

Geometric Representations of Hypergraphs for Prior Specification and Posterior Sampling

Simón Lunagómez, Sayan Mukherjee and Robert L. Wolpert

Simón Lunagómez¹, Sayan Mukherjee^{1,2,3,4}, and Robert L. Wolpert^{1,5}

Department of Statistical Science¹

Department of Computer Science²

Institute for Genome Sciences & Policy³

Department of Mathematics⁴

Nicholas School of the Environment⁵

Duke University

Durham, NC 27708, USA

e-mail: s165@duke.edu; sayan@stat.duke.edu; wolpert@stat.duke.edu

Abstract: A parametrization of hypergraphs based on the geometry of points in \mathbb{R}^d is developed. Informative prior distributions on hypergraphs are induced through this parametrization by priors on point configurations via spatial processes. This prior specification is used to infer conditional independence models or Markov structure of multivariate distributions. Specifically, we can recover both the junction tree factorization as well as the hyper Markov law. This approach offers greater control on the distribution of graph features than Erdős-Rényi random graphs, supports inference of factorizations that cannot be retrieved by a graph alone, and leads to new Metropolis/Hastings Markov chain Monte Carlo algorithms with both local and global moves in graph space. We illustrate the utility of this parametrization and prior specification using simulations.

AMS 2000 subject classifications: Primary 60K35, 60K35; secondary 60K35.

Keywords and phrases: Abstract simplicial complex, Computational topology, Copulas, Factor models, Graphical models, Random geometric graphs.

1. Introduction

It is common to model the joint probability distribution of a family of n random variables $\{X_1, \dots, X_n\}$ in two stages: First to specify the *conditional dependence structure* of the distribution, then to specify details of the conditional distributions of the variables within that structure [see p.1274 of Dawid and Lauritzen 8, or p. 180 of Besag 3, for example]. The structure may be summarized in a variety of ways in the form of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ whose vertices $\mathcal{V} = \{1, \dots, n\}$ index the variables $\{X_i\}$ and whose edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ in some way encode conditional dependence. We follow the Hammersley-Clifford approach [2, 14], in which $(i, j) \in \mathcal{E}$ if and only if the conditional distribution of X_i given all other variables $\{X_k : k \neq i\}$ depends on X_j , *i.e.*, differs from the conditional distribution of X_i given $\{X_k : k \neq i, j\}$. In this case the distribution is said to be Markov with respect to the graph. One can show that this graph is symmetric or *undirected*, *i.e.*, all the elements of \mathcal{E} are unordered pairs.

Our primary goal is the construction of informative prior distributions on undirected graphs, motivated by the problem of Bayesian inference of the dependence structure of families of observed random variables. As a side benefit our approach also yields estimates of the conditional distributions given the graph. The model space of undirected graphs grows quickly with the dimension of the vector (there are $2^{n(n-1)/2}$ undirected graphs on n vertices) and is difficult to parametrize. We propose a novel parametrization and a simple, flexible family of prior distributions on \mathcal{G} and on Markov probability distributions with respect to \mathcal{G} [8]; this parametrization is based on computing the intersection pattern of a system of convex sets in \mathbb{R}^d . The novelty and main contribution of this paper is structural inference for graphical models, specifically, the proposed representation of graph spaces allows for flexible prior distributions and new Markov chain Monte Carlo (MCMC) algorithms.

The simultaneous inference of a decomposable graph and marginal dis-

tributions in a fully Bayesian framework was approached in [13] using local proposals to sample graph space. A promising extension of this approach called Shotgun Stochastic Search (SSS) takes advantage of parallel computing to select from a batch of local moves [19]. A stochastic search method that incorporates both local moves and more aggressive global moves in graph space has been developed by Scott and Carvalho [31]. These stochastic search methods are intended to identify regions with high posterior probability, but their convergence properties are still not well understood. Bayesian models for non-decomposable graphs have been proposed by Roverato [30] and by Wong, Carter and Kohn [34]. These two approaches focus on Monte Carlo sampling of the posterior distribution from specified hyper Markov prior laws. Their emphasis is on the computational problem of Monte Carlo simulation, not on that of constructing interesting informative priors on graphs. We think there is need for methodology that offers both efficient exploration of the model space and a simple and flexible family of distributions on graphs that can reflect meaningful prior information.

Erdős-Rényi random graphs (those in which each of the $\binom{n}{2}$ possible undirected edges (i, j) is included in \mathcal{E} independently with some specified probability $p \in [0, 1]$), and variations where the edge inclusion probabilities p_{ij} are allowed to be edge-specific, have been used to place informative priors on decomposable graphs [16, 21]. The number of parameters in this prior specification can be enormous if the inclusion probabilities are allowed to vary, and some interesting features of graphs (such as decomposability) cannot be expressed solely through edge probabilities. Mukherjee and Speed [22] developed methods for placing informative distributions on directed graphs by using *concordance functions* (functions that increase as the graph agrees more with a specified feature) as potentials in a Markov model. This approach is tractable, but it is still not clear how to encode certain common assumptions (*e.g.*, decomposability) within such a framework.

For the special case of jointly Gaussian variables $\{X_j\}$, or those with

arbitrary marginal distributions $F_j(\cdot)$ whose dependence is adequately represented in Gaussian copula form $X_j = F_j^{-1}(\Phi(Z_j))$ for jointly Gaussian $\{Z_j\}$ with zero mean and unit-diagonal covariance matrix C_{ij} , the problem of studying conditional independence reduces to a search for zeros in the precision matrix C^{-1} . This approach [see 17, for example] is faster and easier to implement than ours in cases where both are applicable, but is far more limited in the range of dependencies it allows. For example, a three-dimensional model in which each pair of variables is conditionally independent given the third cannot be distinguished from a model with complete joint dependence of the three variables (we return to this example in Section 5.3).

In this article we establish a novel approach to parametrize spaces of graphs. For any integers $n, d \in \mathbb{N}$, we show in Section 2 how to use the geometrical configuration of a set $\{v_i\}$ of n points in Euclidean space \mathbb{R}^d to determine a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ on $\mathcal{V} = \{v_1, \dots, v_n\}$. Any prior distribution on point sets $\{v_i\}$ induces a prior distribution on graphs, and sampling from the posterior distribution of graphs is reduced to sampling from spatial configurations of point sets—a standard problem in spatial modeling. Relations between graphs and finite sets of points have arisen earlier in the fields of computational topology [9] and random geometric graphs [25]. From the former we borrow the idea of *nerves*, *i.e.*, simplicial complexes computed from intersection patterns of convex subsets of \mathbb{R}^d ; the 1-skeletons (collection of 1-dimensional simplices) of nerves are geometric graphs.

From the random geometric graph approach we gain understanding about the induced distribution on graph features when making certain features of a geometric graph (or hypergraph) stochastic.

1.1. Graphical models

The graphical models framework is concerned with the representation of conditional dependencies for a multivariate distribution in the form of a graph or hypergraph. We first review relevant graph theoretical concepts

and then relate these concepts to factorizing distributions.

A *graph* \mathcal{G} is an ordered pair $(\mathcal{V}, \mathcal{E})$ of a set \mathcal{V} of *vertices* and a set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ of *edges*. If all edges are unordered (resp., ordered), the graph is said to be *undirected* (resp., *directed*). All graphs considered in this paper are undirected, unless stated otherwise. A *hypergraph*, denoted \mathcal{H} , consists of a vertex set \mathcal{V} and a collection \mathcal{K} of unordered subsets of \mathcal{V} (known as *hyperedges*); a graph is the special case where all the subsets are vertex pairs. A graph is *complete* if $\mathcal{E} = \mathcal{V} \times \mathcal{V}$ contains all possible edges; otherwise it is *incomplete*. A complete subgraph that is maximal with respect to inclusion is a *clique*. Denote by $\mathcal{C}(\mathcal{G})$ and $\mathcal{Q}(\mathcal{G})$, respectively, the collection of complete sets and cliques of \mathcal{G} . A *path* between two vertices $\{v_i, v_j\} \in \mathcal{V}$ is a sequence of edges connecting v_i to v_j . A graph such that any pair of vertices can be joined by a unique path is a *tree*. A *weak decomposition* of an incomplete graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a partition of \mathcal{V} into disjoint nonempty sets (A, B, S) such that S is complete in \mathcal{G} and *separates* A and B , *i.e.*, any path from a vertex in A to a vertex in B must pass through S . Iterative weak decomposition of a graph \mathcal{G} such that at each step the separator S_i is minimal and the subsets A_i and B_i are nonempty generates the *prime components* of \mathcal{G} , the collection of subgraphs that cannot be further weakly decomposed. If all prime components of a graph \mathcal{G} are complete, then \mathcal{G} is said to be *weakly decomposable*. Any graph \mathcal{G} can be represented as a tree \mathcal{T} whose vertices are its prime components $\mathcal{P}(\mathcal{G})$; this is called its *junction tree* representation. A junction tree is a hypergraph.

Let \mathcal{P} be a probability distribution on \mathbb{R}^n and $X = (X_1, \dots, X_n)$ a random vector with distribution \mathcal{P} . *Graphical modeling* is the representation of the Markov or conditional dependence structure among the components $\{X_i\}$ in the form of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Denote by $f(x)$ the joint density function of $\{X_i\}$ (or probability mass function for discrete distributions—more generally, density for an arbitrary reference measure). The distribution \mathcal{P} (and hence its density $f(x)$) may depend implicitly on a vector θ of pa-

rameters, taking values in some set $\Theta_{\mathcal{G}}$, which in some cases will depend on the graph \mathcal{G} ; write $\Theta = \sqcup \Theta_{\mathcal{G}}$ for the disjoint union of the parameter spaces for all graphs on \mathcal{V} .

Each vertex $v_i \in \mathcal{V}$ is associated with a variable X_i , and the edges \mathcal{E} determine how the distribution factors. The density $f(x)$ for the distribution can be factored in a variety of ways associated with the graph \mathcal{G} [20, p. 35]. It may be factored in terms of complete sets $a \in \mathcal{C}(\mathcal{G})$:

$$f(x) = \prod_{a \in \mathcal{C}(\mathcal{G})} \phi_a(x_a | \theta_a), \quad (1.1a)$$

or similarly in terms of cliques $a \in \mathcal{Q}$; if \mathcal{G} is decomposable then $f(x)$ may also be factored in junction-tree form as:

$$f(x) = \frac{\prod_{a \in \mathcal{P}(\mathcal{G})} \psi_a(x_a | \theta_a)}{\prod_{b \in \mathcal{S}(\mathcal{G})} \psi_b(x_b | \theta_b)}, \quad (1.1b)$$

where $\mathcal{P}(\mathcal{G})$ and $\mathcal{S}(\mathcal{G})$ denote the prime factors and separators of \mathcal{G} , respectively, and where $\psi_a(x_a | \theta_a)$ denotes the marginal joint density for the components x_a for prime factors $a \in \mathcal{P}(\mathcal{G})$ and $\psi_b(x_b | \theta_b)$ that for separators $b \in \mathcal{S}(\mathcal{G})$ [8, Eqn. (6)]. In the Gaussian case, a similar factorization to (1.1b) holds even for non-decomposable graphs [30, Prop. 2], but $\psi_a(x_a)$ and $\psi_b(x_b)$ lose their simple interpretation.

The prior distributions required for Bayesian inference about models of the form (1.1) may be specified by giving a marginal distribution on the set of all graphs $\mathcal{G} \in \mathcal{G}_n$ on n vertices and conditional distributions on each $\Theta_{\mathcal{G}}$, the space of parameters for that graph:

$$p(\mathcal{G}, \theta) = p(\mathcal{G}) p(\theta | \mathcal{G}), \quad \mathcal{G} \in \mathcal{G}_n, \theta \in \Theta_{\mathcal{G}} \quad (1.2)$$

where $\theta \in \Theta_{\mathcal{G}}$ determines the parameters $\{\theta_a : a \in \mathcal{C}(\mathcal{G})\}$ or $\{\theta_a : a \in \mathcal{P}(\mathcal{G})\}$ and $\{\theta_b : b \in \mathcal{S}(\mathcal{G})\}$. Giudici and Green [12] pursue this approach in the Gaussian case, while Dawid and Lauritzen [8] offer a rigorous framework for specifying more general prior distributions on $\Theta_{\mathcal{G}}$. Such priors, called *hyper*

Markov laws, inherit the conditional independence structure from the sampling distribution, now at the parameter level. The hyper Inverse Wishart, useful when the factors are multivariate normal, is by far the most studied hyper Markov law. Most previously studied models of the form (1.2) specify very little structure on $p(\mathcal{G})$ [12, 16, 30]— typically $p(\mathcal{G})$ is taken to be a uniform distribution on the space of decomposable (or unrestricted) graphs, or perhaps an Erdős-Rényi prior to encourage sparsity [21], with no additional structure or constraints and hence no opportunity to express prior knowledge or belief.

Two inference problems arise for the model specified in (1.2): inference of the entire joint posterior distribution of the graph and factor parameters, (θ, \mathcal{G}) , or inference of only the *conditional independence structure*, which entails comparing different graphs via the marginal likelihood

$$\Pr \{ \mathcal{G} \mid x \} \propto \int_{\Theta_{\mathcal{G}}} f(x \mid \theta, \mathcal{G}) p(\mathcal{G}) p(\theta \mid \mathcal{G}) d\theta.$$

Inference about \mathcal{G} may now be viewed as a Bayesian model selection procedure [see 28, p. 348].

2. Geometric Graphs

Most methodology for structural inference in graphical models either assumes little prior structure on graph space, or else represents graphs using high dimensional discrete spaces with no obvious geometry or metric. In either case prior elicitation and posterior sampling can be challenging. In this section we propose parametrizations of graph space that will be used in Section 3 to specify flexible prior distributions and to construct new Metropolis/Hastings MCMC algorithms with local and global moves. The key idea for this parametrization is to construct graphs and hypergraphs from intersections of convex sets in \mathbb{R}^d .

We illustrate the approach with an example. Fix a convex region $A \subset \mathbb{R}^d$ and let $\mathcal{V} \subset A$ be a finite set of n points. For each number $r \geq 0$, the

proximity graph $\text{Prox}(\mathcal{V}, r)$ (see Figure 1) is formed by joining every pair of (unordered) elements in \mathcal{V} whose distance is $2r$ or less, *i.e.*, whose closed balls of radius r intersect. As r ranges from 0 to half the diameter of A , the graph $\text{Prox}(\mathcal{V}, r)$ ranges from the totally disconnected graph to the complete graph. This example is a particular case of a more general construction illustrated in Figure 2; hypergraphs can be computed from properties of intersections of classes of convex subsets in Euclidean space. The convex sets we consider are subsets of \mathbb{R}^d that are simple to parametrize and compute. The key concept in our construction is the *nerve*:

Definition 2.1 (Nerve). Let $F = \{A_j, j \in I\}$ be a finite collection of distinct nonempty convex sets. The *nerve* of F is given by

$$\text{Nrv}(F) = \left\{ \sigma \subseteq I : \bigcap_{j \in \sigma} A_j \neq \emptyset \right\}.$$

The nerve of a family of sets uniquely determines a hypergraph. We use the following three nerves in this paper to construct hypergraphs [for more details, see 9].

Definition 2.2 (Čech Complex). Let \mathcal{V} be a finite set of points in \mathbb{R}^d and $r > 0$. Denote by \mathbb{B}^d the closed unit ball in \mathbb{R}^d . The *Čech complex* corresponding to \mathcal{V} and r is the nerve of the sets $B_{v,r} = v + r\mathbb{B}^d$, $v \in \mathcal{V}$. This is denoted by $\text{Nrv}(\mathcal{V}, r, \check{\text{Cech}})$.

Definition 2.3 (Delaunay Triangulation). Let \mathcal{V} be a finite set of points in \mathbb{R}^d . The *Delaunay triangulation* corresponding to \mathcal{V} is the nerve of the sets $C_v = \{x \in \mathbb{R}^d : \|x - v\| \leq \|x - u\|, u \in \mathcal{V}\}$ for $v \in \mathcal{V}$. This is denoted by $\text{Nrv}(\mathcal{V}, \text{Delaunay})$, and the sets C_v are called *Voronoi cells*.

Definition 2.4 (Alpha Complex). Let \mathcal{V} be a finite set of points in \mathbb{R}^d and $r > 0$. The *Alpha complex* corresponding to \mathcal{V} and r is the nerve of the sets $B_{v,r} \cap C_v$, $v \in \mathcal{V}$. This is denoted by $\text{Nrv}(\mathcal{V}, r, \text{Alpha})$.

The nerve of a family of sets is a particular class of hypergraphs known as (abstract) simplicial complexes.

Definition 2.5 (Abstract simplicial complex). Let \mathcal{V} be a finite set. A *simplicial complex* with base set \mathcal{V} is a family \mathcal{K} of subsets of \mathcal{V} such that $\tau \in \mathcal{K}$ and $\sigma \subseteq \tau$ implies $\sigma \in \mathcal{K}$. The elements of \mathcal{K} are called simplices, and the number of connected components of \mathcal{K} is denoted $\sharp(\mathcal{K})$.

The nerve of a collection of sets is always a hypergraph; in simple cases, only vertex pairs arise so the 1-skeleton determines a unique graph.

Definition 2.6 (p -skeleton). Let \mathcal{K} be a simplicial complex, and denote by $|\tau|$ the cardinality of a simplex $\tau \in \mathcal{K}$. The p -skeleton of \mathcal{K} is the collection of all $\tau \in \mathcal{K}$ such that $|\tau| \leq p + 1$. The elements of the p -skeleton are called p -simplices and the 1-skeleton is just a graph (more precisely, it is $\mathcal{V} \cup \mathcal{E}$ for a uniquely determined graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$).

The 1-skeleton of a nerve is the graph obtained by considering only nonempty *pairwise* intersections. The process of obtaining the nerve and the 1-skeleton from a family of sets is illustrated in Figure 3. Different families of convex sets in \mathbb{R}^d induce different restrictions in graph space: for the Delaunay triangulation and the Alpha complex, for example, clique sizes cannot exceed $d + 1$. Although no such blanket restriction applies to the Čech complex, for this complex some graphs are still unattainable— for example, no Čech complex can include a star graph whose central node has degree higher than the “kissing number,” *i.e.*, maximal number of disjoint unit hyperspheres touching a given hypersphere, 6 for $d = 2$, 12 for $d = 3$, *etc.*

The Čech and Alpha complexes are hypergraphs indexed by a finite set $\mathcal{V} = \{V_1, \dots, V_n\} \subset \mathbb{R}^d$ and a size parameter $r \geq 0$. Each induces a parametrization on the space of hypergraphs $(\mathcal{V}, r) \mapsto \mathcal{H}(\mathcal{V}, r)$. The class \mathcal{A} of convex sets used to compute the nerve determines the space of hypergraphs. To keep the notation simple, \mathcal{A} will be implicit whenever obvious by the context. We will use $A(\mathcal{V}, r)$ to denote a generic element of \mathcal{A} for either the Čech or the Alpha complex. Similarly, 1-skeletons of nerves induce a parametrization of the spaces of graphs $(\mathcal{V}, r) \mapsto \mathcal{G}(\mathcal{V}, r)$.

Two principal advantages of this approach are:

1. For each family of convex sets $\{\mathcal{A}\}$, the number of parameters needed to specify the graph \mathcal{G} or hypergraph \mathcal{H} grows only linearly with the number of vertices;
2. The hypergraph parameter space will be a subset of \mathbb{R}^d , a very convenient parameter space for MCMC sampling.

2.1. Example

Here we use a junction tree factorization with each univariate marginal X_i associated to a point $V_i \in \mathbb{R}^d$ (the standard graphical models approach). In this case, specifying the class of sets to compute the nerve and the value for r determines a factorization for the joint density of $\{X_1, \dots, X_n\}$. We illustrate with $n = 5$ points in Euclidean space of dimension $d = 2$.

Let $(X_1, X_2, X_3, X_4, X_5) \in \mathbb{R}^5$ be a random vector with density $f(x)$ and consider the vertex set displayed in Table 1 (also shown as solid dots in Figures 4 and 5). For an Alpha complex with $r = 0.5$ the junction tree

Coordinate	V_1	V_2	V_3	V_4	V_5
x	0.2065	0.6383	0.9225	-0.8863	0.3043
y	0.3149	-0.1193	-0.2544	0.0816	-0.9310

TABLE 1

Vertex set used for generating a factorization based on nerves.

factorization (1.1b) corresponding to the graph in Figure 4 is

$$f(x) = \frac{\psi_{12}(x_1, x_2)\psi_{235}(x_2, x_3, x_5)\psi_4(x_4)}{\psi_2(x_2)},$$

we will denote the factorization as $[1, 2][2, 3, 5][4]$. In the case where the factors are potential functions rather than marginals we will use $\{\cdot\}$ instead of $[\cdot]$. Similarly, for the Čech complex and $r = 0.7$ the factorization corresponding to the graph in Figure 5 is

$$f(x) = \frac{\psi_{1235}(x_1, x_2, x_3, x_5)\psi_{14}(x_1, x_4)}{\psi_1(x_1)}.$$

2.2. Filtrations and Decomposability

A *filtration* is a sequence of simplicial complexes that properly include their predecessors:

Definition 2.7 (Filtration). A *filtration* for a simplicial complex \mathcal{K} is a sequence $\mathcal{L} = \{\mathcal{K}^0, \mathcal{K}^1, \dots, \mathcal{K}^k\}$ of simplicial complexes such that

$$\emptyset = \mathcal{K}^0 \subset \mathcal{K}^1 \subset \dots \subset \mathcal{K}^k = \mathcal{K},$$

with all inclusions proper.

Filtrations are commonly used in computational geometry and topology to construct efficient algorithms for computing specific nerves including the Alpha complex [11]. The simplicial complexes constructed in Section 2 from families of convex sets lead to filtrations as the convex sets are enlarged, by increasing the radius parameter r .

Although much of the graphical models literature focuses on Markov structures derived from weakly decomposable graphs, those constructed in Section 2 from 1-skeletons of Čech and Alpha complexes need not be weakly decomposable (see Figure 6). In Algorithm 1 we present an adaptation of this construction that generates weakly decomposable graphs, for use in applications that require them. In Sections 4 and 5 we present methodology and examples for both weakly decomposable and unrestricted model spaces.

The central idea for generating decomposable graphs from filtrations is to note that the complex $\mathcal{A}(\mathcal{V}, 0)$ for $r = 0$ is disconnected and hence weakly decomposable; as the radius r increases, if one adds edges *only* if the resulting graph is weakly decomposable, then decomposability will hold for all $r \geq 0$. This procedure is formalized in Algorithm 1.

Proposition 2.1. *The graph \mathcal{G} produced by Algorithm 1 is weakly decomposable.*

Proof. The algorithm is initialized with the weakly decomposable empty graph, and weak decomposability is tested with each proposed addition of

Algorithm 1: The algorithm takes as input a filtration of k abstract complexes and returns a decomposable graph that is a subset of the 1-skeleton of the k -th complex.

```

input : a filtration  $\mathcal{L} = \{\mathcal{K}^0, \mathcal{K}^1, \dots, \mathcal{K}^k\}$ 
return: a weakly decomposable graph  $\mathcal{G}$ 
 $m = 0$ ;  $i = 0$ ;  $\mathcal{G}_0 = \emptyset$ ;
while  $i < k$  and  $\mathcal{K}^{i+1} \neq \emptyset$  do
   $\tau_i = \{\kappa \in \mathcal{K}^{i+1} \setminus \mathcal{K}^i \text{ such that } |\kappa| = 2\}$ ; // the edges in the set difference
  and denote  $\tau_{i,s}$  as the  $s$ th edge
  if  $\tau_i \neq \emptyset$  then
     $P = |\tau_i|$ ;
    for  $s = 1$  to  $P$  do
       $\mathcal{G}' = \mathcal{G}_m \cup \tau_{i,s}$ ; // propose adding the edge
      if  $\#\mathcal{G}' < \#\mathcal{G}_m$  // fewer connected components?
      then
         $\mathcal{G}_{m+1} = \mathcal{G}'$ ; // accept the proposal
         $m = m + 1$ ;
      else
         $[c_i] = \mathcal{C}(\mathcal{G}_m)$ ; // the cliques
         $[s_i] = \mathcal{S}(\mathcal{G}_m)$ ; // the separators
         $[v_1, v_2] = \tau_{i,s}$ ; // the proposed edge
        if  $\exists i \neq j, k$  with  $v_1 \in c_i, v_2 \in c_j$  and  $c_i \cap c_j = s_k$  then
           $\mathcal{G}_{m+1} = \mathcal{G}'$ ; // accept the proposal
           $m = m + 1$ ;
     $i = i + 1$ ;
 $\mathcal{G} = \mathcal{G}_m$ 

```

an edge (*i.e.*, a 1-simplex in \mathcal{L}). The decomposability test is taken from [12, Theorem 2]. Since only finitely many edges may possibly be added, \mathcal{G} is weakly decomposable by construction. ■

2.2.1. Example

We first illustrate the algorithm on a simple example, based on the five points in \mathbb{R}^2 shown in Figure 7 and given in Table 2. The graph induced by a Čech complex with $r = 0.5$ will not be decomposable. Table 3 presents the evolution of cliques and separators with increasing r , as edges are proposed for inclusion in Algorithm 1. The first three proposed edge additions are accepted, but the proposal to add edge $(1, 2)$ at radius $r = 0.474$ is rejected, since the intersection of prime components $\{1, 3\}$ and $\{2, 4, 5\}$ is empty, and therefore not a separator.

Coordinate	V_1	V_2	V_3	V_4	V_5
x	0.686	0.214	0.846	0.411	0.089
y	0.151	0.194	0.420	0.567	0.553

TABLE 2
Vertex set used to illustrate Algorithm 1.

Cliques	Separators	r	Update
[1][2][3][4][5]	–	0	–
[1, 3][2][4][5]	–	0.313	(1, 3)
[1, 3][2][4, 5]	–	0.321	(4, 5)
[1, 3][2, 5][4, 5]	[5]	0.379	(2, 5)
[1, 3][2, 4, 5]	–	0.421	(2, 4)
[1, 3][3, 4][2, 4, 5]	[3][4]	0.459	(3, 4)
[1, 3][3, 4][2, 4, 5]	[3][4]	0.474	\sim (1, 2)
[1, 3, 4][2, 4, 5]	[4]	0.498	(1, 4)

TABLE 3
Evolution of cliques and separators in the junction tree representation of \mathcal{G} as edges are added according to Algorithm 1. The proposed addition of edge $(1, 2)$ is rejected.

2.2.2. Algorithm deletes few edges

It is interesting to note that the number of proposed edge additions that are rejected by the algorithm is typically quite small. To illustrate this we applied Algorithm 1 to a filtration of Čech complexes corresponding to 100 points sampled uniformly from the unit square, with radius $r = 0.05$. In Figure 8 the graph \mathcal{G} output by the algorithm is compared to the 1-skeleton of the Čech complex (with no decomposability restriction) with the same radius $r = 0.05$. Few edges appear in the Čech complex but not in \mathcal{G} . This occurs because geometric graphs tend to be triangulated, in the sense that if edges (v_1, v_2) and (v_2, v_3) belong to a geometric graph, then very likely the edge (v_1, v_3) will also be in the graph, preserving decomposability.

3. Random Geometric Graphs

In Section 2 we demonstrated how the geometry of a set \mathcal{V} of n points in \mathbb{R}^d can be used to induce a graph \mathcal{G} . In this section we explore the relation between prior distributions on *random* sets \mathbf{V} of points in \mathbb{R}^d and features of the induced distribution on graphs \mathcal{G} , with the goal of learning how to tailor a point process model to obtain graph distributions with desired features.

Definition 3.1 (Random Geometric Graph). Fix integers $n, d \in \mathbb{N}$ and let $\mathbf{V} = (V_1, \dots, V_n)$ be drawn from a probability distribution \mathcal{Q} on $(\mathbb{R}^d)^n$. For any class \mathcal{A} of convex sets in \mathbb{R}^d and radius $r > 0$, the graph $\mathcal{G}(\mathbf{V}, r, \mathcal{A})$ is said to be a *Random Geometric Graph* (RGG).

While Definition 3.1 is more general than that of [25, p. 2], it still cannot describe all the random graphs discussed in [26] (for example, those based on k -neighbors cannot in general be generated by nerves). For \mathcal{A} we will use closed balls in \mathbb{R}^d or intersections of balls and Voronoi cells; most often \mathcal{Q} will be a product measure under which the $\{V_i\}$ will be n independent identically distributed (*iid*) draws from some marginal distribution \mathcal{Q}_M on \mathbb{R}^d , such as the uniform distribution on the unit cube $[0, 1]^d$ or unit ball \mathbb{B}^d ,

but we will also explore the use of repulsive processes for \mathbf{V} under which the points $\{V_i\}$ are more widely dispersed than under independence. It is clear that different choices for \mathcal{A} , \mathcal{Q} and r will have an impact on the support of the induced RGG distribution. To make this notion precise we define feasible graphs. First,

Definition 3.2 (Isomorphic). Write $\mathcal{G}_1 \cong \mathcal{G}_2$ for two graphs $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)$ and call the graphs *isomorphic* if there is a 1:1 mapping $\chi : \mathcal{V}_1 \rightarrow \mathcal{V}_2$ such that $(v_i, v_j) \in \mathcal{E}_1 \Leftrightarrow (\chi(v_i), \chi(v_j)) \in \mathcal{E}_2$ for all $v_i, v_j \in \mathcal{V}_1$.

Definition 3.3 (Feasible Graph). Fix numbers $d, n \in \mathbb{N}$, a class \mathcal{A} of convex sets in \mathbb{R}^d , and a distribution \mathcal{Q} on the random vectors \mathbf{V} in $(\mathbb{R}^d)^n$. A graph Γ is said to be *feasible* if for some number $r > 0$,

$$\Pr \{ \mathcal{G}(\mathbf{V}, r, \mathcal{A}) \cong \Gamma \} > 0.$$

In contrast to Erdős-Rényi models, where the inclusion of graph edges are independent events, the RGG models exhibit edge dependence that depends on the metric structure of \mathbb{R}^d and the class \mathcal{A} of convex sets used to construct the nerves.

There is an extensive literature describing asymptotic distributions for a variety of graph features such as: subgraph counts, vertex degree, order of the largest clique, and maximum vertex degree [for an encyclopedic account of results for the important case of 1-skeletons of Čech complexes, see 25]. Several results for the Delaunay triangulation, some of which generalize to the Alpha complex, are reported in [26].

Penrose [25], Chap. 3 gives conditions which guarantee the asymptotic normality of the joint distribution of the numbers Q_j of j -simplices (edges, triads, etc.), for *iid* samples $\mathbf{V} = (V_1, \dots, V_n)$ from some marginal distribution \mathcal{Q}_M on \mathbb{R}^d , as the number $n = |\mathbf{V}|$ of vertices grows and the radius r_n shrinks.

Simulation studies suggest that the asymptotic results apply approximately for $n \geq 24$ –100.

3.1. Simulation Study of Subgraph Counts for RGGs

In this subsection we study the distribution of particular graph features as a function of the sampling distribution of the random point set \mathbf{V} and contrast this with Erdős-Rényi models. Specifically we will focus on the number of edges (2-cliques) Q_2 and the number of 3-cliques Q_3 .

The two spatial processes we study for \mathcal{Q} are *iid* uniform draws from the unit square $[0, 1]^2$ in the plane, and dependent draws from the Matérn type III hard-core repulsive process [18], using Čech complexes with radius $r = 1/\sqrt{150} \approx 0.082$ in both cases to ensure asymptotic normality [25, Thm. 3.13]. In our simulations we vary both the number of variables (graph size) n and the Matérn III hard core radius ρ . Comparisons are made with an Erdős-Rényi model with a common edge inclusion parameter. Table 4 displays the quartiles for Q_2 and Q_3 as a function of the graph size n , hard core radius ρ , and Erdős-Rényi edge inclusion probability p . Figures 9, 10, and 11 show the joint distribution of (Q_2, Q_3) for $\{V_i\} \stackrel{\text{iid}}{\sim} \text{Un}([0, 1]^2)$, for a Matérn III process with hard core radius $\rho = 0.35$, and for draws from an Erdős-Rényi model with inclusion probability $p = 0.065$, respectively.

These simulations illustrate that by varying the distribution \mathcal{Q} we can control the joint distribution of graph features. The repulsive and *iid* uniform distribution have very similar edge distributions, for example (see Figures 9 and 10), while (as anticipated) the repulsive process penalizes large cliques. Joint control of these features is not possible with an Erdős-Rényi model with a common edge inclusion probability and it is not obvious how to encode this type of information in the concordance function approach of Mukherjee and Speed [22].

In Section 2.2 we proposed a procedure for generating decomposable graphs, and noted that the graphs induced by this algorithm are similar to those constructed without the decomposability restriction. In Figure 12 we display a simulation study of the distribution of edge counts for a RGG and the restriction to decomposable graphs. These distributions are very

Graph	V	Edges			3-Cliques		
		25%	50%	75%	25%	50%	75%
Uniform	75	161	171	182	134	160	190
Matérn (0.035)	75	154	161	170	110	124	144
ER (0.050)	75	130	138	146	6	8	11
ER (0.065)	75	172	181	189	14	18	22
Uniform	50	69	75	81	34	43	57
Matérn (0.035)	50	66	71	76	27	35	43
Matérn (0.050)	50	62	67	71	22	27	33
ER (0.050)	50	56	61	67	1	2	4
ER (0.065)	50	74	79	85	3	5	7
Uniform	20	9	12	14	1	2	4
Matérn (0.035)	20	9	11	13	1	1	3
Matérn (0.050)	20	8	10	12	0	1	2
ER (0.050)	20	8	9	11	0	0	0
ER (0.065)	20	10	12	15	0	0	1

TABLE 4

Summaries of the empirical distribution of edge and 3-clique counts for Čech complex random geometric graphs with radius $r = 0.082$, for vertex sets sampled from iid draws from the unit square from: a uniform distribution, a hard core process with radius $\rho = 0.035$, and from Erdős-Rényi (ER) with common edge inclusion probabilities of $p = 0.050$ and $p = 0.065$.

similar.

4. Model specification

4.1. General Setting

We offer a Bayesian approach to the problem of inferring factorizations of distributions of the forms of (1.1),

$$f(x) = \prod_{a \in \mathcal{L}(\mathcal{G})} \phi_a(x_a | \theta_a) \quad \text{or} \quad \frac{\prod_{a \in \mathcal{P}(\mathcal{G})} \psi_a(x_a | \theta_a)}{\prod_{b \in \mathcal{S}(\mathcal{G})} \psi_b(x_b | \theta_b)}.$$

In each case we specify the prior density function as a product

$$p(\theta, \mathcal{G}) = p(\theta | \mathcal{G}) p(\mathcal{G}) \tag{4.1}$$

of a conditional hyper Markov law for $\theta \in \Theta$ and a marginal RGG law on \mathcal{G} . We use conventional methods to select the specific hyper Markov distribution

(hyper Inverse Wishart for multivariate normal sampling distributions, for example) since our principal focus is on prior distributions for the graphs, $p(\mathcal{G})$. We also present MCMC algorithms for sampling from the posterior distribution on \mathcal{G} , for observed data.

4.2. Prior Specification

All the graphs in our statistical models are built from nerves constructed in Section 2 from a random vertex set $\mathcal{V} = \{V_i\}_{i=1}^n \subset \mathbb{R}^d$ and radius $r > 0$. Since the nerve construction is invariant under rigid transformations of \mathcal{V} or simultaneous scale changes in \mathcal{V} and r , restricting the support of the prior distribution on \mathcal{V} to the unit ball \mathbb{B}^d does not reduce the model space:

Proposition 4.1. *Every feasible graph in \mathbb{R}^d may be represented in the form $\mathcal{G}(\mathcal{V}, r, \mathcal{A})$ for a collection \mathcal{V} of n points in the unit ball \mathbb{B}^d and for $r = \frac{1}{n}$.*

Proof. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}) \cong \mathcal{G}(\mathcal{V}, r, \mathcal{A})$ be a feasible graph with $|\mathcal{V}| = n$ vertices. Every edge $(v_i, v_j) \in \mathcal{E}$ has length $\text{dist}(v_i, v_j) \leq 2r$ so, by the triangle inequality, every connected component Γ_i of \mathcal{G} with n_i vertices must have diameter no greater than the longest possible path length, $2r(n_i - 1)$, and so fits in a ball B_i of diameter $2r(n_i - 1)$. The union of these balls, centered on a line segment and separated by $r(2 + 1/n)$, will have diameter less than $r(2n - 1)$. By translation and linear rescaling we may take $r = 1/n$ and bound the diameter by 2, completing the proof. ■

We fix $r = \frac{1}{n}$ and simplify the notation by writing $\mathcal{G}(\mathcal{V}, \mathcal{A})$ instead of $\mathcal{G}(\mathcal{V}, r, \mathcal{A})$ for $\mathcal{A} = \check{\text{C}}\text{ech}$ or $\mathcal{A} = \text{Alpha}$ or, if \mathcal{A} is understood, simply $\mathcal{G}(\mathcal{V})$. Thus we can induce prior distributions on the space of feasible graphs from distributions on configurations of n points in the unit ball in \mathbb{R}^d .

For *iid* uniform draws $\mathbf{V} = (V_1, \dots, V_n)$ from \mathbb{B}^d , the expected number of edges of the graph $\mathcal{G}(\mathbf{V}, r, \mathcal{A})$ is bounded above by $\mathbb{E}[\#\mathcal{E}] \leq \binom{n}{2}(2r)^d$; for $r = \frac{1}{n}$ in dimension $d = 2$ this is less than $\mathbb{E}[\#\mathcal{E}] < 2$, leading to relatively sparse graphs. We often take larger values of r (still small enough for empty

graphs to be feasible), to generate richer classes of graphs. A limit to how large r may be is given by the partial converse to Prop. 4.1,

Proposition 4.2. *The empty graph on n vertices cannot be expressed as $\mathcal{G}(\mathcal{V}, r, \check{\text{Cech}})$ for any $\mathcal{V} \subset \mathbb{B}^d$ with $r \geq (n^{1/d} - 1)^{-1}$.*

Proof. Let $\mathcal{V} = \{V_1, \dots, V_n\} \subset \mathbb{B}^d$ be a set of points and $r > 0$ a radius such that $\mathcal{G}(\mathcal{V}, r, \check{\text{Cech}})$ is the empty graph. Then the balls $\mathcal{V}_i + r\mathbb{B}^d$ are disjoint and their union with d -dimensional volume $n\omega_d r^d$ lies wholly within the ball $(1+r)\mathbb{B}^d$ of volume $\omega_d(1+r)^d$ (where $\omega_d = \pi^{d/2}/\Gamma(1+d/2)$ is the volume of the unit ball), so $n < (1 + \frac{1}{r})^d$. ■

Slightly stronger, the empty graph may not be attained as $\mathcal{G}(\mathcal{V}, r, \check{\text{Cech}})$ for any $r \geq 1/[(n/p_d)^{1/d} - 1]$ where p_d is the maximum spherical packing density in \mathbb{R}^d . For $d = 2$, this gives the asymptotically sharp bound $r < 1/\left[\sqrt{n\sqrt{12}/\pi} - 1\right]$.

4.3. Sampling from Prior and Posterior Distributions

Let \mathcal{Q} be a probability distribution on n -tuples in \mathbb{R}^d , $p(\mathcal{G})$ the induced prior distribution on graphs $\mathcal{G}(\mathbf{V}, \check{\text{Cech}})$ for $\mathbf{V} \sim \mathcal{Q}$ with $r = \frac{1}{n}$, and let $p(\theta | \mathcal{G})$ be a conventional hyper Markov law (see below). We wish to draw samples from the prior distribution $p(\theta, \mathcal{G})$ of (4.1) and from the posterior distribution $p(\theta, \mathcal{G} | \mathbf{x})$, given a vector $\mathbf{x} = (x_1, \dots, x_N)$ of *iid* observations $x_j \stackrel{\text{iid}}{\sim} f(x | \theta)$, using the Metropolis/Hastings approach to MCMC [15, 29, Ch. 7].

We begin with a random walk proposal distribution in \mathbb{B}^d starting at an arbitrary point $v \in \mathbb{B}^d$, that approximates the steps $\{V^{(0)}, V^{(1)}, V^{(2)}, \dots\}$ of a diffusion $V^{(t)}$ on \mathbb{B}^d with uniform stationary distribution and reflecting boundary conditions at the unit sphere $\partial\mathbb{B}^d$.

The random walk is conveniently parametrized in spherical coordinates with radius $\rho^{(t)} = |V^{(t)}|$ and Euler angles— in $d=2$ dimensions, angle $\varphi^{(t)}$ — at step t . Informally, we take independent radial random walk steps such that

$(\rho^{(t)})^d$ is reflecting Brownian motion on the unit interval (this ensures that the stationary distribution will be $\text{Un}(\mathbb{B}^d)$) and, conditional on the radius, angular steps from Brownian motion on the d -sphere of radius $\rho^{(t)}$.

Fix some $\eta > 0$. In $d = 2$ dimensions the reflecting random walk proposal (ρ^*, φ^*) we used for step $(t + 1)$, beginning at $(\rho^{(t)}, \varphi^{(t)})$, is:

$$\rho^* = R\left([\rho^{(t)}]^2 + \zeta_\rho^{(t)} \eta\right)^{1/2}, \quad \varphi^* = \varphi^{(t)} + \zeta_\phi^{(t)} \eta / \rho^{(t)}$$

for *iid* standard normal random variables $\{\zeta_\rho^{(t)}, \zeta_\phi^{(t)}\}$, where

$$R(x) = \left| x - 2 \left\lfloor \frac{1}{2}(x + 1) \right\rfloor \right|$$

is x reflected (as many times as necessary) to the unit interval. Similar expressions work in any dimension d , with $\rho^* = R([\rho^{(t)}]^d + \zeta_\rho^{(t)} \eta)^{1/d}$ and appropriate step sizes for the $(d - 1)$ Euler angles.

For small $\eta > 0$ this diffusion-inspired random walk generates local moves under which the proposed new point (ρ^*, φ^*) is quite close to $(\rho^{(t)}, \varphi^{(t)})$ with high probability. To help escape local modes, and to simplify the proof of ergodicity below, we add the option of more dramatic “global” moves by introducing at each time step a small probability of replacing $(\rho^{(t)}, \varphi^{(t)})$ with a random draw (ρ^*, φ^*) from the uniform distribution on \mathbb{B}^d (see Figure 13). Let $q(\mathcal{V}^* | \mathcal{V})$ denote the Lebesgue density at $\mathcal{V}^* \in (\mathbb{B}^d)^n$ of one step of this hybrid random walk for $\mathcal{V} = (V_1, \dots, V_n)$, starting at $\mathcal{V} \in (\mathbb{B}^d)^n$.

4.3.1. Prior Sampling

To draw sample graphs from the prior distribution begin with $\mathbf{V}^{(0)} \sim \mathcal{Q}(d\mathbf{V})$ and, after each time step $t \geq 0$, propose a new move to $\mathbf{V}^* \sim q(\mathbf{V}^* | \mathbf{V}^{(t)})$. The proposed move from $\mathbf{V}^{(t)}$ (with induced graph $\mathcal{G}^{(t)} = \mathcal{G}(\mathbf{V}^{(t)})$) to \mathbf{V}^* (and \mathcal{G}^*) is accepted (whereupon $\mathbf{V}^{(t+1)} = \mathbf{V}^*$) with probability $1 \wedge H^{(t)}$, the minimum of one and the Metropolis/Hastings ratio

$$H^{(t)} = \frac{p(\mathbf{V}^*) q(\mathbf{V}^{(t)} | \mathbf{V}^*)}{p(\mathbf{V}^{(t)}) q(\mathbf{V}^* | \mathbf{V}^{(t)})}.$$

Otherwise $\mathbf{V}^{(t+1)} = \mathbf{V}^{(t)}$; in either case set $t \leftarrow t+1$ and repeat. Note the proposal distribution $q(\cdot | \cdot)$ leaves the uniform distribution invariant, so $H^{(t)} \equiv 1$ for $\mathcal{Q}(d\mathbf{V}) \propto d\mathbf{V}$ and in that case every proposal is accepted.

4.3.2. Posterior Sampling

After observing a random sample $X = \mathbf{x} = (x_1, \dots, x_N)$ from the distribution $x_j \sim f(x | \theta, \mathcal{G})$, let

$$f(\mathbf{x} | \theta, \mathcal{G}) = \prod_{i=1}^N f(x_i | \theta, \mathcal{G})$$

denote the likelihood function and

$$\mathcal{M}(\mathcal{G}) = \int_{\Theta_{\mathcal{G}}} f(\mathbf{x} | \theta, \mathcal{G}) p(\theta | \mathcal{G}) d\theta \quad (4.2)$$

the marginal likelihood for \mathcal{G} . For posterior sampling of graphs, a proposed move from $\mathbf{V}^{(t)}$ to \mathbf{V}^* is accepted with probability $1 \wedge H^{(t)}$ for

$$H^{(t)} = \frac{\mathcal{M}(\mathcal{G}^*) p(\mathbf{V}^*) q(\mathbf{V}^{(t)} | \mathbf{V}^*)}{\mathcal{M}(\mathcal{G}^{(t)}) p(\mathbf{V}^{(t)}) q(\mathbf{V}^* | \mathbf{V}^{(t)})}. \quad (4.3)$$

For multivariate normal data X and hyper inverse Wishart hyper Markov law $p(\theta | \mathcal{G})$, $\mathcal{M}(\mathcal{G})$ from (4.2) can be expressed in closed form for decomposable graphs $\mathcal{G}(\mathbf{V})$. efficient algorithms for evaluating (4.2) are still available even if this condition fails.

The model will typically be of variable dimension, since the parameter space $\Theta_{\mathcal{G}}$ for the factors may depend on the graph $\mathcal{G} = \mathcal{G}(\mathbf{V})$. Not all proposed moves of the point configuration $\mathbf{V}^{(t)} \rightsquigarrow \mathbf{V}^*$ will lead to a change in $\mathcal{G}(\mathbf{V})$; for those that do we implement reversible-jump MCMC [13, 32] using the auxiliary variable approach of Brooks *et al.* [4] to simplify the book-keeping needed for non-nested moves $\Theta_{\mathcal{G}} \rightsquigarrow \Theta_{\mathcal{G}^*}$.

4.4. Convergence of the Markov chain

Denote by $\dot{\mathcal{G}}(n, d, \mathcal{A})$ the finite set of feasible graphs with n vertices in \mathbb{R}^d , *i.e.*, those generated from 1-skeletons of \mathcal{A} -complexes. For each $\mathcal{G} \in \dot{\mathcal{G}}(n, d, \mathcal{A})$ let $V_{\mathcal{G}} \subset (\mathbb{B}^d)^n$ denote the set of all points $\mathbf{V} = \{V_1, \dots, V_n\} \in (\mathbb{B}^d)^n$ for which $\mathcal{G} \cong \mathcal{G}(\mathbf{V}, \frac{1}{n}, \mathcal{A})$, and set $\mu(\mathcal{G}) = \mathcal{Q}(V_{\mathcal{G}})$. Then

Proposition 4.3. *The sequence $\mathcal{G}^{(t)} = \mathcal{G}(\mathbf{V}^{(t)}, \frac{1}{n}, \mathcal{A})$ induced by the prior MCMC procedure described in Section 4.3.1 samples each feasible graph $\mathcal{G} \in \dot{\mathcal{G}}(n, d, \mathcal{A})$ with asymptotic frequency $\mu(\mathcal{G})$. The posterior procedure described in Section 4.3.2 samples each feasible graph with asymptotic frequency $\mu(\mathcal{G} \mid \mathbf{x})$, the posterior distribution of \mathcal{G} given the data \mathbf{x} and hyper Markov prior $p(\theta \mid \mathcal{G})$.*

Proof. Both statements follow from the Harris recurrence of the Markov chain $\mathbf{V}^{(t)}$ constructed in Section 4.3. For this it is enough to find a strictly positive lower bound for the probability of transitioning from an arbitrary point $V \in (\mathbb{B}^d)^n$ to any open neighborhood of another arbitrary point $V^* \in (\mathbb{B}^d)^n$ [29, Theorem 6.38, pg. 225]. This follows immediately from our inclusion of the global move in which all n points $\{V_i\}$ are replaced with uniform draws from $(\mathbb{B}^d)^n$. ■

It is interesting to note that while the sequence $\mathcal{G}^{(t)} = \mathcal{G}(\mathbf{V}^{(t)}, \frac{1}{n}, \mathcal{A})$ is a hidden Markov process, it is not itself Markovian on the finite state space $\dot{\mathcal{G}}(n, d, \mathcal{A})$; nevertheless it is ergodic, by Prop. 4.3.

5. Four Simulation Examples

We illustrate the use of the proposed parametrization in Bayesian inference for graphical models: this is done by specifying priors that encourage sparsity and the design of MCMC algorithms that allow for local as well as global moves. We offer four examples. The first example illustrates that our method

works when the graph encoding the Markov structure of underlying density is contained in the space of graphs spanned by the nerve used to fit the model. In the second example we apply our method to Gaussian Graphical Models. The third example shows that the nerve hypergraph (not just the 1-skeleton) can be used to induce different groupings in the terms of a factorization, and therefore a way to encode dependence features that go beyond pairwise relationships. In the fourth example we compare results obtained by using different filtrations to induce priors over different spaces of graphs.

5.1. Example 1: \mathcal{G} is in the Space Generated by \mathcal{A}

Let (X_1, \dots, X_{10}) be a random vector whose distribution has factorization:

$$f_{\theta}(\mathbf{x}) = \frac{\psi_{\theta}(x_1, x_4, x_{10})\psi_{\theta}(x_1, x_8, x_{10})\psi_{\theta}(x_4, x_5)\psi_{\theta}(x_8, x_9)\psi_{\theta}(x_2, x_3, x_9)\psi_{\theta}(x_6)\psi_{\theta}(x_7)}{\psi_{\theta}(x_4)\psi_{\theta}(x_8)\psi_{\theta}(x_9)\psi_{\theta}(x_1, x_{10})} \quad (5.1a)$$

The Markov structure of (5.1a) can be encoded by the geometric graph displayed in Figure 14. We transform variables if necessary to achieve standard $\text{Un}(0, 1)$ marginal distributions for each X_i , and model clique joint marginals with a Clayton copula [6, or 24, §4.6], the exchangeable multivariate model with joint distribution function

$$\Psi_{\theta}(\mathbf{x}_I) = \left(1 - n_I + \sum_{i \in I} x_i^{-\theta}\right)^{-1/\theta}$$

and density function

$$\psi_{\theta}(\mathbf{x}_I) = \theta^{n_I} \frac{\Gamma(n_I + 1/\theta)}{\Gamma(1/\theta)} \left(1 - n_I + \sum_{i \in I} x_i^{-\theta}\right)^{-n_I - 1/\theta} \left(\prod_{i \in I} x_i\right)^{-1 - \theta} \quad (5.1b)$$

on $[0, 1]^{n_I}$ for some $\theta \in \Theta = (0, \infty)$, for each clique $[v_i : i \in I]$ of size n_I .

We drew 250 samples from model (5.1) with $\theta = 4$. For inference about \mathcal{G} we set $\mathcal{A} = \text{Alpha}$ and $r = 0.30$, with independent uniform prior distributions for the vertices $V_i \stackrel{\text{iid}}{\sim} \text{Un}(\mathbb{B}^2)$ on the unit ball in the plane. We

used the random walk described in Section 4.3 to draw posterior samples with Algorithm 1 applied to enforce decomposability. To estimate θ we take a unit Exponential prior distribution $\theta \sim \text{Ex}(1)$ and employ a Metropolis/Hastings approach using a symmetric random walk proposal distribution with reflecting boundary conditions at $\theta = 0$,

$$\theta^* = |\theta^{(t)} + \varepsilon|,$$

with $\varepsilon_t \sim \text{Un}(-\beta, \beta)$ for fixed $\beta > 0$. We drew 1 000 samples after a burn-in period of 25 000 draws. The three models with the highest posterior probabilities are displayed in Table 5. The geometric graphs computed from nine posterior samples (one every 100 draws) are shown in Figure 15; note that the computed nerves appear to stabilize after a few hundred iterations while the actual position of the vertex set continues to vary.

Graph Topology	Posterior Probability
[1, 4, 10][1, 8, 10][4, 5][8, 9][2, 3, 9][6][7]	0.963
[1, 4, 10][1, 8, 10][4, 5][8, 9][2, 3, 9][6][5, 7]	0.021
[1, 4, 10][1, 8][4, 5][8, 9][2, 3, 9][6][7]	0.010

TABLE 5

The three models with highest estimated posterior probability. The true model is shown in bold (see Figure 14). Here $\theta = 4$.

5.2. Example 2: Gaussian graphical model

We use our procedure to perform model selection for the Gaussian graphical model $X \sim \text{No}(0, \Sigma_{\mathcal{G}})$, where \mathcal{G} encodes the zeros in Σ^{-1} . We adopt a Hyper Inverse Wishart (HIW) prior distribution for $\Sigma \mid \mathcal{G}$. The marginal likelihood [in the parametrization of 1, Eqn (12)] is given by

$$\mathcal{M}(\mathbf{V}) = (2\pi)^{-nN/2} \frac{I_{\mathcal{G}(\mathbf{V})}(\delta + N, D + X^T X)}{I_{\mathcal{G}(\mathbf{V})}(\delta, D)}, \quad (5.2)$$

where

$$I_{\mathcal{G}}(\delta, D) = \int_{M^+(\mathcal{G})} |\Sigma|^{(\delta-2)/2} e^{-\frac{1}{2}\langle \Sigma, D \rangle} d\Sigma$$

denotes the HIW normalizing constant. This quantity is available in closed form for weakly decomposable graphs $\mathcal{G}(\mathbf{V})$, but for our unrestricted graphs (5.2) must be approximated via simulation. For our low-dimensional examples the method of [1] suffices; for larger numbers of variables we recommend that of [5]. We set $\delta = 3$ and $D = 0.4I_6 + 0.6J_6$ (I_6 and J_6 denote the identity matrix and the matrix of all ones, respectively).

We sampled 300 observations from a Multivariate Normal with mean zero and precision matrix

$$\begin{pmatrix} 18.18 & -6.55 & 0 & 2.26 & -6.27 & 0 \\ -6.55 & 14.21 & 0 & -4.90 & 0 & 0 \\ 0 & 0 & 10.47 & 0 & 0 & -3.65 \\ 2.26 & -4.90 & 0 & 10.69 & 0 & 0 \\ -6.27 & 0 & 0 & 0 & 27.26 & 0 \\ 0 & 0 & -3.65 & 0 & 0 & 7.41 \end{pmatrix} \quad (5.3)$$

whose conditional independence structure is given by the graph in Figure 16. We fit the model described in Section 4 using a uniform prior for each $V_i \in \mathbb{B}^2$ and $r = 0.25$. We employed hybrid random walk proposals in which we move all five vertices $\{V_i\}$ independently according to the diffusion-inspired random walk described in Section 4.3 with probability 0.85; replace one uniformly selected vertex V_i with a uniform draw from $\text{Un}(\mathbb{B}^2)$ with probability 0.05; and replace all five vertices with independent uniform draws from $\text{Un}(\mathbb{B}^2)$ with probability 0.10. We sampled 1 000 observations from the posterior after a burn in of 750 000. Results are summarized in Table 6

5.3. Example 3: Factorization Based on Nerves

While Gaussian joint distributions are determined entirely by the bivariate marginals, and so only edges appear in their complete-set factorizations (see

Graph Topology	Posterior Probability
[1, 2, 4][1, 5][3, 6]	0.152
[1, 5][2, 3, 4][2, 3, 6]	0.072
[1, 2, 3, 4, 6][1, 5]	0.069
[1, 4][2, 4][2, 3, 6]	0.055
[1, 2, 4][2, 3, 4][1, 5][3, 6]	0.052

TABLE 6

The five models with highest estimated posterior probability. The true model is shown in bold.

(1.1a)); more complex dependencies are possible for other distributions. The familiar example of the joint distribution of three Bernoulli variables X_1, X_2, X_3 each with mean $1/2$, with X_1 and X_2 independent but $X_3 = (X_1 - X_2)^2$ (so that $\{X_i\}$ are only *pairwise* independent) has only the complete set $\{1, 2, 3\}$ in its factorization.

Consider now a model with the graphical structure illustrated in Figure 17 whose density function, if it is continuous and strictly positive ([see 20, Prop. 3.1]), admits the complete-set factorization:

$$f(x \mid \mathcal{G}, \theta) = c_{\mathcal{G}} \phi(x_1, x_2) \phi(x_1, x_6) \phi(x_2, x_6) \cdot \phi(x_3, x_4, x_5). \quad (5.4a)$$

For illustration we will take each $\phi(\cdot)$ to be a Clayton copula density (see (5.1b)). For simplicity we specify the same value $\theta = 4$ for each association parameter, so $f(x \mid \mathcal{G}, \theta)$ is given by (5.4a) with

$$\phi(x, y) = 5 (x^{-4} + y^{-4} - 1)^{-9/4} (xy)^{-5} \quad (5.4b)$$

$$\phi(x, y, z) = 30(x^{-4} + y^{-4} + z^{-4} - 2)^{-13/4} (xyz)^{-5}. \quad (5.4c)$$

In earlier examples we associated graphical structures (*i.e.*, edge sets) with 1-skeletons of nerves. We now associate *hypergraphical* structures (*i.e.*, abstract simplicial complexes that may include higher-order simplexes) with the entire nerves, with maximal simplices associated with complete-set factors. For example: the Alpha complex computed from the vertex set displayed in Table 7 with $r = 0.40$ has $\{3, 4, 5\}\{1, 2\}\{1, 6\}\{2, 6\}$ as its maximal

Coordinate	V_1	V_2	V_3	V_4	V_5	V_6
x	-0.0936	-0.4817	0.0019	0.0930	0.2605	-0.5028
y	0.6340	0.7876	0.0055	0.0351	-0.0702	0.2839

TABLE 7

Vertex set used for generating a factorization based on nerves.

simplices (Figure 18). By associating a Clayton copula to each of these hyperedges we recover the model shown in (5.4).

We use the same prior and proposal distributions constructed in Section 4.3 from point distributions in \mathbb{R}^d ; what has changed is the way the nerve is being used: as a hypergraph whose maximal hyperedges represent factors. One complicating factor is the need to evaluate the normalizing factor $c_{\mathcal{G}}$ for each graph \mathcal{G} we encounter during the simulation; unavailable in closed form, we use Monte Carlo importance sampling to evaluate $c_{\mathcal{G}}$ for each new graph, and store the result to be reused when \mathcal{G} recurs.

We anticipate that uniform draws $V_i \stackrel{\text{iid}}{\sim} \text{Un}(\mathbb{B}^2)$ will give high probability to clusters of three or more points within a ball of radius r , favoring higher-dimensional features (triangles and tetrahedra) in the induced hypergraph encoding the Markov structure of $\{X_i\}$. To explore this phenomenon, we compare results for uniform draws with those from a repulsive process under which clusters of three or more points are unlikely to lie within a ball of radius r , hence favoring hypergraphs with only edges.

We began by sampling 650 observations from model (5.4) with $\mathcal{A} = \text{Alpha}$ and $r = 0.40$, with independent uniform prior distributions for the vertices $V_i \stackrel{\text{iid}}{\sim} \text{Un}(\mathbb{B}^2)$ on the unit ball in the plane. The Metropolis/Hastings proposals for the vertex set are given by a mixture scheme:

- A random walk for each V_i as described in Section 4.3, with step size $\eta = 0.020$. This proposal is picked with probability 0.94
- An integer $1 \leq k \leq 6$ is chosen uniformly and, given k , a subset of size k from $\{1, 2, 3, 4, 5, 6\}$ is sampled uniformly; the vertices corresponding to those indices are replaced with random independent draws from

$\text{Un}(\mathbb{B}^2)$. This proposal is picked with probability 0.06, 0.01 for each k .

For θ we used the same standard exponential prior distribution and reflecting uniform random walk proposals described in Example 5.1.

Using 5 000 posterior samples after a burn-in period of 95 000 iterations, the models with highest posterior probability are summarized in Table 8.

Maximal Simplices	Posterior Probability
{3, 4, 5}{1, 2}{2, 6}{1, 6}	0.609
{1, 2, 6}{3, 4}{4, 5}{3, 5}	0.161
{3, 5}{1, 6}{3, 4}{1, 2}{2, 6}	0.137

TABLE 8

Highest posterior factorizations with uniform prior for model of (5.4) and Figure 17 (true model is shown in bold).

To penalize higher-order simplexes we used a Strauss repulsive process [33] conditioned to have n points in \mathbb{B}^d as prior distribution for the vertex set, with Lebesgue density

$$g(v) \propto \gamma^{\#\{(i,j): \text{dist}(v_i, v_j) < 2R\}}$$

for some $0 < \gamma \leq 1$, penalizing each pair of points closer than $2R$. Simulation results for this prior (with $R = 0.7r$ and $\gamma = 0.75$) are summarized in Table 9. The posterior mode is far more distinct for this prior than for the uniform prior shown in Table 8.

Maximal Simplices	Posterior Probability
{3, 4, 5}{1, 2}{2, 6}{1, 6}	0.824
{1, 2, 6}{3, 4, 5}	0.111
{1, 2, 6}{3, 4}{3, 5}{4, 5}	0.002

TABLE 9

Highest posterior factorizations with Strauss prior (true model is shown in bold).

In a further experiment with $\gamma = 0.35$, the posterior was concentrated on factorizations without any triads.

5.4. Example 4: \mathcal{G} Outside the Space Generated by \mathcal{A}

In the simulation studies of Sections 5.1–5.3 the class of sets \mathcal{A} used to compute the nerve was known. In this example we investigate the behavior of our methodology when the class of convex sets used to fit the model differs from that used to generate the true graph. We consider three possibilities: $\mathcal{A} = \text{Alpha in } \mathbb{R}^2$, $\mathcal{A} = \text{Alpha in } \mathbb{R}^3$ and $\mathcal{A} = \check{\text{Cech in } \mathbb{R}^2}$. We performed two experiments: one when the graph is feasible for each the three classes, and another example where the graph could be generated by only two of the classes.

First consider a model with junction tree factorization:

$$f_{\theta}(\mathbf{x}) = \frac{\psi_{\theta}(x_1, x_3)\psi_{\theta}(x_2, x_3, x_4)\psi_{\theta}(x_5)}{\psi_{\theta}(x_3)}, \quad (5.5)$$

whose conditional independence structure given by the graph of Figure 19. Again, the clique marginals are specified as a Clayton copula with $\theta = 4$. We simulated 300 samples from this distribution.

We fitted the model with each of the three classes of convex sets using the Metropolis Hastings algorithm of Section 4.3 with random walk proposals on \mathbb{B}^d (where $d = 2$ or 3 , depending on \mathcal{A}). Algorithm 1 was used to enforce decomposability, using $r = 0.40$ and $\eta = 0.020$. The same exponential prior and uniform reflecting random-walk proposals for θ were used as in Example 5.1. Results of 1000 samples after a burn-in period of 50000 draws are summarized in Table 10. Not surprisingly, the posterior mode coincided with the true model in all three cases.

The second model we considered has junction tree factorization:

$$f_{\theta}(\mathbf{x}) = \frac{\psi_{\theta}(x_1, x_2, x_4)\psi_{\theta}(x_1, x_3, x_4)\psi_{\theta}(x_1, x_4, x_5)}{(\psi_{\theta}(x_1, x_4))^2}. \quad (5.6)$$

The corresponding graph cannot be obtained from an Alpha complex in \mathbb{R}^2 , but it is feasible for an Alpha complex in \mathbb{R}^3 (Figure 20) or a $\check{\text{Cech}}$ complex in \mathbb{R}^2 . Using the same Clayton clique marginals before, we sampled 300 observations from this distribution and fitted the model using the three

Nerve	HPP Models	Posterior
Alpha in \mathbb{R}^2	[1, 3][2, 3, 4][5]	0.964
	[1, 3][2, 3, 4][1, 5]	0.012
	[1, 3, 4][2, 3, 4][5]	0.012
Alpha in \mathbb{R}^3	[1, 3][2, 3, 4][5]	0.982
	[1, 3][2, 3][3, 4][5]	0.011
	[1][2, 3, 4][5]	0.003
Čech in \mathbb{R}^2	[1, 3][2, 3, 4][5]	0.595
	[1, 2, 3, 4][5]	0.179
	[1, 2, 3, 4, 5]	0.168

TABLE 10

Models with highest posterior probability. The table is divided according to the class of convex sets used when fitting the model. The true model is shown in bold.

classes of convex sets. Results from 1 000 samples after a burn-in period of 75 000 are summarized in Table 11. Observe that for Alpha complexes in \mathbb{R}^2 , there is no clear posterior mode (unlike the previous example, or Sections 5.1 and 5.3). The posterior mode for the Čech complex in \mathbb{R}^2 and the Alpha complex in \mathbb{R}^3 both match the true model.

Nerve	HPP Models	Posterior
Alpha in \mathbb{R}^2	[1,2][1,3,4][1,4,5]	0.214
	[1,2,4][1,3,4][1,3,5]	0.115
	[1,2,4][1,3,4][3,4,5]	0.112
Alpha in \mathbb{R}^3	[1,2,4][1,3,4][1,4,5]	0.976
	[1,2,3,4][1,4,5]	0.016
	[1,2,4][1,3][1,4,5]	0.009
Čech in \mathbb{R}^2	[1,2,4][1,3,4][1,4,5]	0.758
	[1,2,4][1,3,4][1,3,5]	0.177
	[1,2,4][1,3,4][4,5]	0.148

TABLE 11

Models with highest posterior probability, for each class of convex sets. The true model (shown in bold) is unattainable for Alpha complexes in \mathbb{R}^2 .

6. Discussion

In this article we present a new parametrization of graphs by associating them with finite sets of points in \mathbb{R}^d . This perspective supports the design of informative prior distributions on graphs using familiar probability dis-

tributions of point sets in \mathbb{R}^d . It also induces new and useful Metropolis/Hastings proposal distributions in graph space that include both local and global moves. As suggested by Helly's Theorem [10] characterizing the sparsity of intersections of convex sets in \mathbb{R}^d , this methodology is particularly well suited for sparse graphs. The simple strategies presented here generalize easily to more detailed and subtle models for both priors and M/H proposals.

An interesting feature of our approach is that the distribution on the space of graphs is modeled directly before the application of any specific hyper Markov law, in contrast to standard approaches in which it is the hyper Markov law that is used to encourage sparsity or other features on the graph. We think that working with the space of graphs explicitly opens a lot of possibilities for prior specification in graphical models, therefore, it is a perspective worth further study.

Interesting questions and extensions of this idea include: (1) achieving a deeper and more detailed understanding of the subspace of graphs spanned by different specific filtrations; (2) designing priors to control the distributions of specific features of graphs such as clique size or tree width; (3) modeling directed acyclic graphs (DAGs), and (4) concrete implementation of novel Markov structures based on nerves.

This methodology generates only graphs that are feasible for the particular filtration chosen. Although we do have some insight about which graphs can and cannot be generated by a specific filtrations, a more complete and formal understanding of this aspect of the problem would be useful.

We used very simple prior distributions for the purpose of illustrating the core idea of the methodology. It is natural in this approach to incorporate tools from point processes into graphical models to define new classes of priors for graph space. Future developments in our research will involve a range of repulsive and cluster processes.

The parametrization we propose can be used to represent Markov struc-

tures on DAGs, but the strategies for obtaining such graphs from nerves will be different and will establish stronger connections between Graphical Models and computational topology.

The present work is related to that of Pistone *et al.* [27] in which a nerve of convex subsets in \mathbb{R}^d is used to obtain Markov structures for a distribution, an extension of the abstract tube theory of Naiman and Wynn [23]. This new perspective allows for constructions that generalize the idea of junction trees. By modifying our methodology according to this framework (personal communication from H. Wynn suggests that this is feasible) we hope to fit models that factorize according to those novel Markov structures.

Another possible extension of this work is to discretize the set from which the vertex set is sampled (*e.g.*, use a grid). Such discretization may improve the behaviour of the MCMC; it would also allow the use of a nonparametric form for the prior on the vertex set, leading to more flexible priors on graph space.

Acknowledgments

We are grateful to Herbert Edelsbrunner, John Harer, and Henry Wynn for helpful conversations. This work was partially supported by National Science Foundation grants DMS-0732260, DMS-0635449 and DMS-0757549, and by National Institutes of Health grants NIH R01 CA123175-01A1 and NIH P50-GM081883-02. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF or the NIH.

References

- [1] ATAY-KAYIS, A. and MASSAM, H. (2005). A Monte Carlo method to compute the marginal likelihood in non decomposable graphical Gaussian models. *Biometrika* **92** 317–335.

- [2] BESAG, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Ser. B: Statistical Methodology* **36** 192–236.
- [3] BESAG, J. E. (1975). Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society, Ser. D: The Statistician* **24** 179–195.
- [4] BROOKS, S. P., GIUDICI, P. and ROBERTS, G. O. (2003). Efficient construction of reversible jump markov chain Monte Carlo proposal distributions. *Journal of the Royal Statistical Society, Ser. B: Statistical Methodology* **65** 3–55.
- [5] CARVALHO, C. M., MASSAM, H. and WEST, M. (2007). Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika* **94** 719–733.
- [6] CLAYTON, D. G. (1978). A model for association in bivariate life tables and its applications in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65** 141–151.
- [7] CLIFFORD, P. (1990). Markov random fields in statistics. In *Disorder in Physical Systems: A Volume in Honour of John M. Hammersley* (G. Grimmett and D. Welsh, eds.), 19–32. Oxford University Press, Oxford, UK.
- [8] DAWID, A. P. and LAURITZEN, S. L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *Annals of Statistics* **21** 1272–1317.
- [9] EDELSBRUNNER, H. and HARER, J. (2008). Persistent homology— a survey. In *Surveys on Discrete and Computational Geometry: Twenty Years Later* (J. E. Goodman, J. Pach and R. Pollack, eds.), *Contemporary Mathematics*, vol. 453, 257–282. American Mathematical Society.
- [10] EDELSBRUNNER, H. and HARER, J. (2009). Lecture notes from the course ‘Computational Topology’. Available on-line at <http://www.cs.duke.edu/courses/fall106/cps296.1/>.
- [11] EDELSBRUNNER, H. and MÜCKE, E. P. (1994). Three-dimensional

- alpha shapes. *ACM Transactions on Graphics* **13** 43–72.
- [12] GIUDICI, P. and GREEN, P. J. (1999). Decomposable graphical Gaussian model determination. *Biometrika* **86** 785–801.
- [13] GREEN, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–732.
- [14] HAMMERSLEY, J. M. and CLIFFORD, P. (1971). Markov fields on finite graphs and lattices. Unpublished; see however Clifford [7].
- [15] HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109.
- [16] HECKERMAN, D., GEIGER, D. and CHICKERING, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning* **20** 197–243.
- [17] HOFF, P. D. (2007). Extending the rank likelihood for semiparametric copula estimation. *Annals of Applied Statistics* **1** 265–283.
- [18] HUBER, M. L. and WOLPERT, R. L. (2009). Likelihood-based inference for Matérn type III repulsive point processes. *Advances in Applied Probability* **41** In press. Preprint available on-line at <http://www.stat.duke.edu/ftp/pub/WorkingPapers/08-27.html>.
- [19] JONES, B., CARVALHO, C., DOBRA, A., HANS, C., CARTER, C. K. and WEST, M. (2005). Experiments in stochastic computation for high-dimensional graphical models. *Statistical Science* **20** 388–400.
- [20] LAURITZEN, S. L. (1996). *Graphical Models, Oxford Statistical Science Series*, vol. 17. Oxford University Press, Oxford, UK.
- [21] MANSINGHA, V. K., KEMP, C., TENENBAUM, J. B. and GRIFFITHS, T. L. (2006). Structured priors for structure learning. In *Uncertainty in Artificial Intelligence: Proceedings of the Twenty-second Annual Conference (UAI 2006)* (L. Bertossi, A. Hunter and T. Schaub, eds.).
- [22] MUKHERJEE, S. and SPEED, T. P. (2008). Network inference using informative priors. *Proceedings of the National Academy of Science* **105**

14313–14318.

- [23] NAIMAN, D. Q. and WYNN, H. P. (1997). Abstract tubes, improved inclusion exclusion identities and inequalities and importance sampling. *Annals of Statistics* **25** 1954–1983.
- [24] NELSEN, R. B. (1999). *An Introduction to Copulas*. Springer-Verlag, New York, NY.
- [25] PENROSE, M. D. (2003). *Random Geometric Graphs*. Oxford University Press, New York, NY.
- [26] PENROSE, M. D. and YUKICH, J. E. (2001). Central limit theorems for some graphs in computational geometry. *Annals of Applied Probability* **11** 1005–1041.
- [27] PISTONE, G., WYNN, H., DE CABEZÓN, G. S. and SMITH, J. Q. (2009). Junction tubes and improved factorisations for Bayes nets. (unpublished).
- [28] ROBERT, C. P. (2001). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer-Verlag, New York, NY, second edn.
- [29] ROBERT, C. P. and CASELLA, G. (2004). *Monte Carlo Statistical Methods*. Springer-Verlag, New York, NY, second edn.
- [30] ROVERATO, A. (2002). Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scandinavian Journal of Statistics* **29** 341–411.
- [31] SCOTT, J. G. and CARVALHO, C. M. (2008). Feature-inclusion stochastic search for Gaussian graphical models. *Journal of Computational and Graphical Statistics* **17** 790–808.
- [32] SISSON, S. A. (2005). Transdimensional Markov chains: A decade of progress and future perspectives. *Journal of the American Statistical Association* **100** 1077–1089.
- [33] STRAUSS, D. J. (1975). A model for clustering. *Biometrika* **62** 467–476.
- [34] WONG, K. K. F., CARTER, C. K. and KOHN, R. (2003). Hyper inverse

Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Biometrika* **90** 809–830.

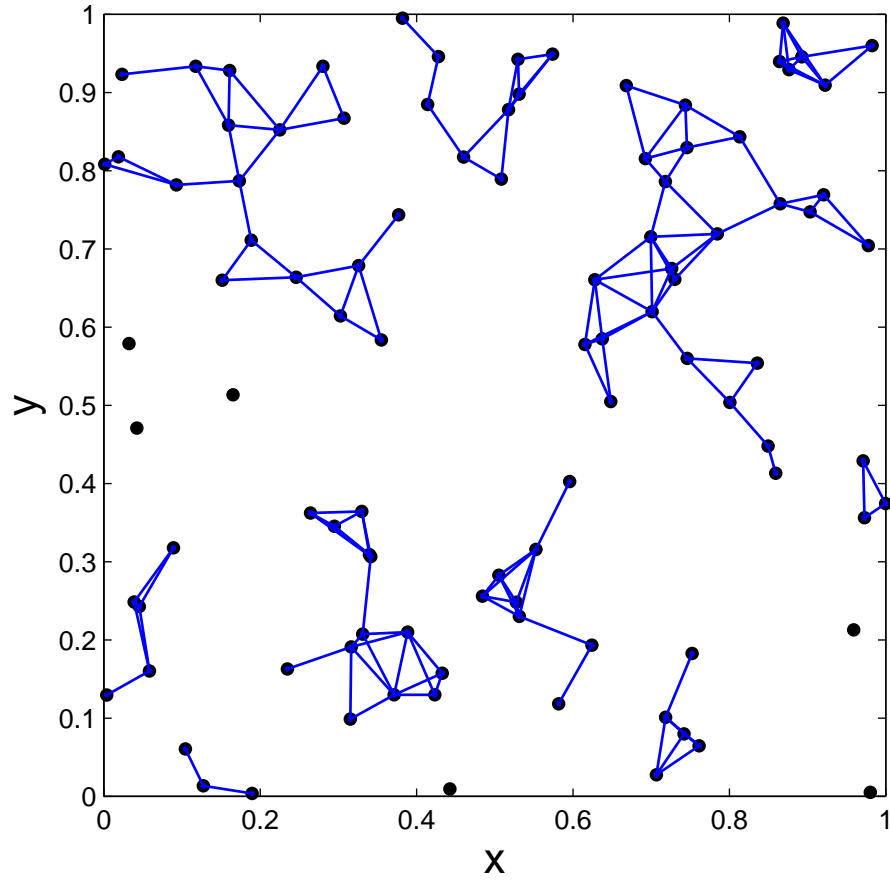


FIG 1. Proximity graph for 100 vertices and radius $r = 0.05$.

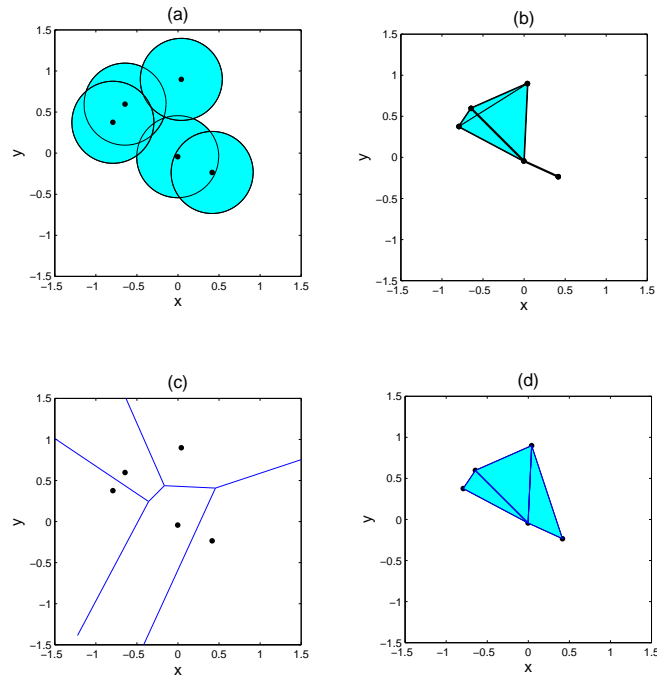


FIG 2. (a) A set of vertices in \mathbb{R}^2 are used to construct a family of disks of radius $r = 0.5$. (b) The nerve of this convex set. This is an example of a Čech complex. (c) For the same vertex set the Voronoi diagram is computed. (d) The nerve of the Voronoi cells is obtained. This is an example of the Delaunay triangulation. Note that the maximum clique size of the Delaunay is bounded by the dimension of the space of the vertex set plus one; such a restriction does not apply to the Čech complex.

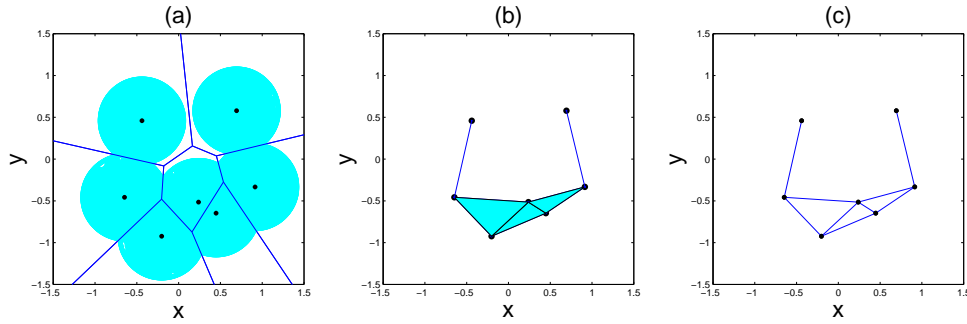


FIG 3. (a) Given a set of vertices and a radius ($r = 0.5$) one can compute $A_i = C_i \cap B_i$, where C_i is the Voronoi cell for vertex i and B_i is the ball of radius r centered at vertex i . (b) The Alpha complex is the nerve of the A_i 's. (c) Often the main interest will be the 1-skeleton of the complex, which is the subset of the nerve that corresponds to (nonempty) pairwise intersections.

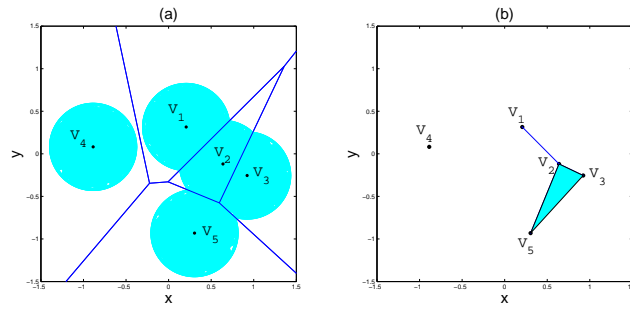


FIG 4. (a) Alpha complex computed using $r = 0.5$ and the vertex set listed in Table 1. (b) The corresponding factorization.

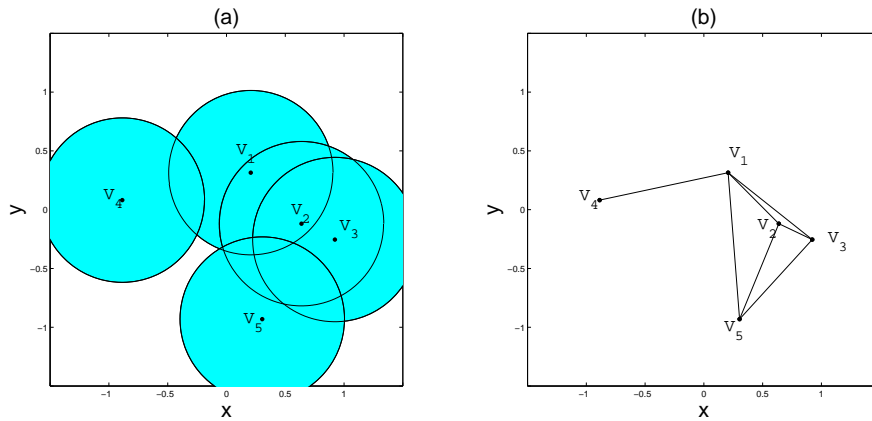


FIG 5. (a) Čech complex computed using $r = 0.7$ and the vertex set listed in Table 1. (b) The 1-Skeleton of the Čech complex.

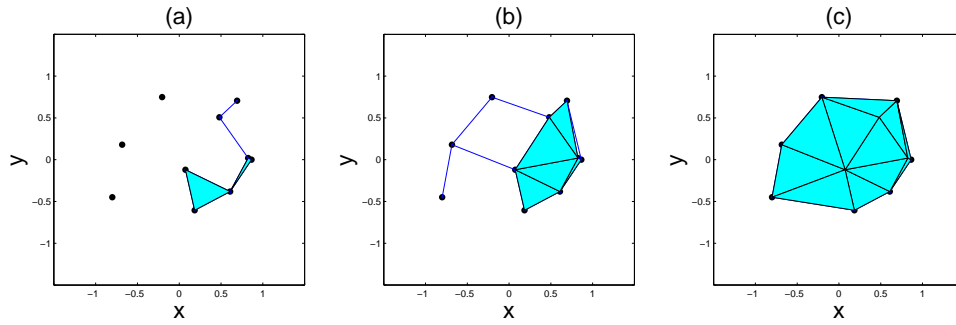


FIG 6. Filtration of Alpha complexes, (a) $r = \sqrt{0.10}$, (b) $r = \sqrt{0.20}$ and (c) $r = \sqrt{0.75}$.

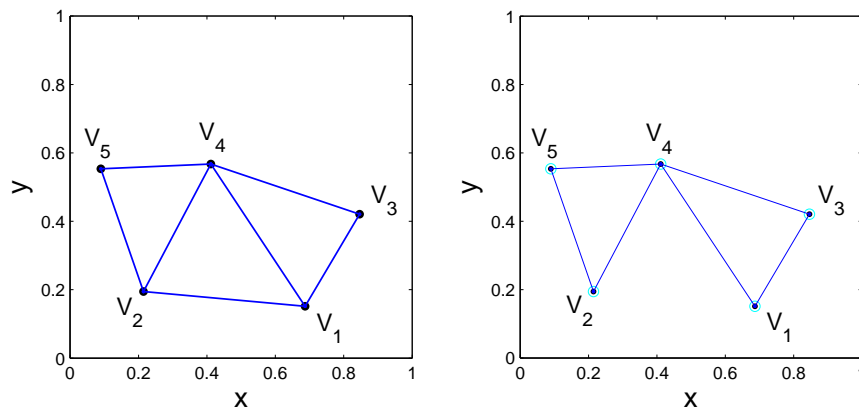


FIG 7. (a) Proximity graph computed from the vertex set given in Table 2. (b) The decomposable graph computed from the same vertex set using Algorithm 1. The edge (1,2) is not included in the decomposable graph.

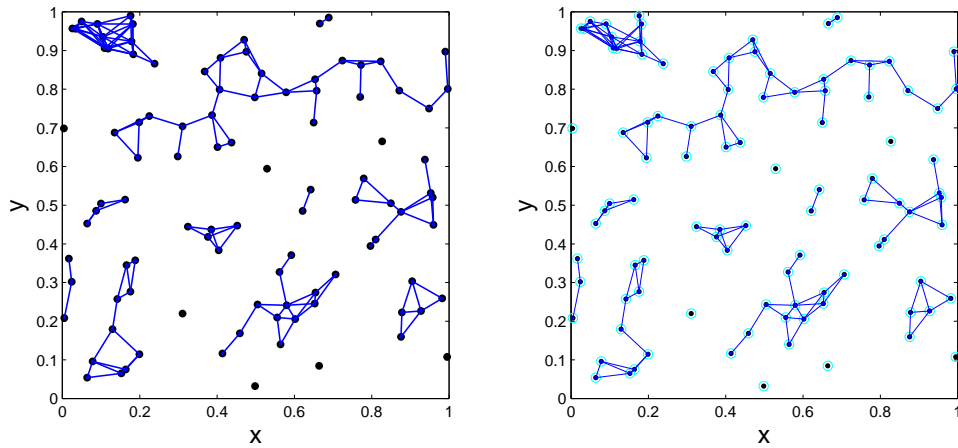


FIG 8. (a) The 1-Skeleton of Čech complex given the displayed point set and $r = 0.05$. (b) The decomposable graph for the same complex, point set, and radius output by Algorithm 1.

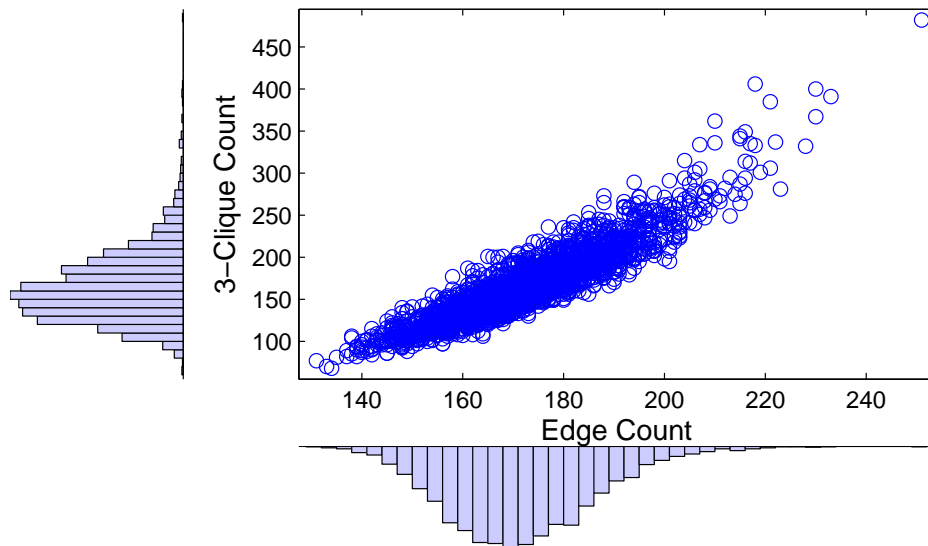


FIG 9. Edge counts and 3-Clique counts from 2,500 simulated samples of $\mathcal{G}(\mathbf{V}, 1/\sqrt{2} \cdot 75, \check{C}ech)$ where $|\mathbf{V}| = 75$ and $V_i \stackrel{iid}{\sim} Un([0, 1]^2)$, $1 \leq i \leq 75$. The multivariate normal appears as a reasonable approximation for the joint distribution, as suggested by [25, Thm. 3.13]. Čech radius is $r_n = 1/\sqrt{2n}$.

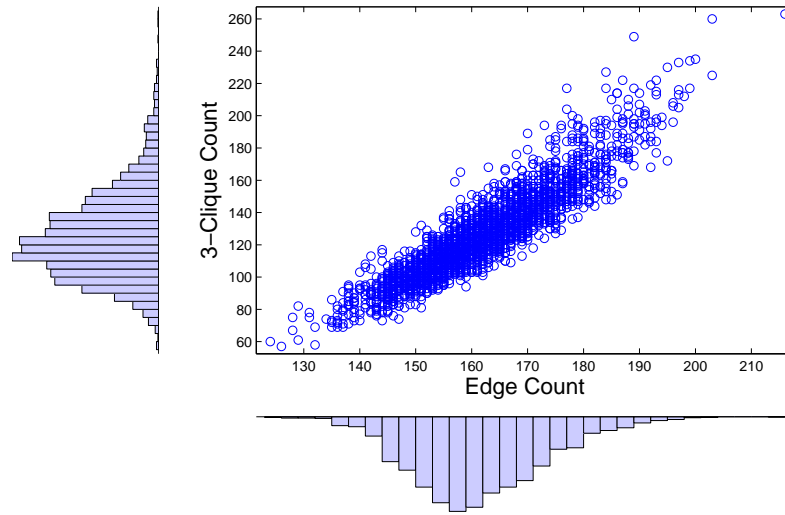


FIG 10. *Edge counts and 3-Clique counts from 2,500 simulated samples of $\mathcal{G}(\mathbf{V}, 1/\sqrt{2} \cdot 75, \check{C}ech)$ where $|\mathbf{V}| = 75$ and V sampled from a Matérn III with hard-core radius $r = 0.35$.*

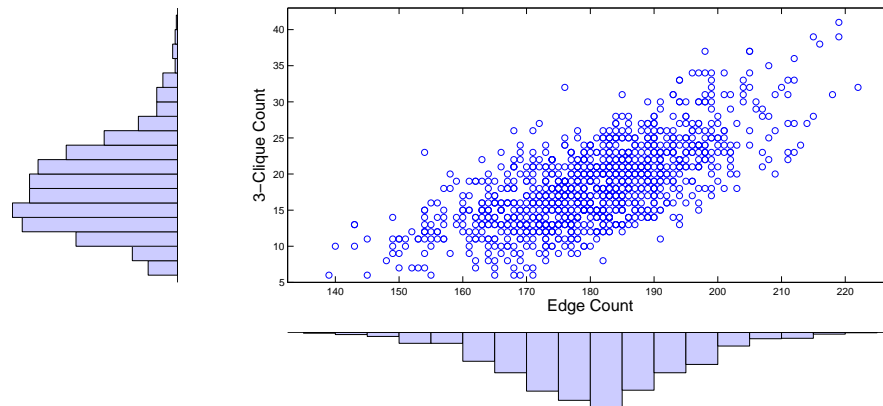


FIG 11. *Edge counts and 3-Clique counts from 1000 simulated samples of an Erdős-Rényi graph with edge inclusion probability of $p = 0.065$.*

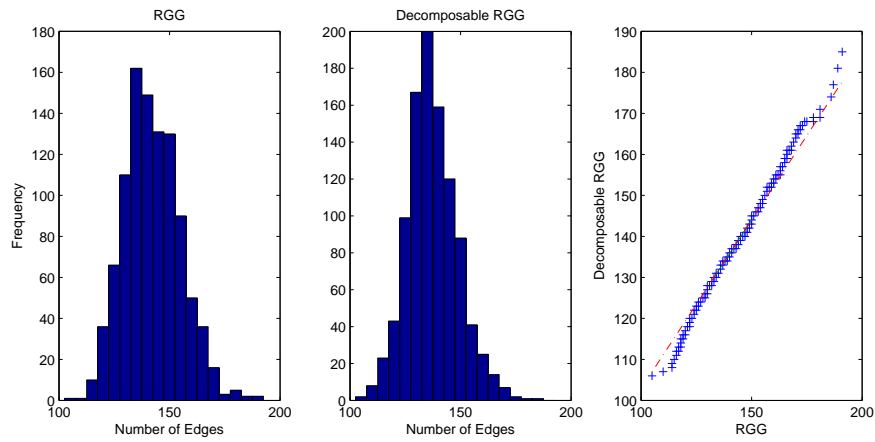


FIG 12. *Distribution of edge counts for both unrestricted and decomposable graphs. Graphs were computed using Čech complex filtrations with $n = 100$ and $V_i \stackrel{iid}{\sim} Un([0, 1]^2)$.*

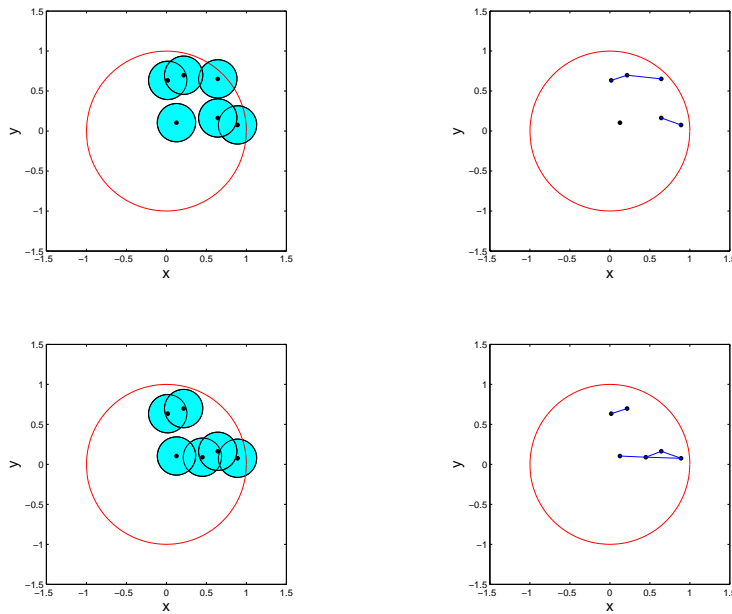


FIG 13. *This figure illustrates a local move. (a) The current configuration of the points. (b) The graph implied by this configuration. (c) The proposal configuration which is obtained by randomly moving one vertex. (d) The graph implied by the proposed move.*

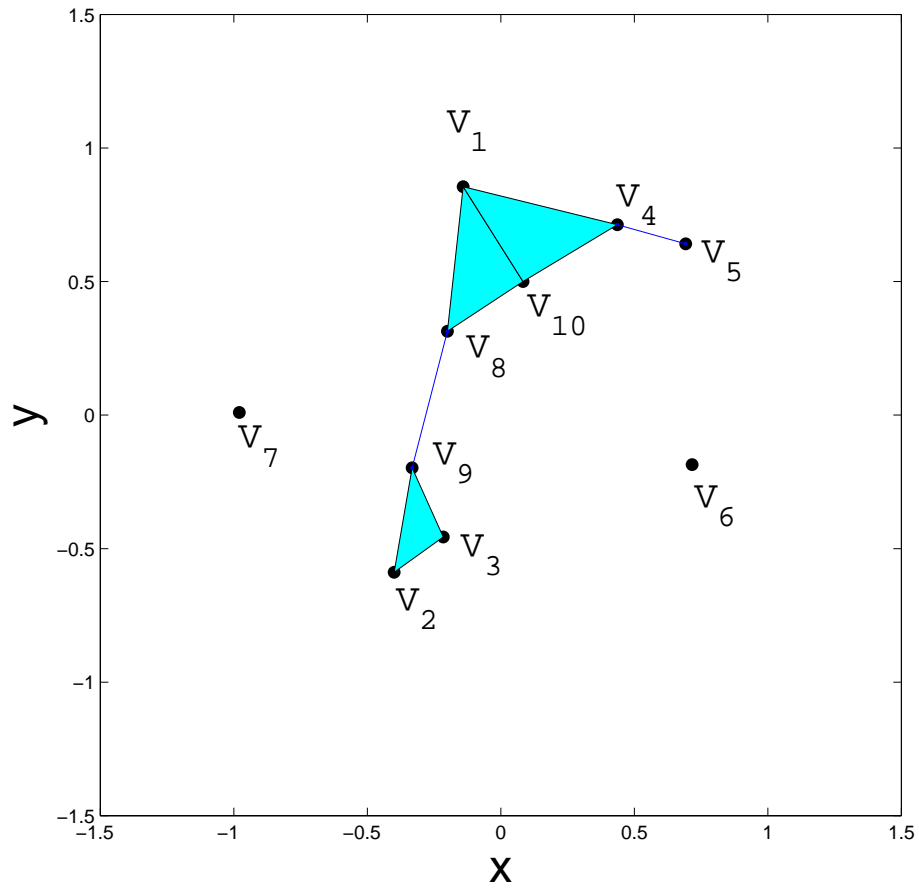


FIG 14. Geometric graph representing the model given in (5.1a). For this example graphs are constructed to be decomposable and the clique marginals are specified as Clayton copulas.

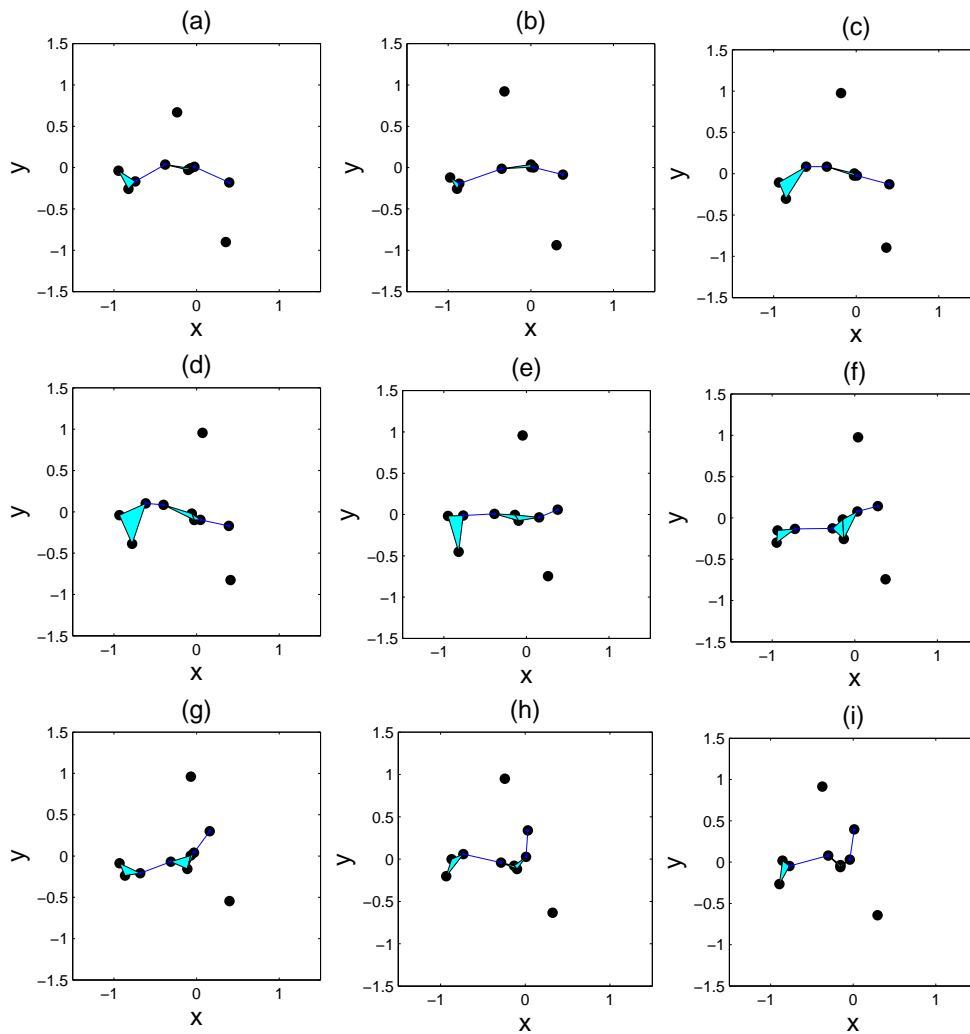


FIG 15. Geometric graphs corresponding to snapshots of posterior samples (one every 100 iterations) from model of (5.1a). For this example graphs are constructed to be decomposable and the clique marginals are specified as Clayton copulas.

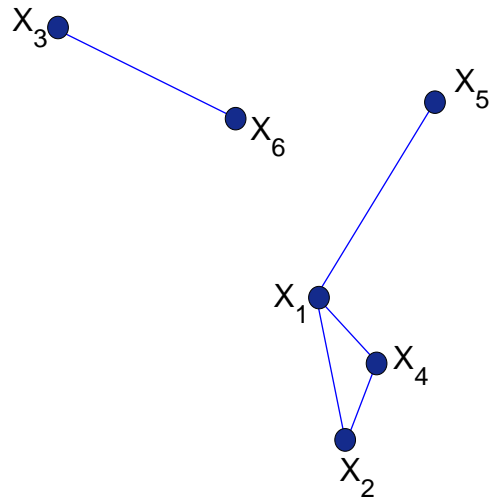


FIG 16. Graph encoding the Markov structure of the model given by precision matrix (5.3).

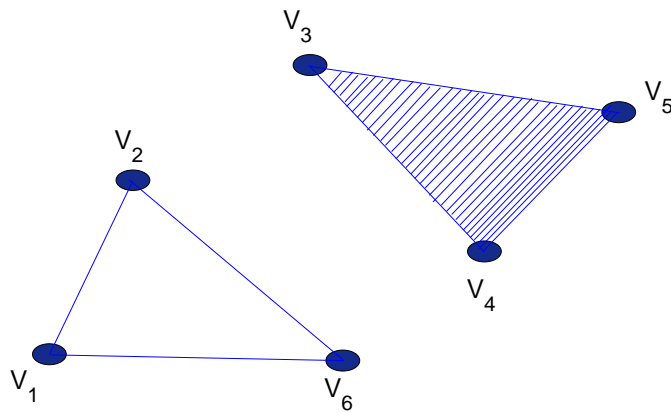


FIG 17. Graph encoding the Markov structure of the model given in (5.4).

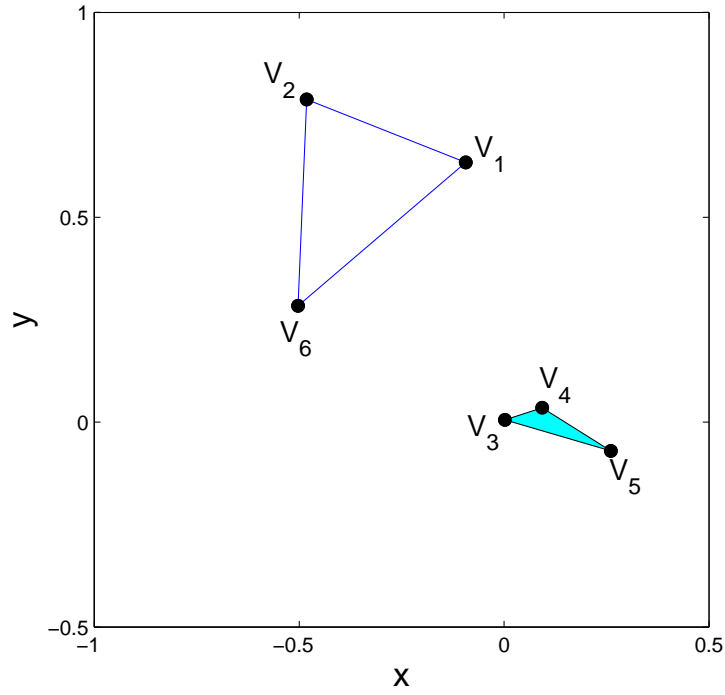


FIG 18. Alpha complex corresponding to the vertex set in Table 7 and $r = \sqrt{0.075}$.

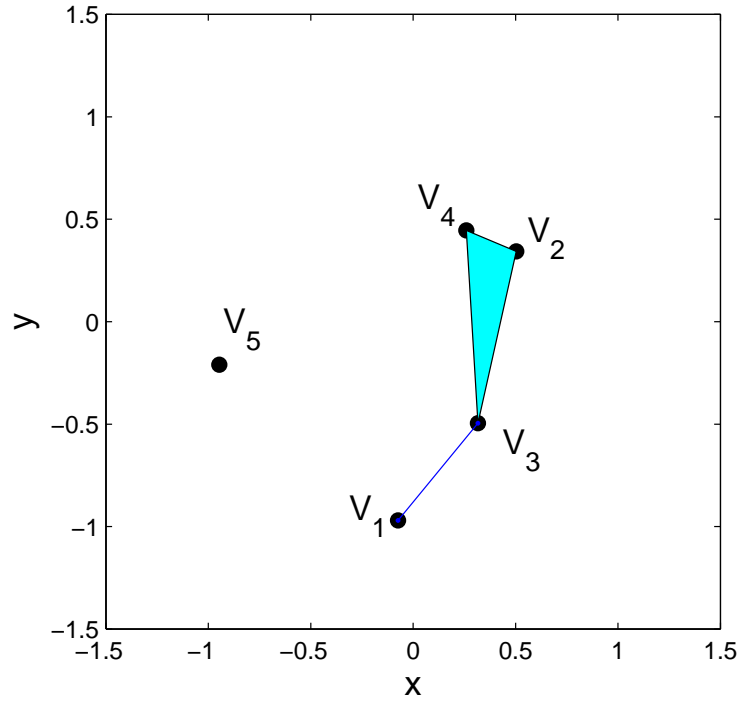


FIG 19. Graph encoding the Markov structure of the model given in (5.5).

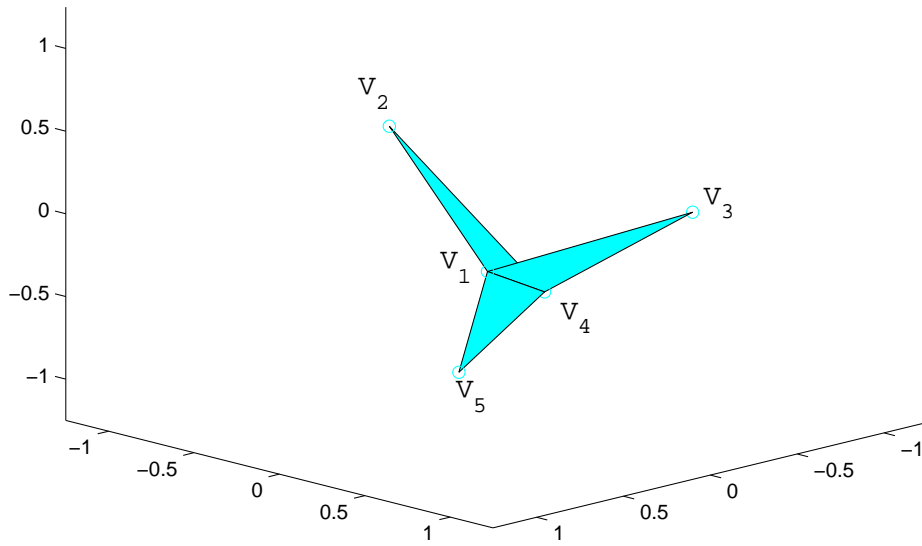


FIG 20. Graph encoding the Markov structure of the model given in (5.6).