bayesian modeling of temporal dependence in large sparse contingency tables. *Journal of the American Statistical Association*, 108:1324–1338, 2013.
[46] S.L. Lauritzen. *Graphical models*, volume 17. Oxford University Press, USA, 1996.
[47] K. Leetaru, S. Wang, G. Cao, A. Padmanabhan, and E. Shook. Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday*, 18(5):4366, 2013.
[48] A. Lenkoski and A. Dobra. Computational aspects related to inference in Gaussian graphical models with the G-Wishart prior. *Journal of Computational and Graphical Statistics*, 20:140–157, 2011.
[49] D. Madigan and A.E. Raftery. Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association*, 89:1535–1546, 1994.
[50] D. Madigan and J. York. Bayesian graphical models for discrete data. *International Statistical Review*, 63:215–232, 1995.
[51] D. Madigan and J. York. Bayesian methods for estimation of the size of a closed population. *Biometrika*, 84:19–31, 1997.
[52] David Madigan, Adrian E Raftery, C Volinsky, and J Hoeting. Bayesian model averaging. In *Proceedings of the AAAI Workshop on Integrating Multiple Learned Models, Portland, OR*, pages 77–83, 1996.
[53] H. Massam, J. Liu, and A. Dobra. A conjugate prior for discrete hierarchical log-linear models. *Annals of Statistics*, 37:3431–3467, 2009.
[54] D. S. Massey. Social structure, household strategies, and the cumulative causation of migration. *Population Index*, 56:3–26, 1990.
[55] D. S. Massey, J. Arango, G. Hugo, A. Kouaouci, A. Pellegrino, and J. E. Taylor. Theories of international migration: A review and appraisal. *Population and Development Review*, 19:431–466, 1993.
[56] D. S. Massey and K. E. Espinosa. What's driving Mexico-U.S. migration? A theoretical, empirical, and policy analysis. *American Journal of Sociology*, 102:939–999, 1997.
[57] D. S. Massey, N. Williams, W. G. Axinn, and D. Ghimire. Community services and out-migration. *International Migration*, 48:1–41, 2010.
[58] A Mohammadi, H Massam, and G Letac. The ratio of normalizing constants for bayesian graphical gaussian model selection. *arXiv preprint arXiv:1706.04416*, 2017.
[59] A. Mohammadi and Ernst C Wit. BDgraph: Bayesian structure learning of graphs in R. *arXiv preprint arXiv:1501.05108v2*, 2015.
[60] Abdolreza Mohammadi, Fentaw Abegaz, Edwin Heuvel, and Ernst C Wit. Bayesian modelling of dupuytren disease by using gaussian copula graphical models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(3):629–645, 2017.
[61] Abdolreza Mohammadi and E. C. Wit. Bayesian structure learning in sparse Gaussian graphical models. *Bayesian Analysis*, 10(1):109–138, 2015.
[62] Abdolreza Mohammadi and Ernst Wit. *BDgraph: Bayesian Structure Learning in Graphical Models using Birth-Death MCMC*, 2017. R package version 2.42.
[63] Y. Nardi and A. Rinaldo. The log-linear group lasso estimator for hierarchical log-linear models and its asymptotic properties. *Bernoulli*, 18:945–974, 2012.
[64] G. Neubauer, H. Huber, A. Vogl, B. Jager, A. Preinerstorfer, S. Schirnhofer, G. Schimak, and D. Havlik. On the volume of geo-referenced tweets and their relationship to events relevant for migration tracking. In R. Denzer, R. M. Argent, G. Schimak, and J. Hřebíček, editors, *Environmental Software Systems. Infrastructures, Services and Applications: 11th IFIP WG 5.11 International Symposium, ISESS 2015, Melbourne, VIC, Australia, March 25-27, 2015. Proceedings*, pages 520–530. Springer International Publishing, Cham, 2015.
[65] OpenMP Architecture Review Board. OpenMP application program interface version 3.0, May 2008.
[66] OpenStreetMap Contributors. Planet dump retrieved from `http://planet.osm.org`. `http://www.openstreetmap.org`, 2017.
[67] J. Pensar, H. Nyman, J. Niiranen, and J. Corander. Marginal pseudo-likelihood learning of discrete Markov network structures. *Bayesian Analysis*, 2017.
[68] C. J. Preston. Spatial birth-and-death processes. *Bulletin of the International Statistical Institute*, 46:371–391, 1977.
[69] P. Ravikumar, M. Wainwright, and J. Lafferty. High-dimensional Ising model selection with l_1-regularized logistic regression. *Annals of Statistics*, 38:1278–1319, 2010.
[70] J. Raymer, G. Abel, and P. W. F. Smith. Combining census and registration data to estimate detailed elderly migration flows in England and Wales. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170:891–908, 2007.
[71] J. Raymer, A. Wiśniowski, J. J. Forster, P. W. F. Smith, and J. Bijak. Integrated modeling of European migration. *Journal of the American Statistical Association*, 108:801–819, 2013.
[72] James G Scott and Carlos M Carvalho. Feature-inclusion stochastic search for Gaussian graphical models. *Journal of Computational and Graphical Statistics*, 17(4):790–808, 2008.
[73] SMaPP. smappR package: Tools for analysis of Twitter data, Social Media and Participation, New York University. Available online at `https://github.com/SMAPPNYU/smappR`, 2017.

[74] P. W. F. Smith, J. Raymer, and C. Giulietti. Combining available migration data in England to study economic activity flows over time. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173:733–753, 2010.
[75] O. Stark and D. E. Bloom. The new economics of labor migration. *American Economic Review*, 75:173–178, 1985.
[76] O. Stark and J. E. Taylor. Migration incentives, migration types: The role of relative deprivation. *The Economic Journal*, 101:1163–1178, 1985.
[77] P. R. Stopher and S. P. Greaves. Household travel surveys: where are we going? *Transportation Research Part A: Policy and Practice*, 41:367–381, 2007.
[78] C. Tarantola. MCMC model determination for discrete graphical models. *Statistical Modelling*, 4:39–61, 2004.
[79] A. J. Tatem. Mapping population and pathogen movements. *International Health*, 6:5–11, 2014.
[80] J. E. Taylor. Undocumented Mexico-U.S. migration and the returns to households in rural Mexico. *American Journal of Agricultural Economics*, 69:616–638, 1987.
[81] M. P. Todaro. A model of labor migration and urban unemployment in less developed countries. *The American Economic Review*, 59:138–148, 1969.
[82] M. P. Todaro and L. Maruszko. Illegal immigration and U.S. immigration reform: A conceptual framework. *Population and Development Review*, 13:101–114, 1987.
[83] Ioannis Tsamardinos, Laura E Brown, and Constantin F Aliferis. The max-min hill-climbing bayesian network structure learning algorithm. *Machine learning*, 65(1):31–78, 2006.
[84] Twitter, Inc. Twitter REST APIs. Available online at `https://dev.twitter.com/rest/public`, 2017.
[85] L. K. VanWey. Land ownership as a determinant of international and internal migration in Mexico and internal migration in Thailand. *International Migration Review*, 39:141–172, 2005.
[86] M. Wainwright and M. Jordan. Graphical models, exponential families and variational inference. *Foundations and Trends in Machine Learning*, 1:1–305, 2008.
[87] H. Wang and S.Z. Li. Efficient Gaussian graphical model determination under G-wishart prior distributions. *Electronic Journal of Statistics*, 6:168–198, 2012.
[88] J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. John Wiley & Sons, 1990.
[89] N. Williams. Education, gender, and migration in the context of social change. *Social Science Research*, 38:883–896, 2009.
[90] N. E. Williams, T. A. Thomas, M. Dunbar, N. Eagle, and A. Dobra. Measures of human mobility using mobile phone records enhanced with gis data. *PLOS ONE*, 10:1–16, 2015.
[91] J. Wolf, M. Oliveira, and M. Thompson. Impact of underreporting on mileage and travel time estimates: Results from global positioning system-enhanced household travel survey. *Transportation Research Record: Journal of the Transportation Research Board*, 1854:189–198, 2003.

Department of Statistics, Department of Biobehavioral Nursing and Health Informatics, Center for Statistics and the Social Sciences and Center for Studies in Demography and Ecology, University of Washington, Box 354322, Seattle, WA 98195
E-mail address: `adobra@uw.edu`

Department of Methodology and Statistics, Tilburg University, Netherlands
E-mail address: `a.mohammadi@uvt.nl`