Applied Probability Trust (16 April 2023)

SUPPLEMENTARY MATERIAL: ON A WIDER CLASS OF PRIOR DISTRIBUTIONS FOR GRAPHICAL MODELS

ABHINAV NATARAJAN,* *University of Oxford*

WILLEM VAN DEN BOOM,** *** *AND*

KRISTOFORUS BRYANT ODANG,** **** *National University of Singapore*

MARIA DE IORIO,***** *National University of Singapore; Singapore Institute for Clinical Sciences, A*STAR; University College London*

S1. Proofs for propositions from the main text

Proof of Proposition 1. Note that the sum of degrees over every vertex of a graph is even. Hence, k is an even number, else the sum of the degrees would be odd. Number the odd-degree vertices as $v_1, v_2, \ldots, v_k$. Now, add (using $\oplus$) edges $(v_i, v_{k/2+i})$, $i = 1, 2 \ldots k/2$, to G to obtain G_C. By construction, G_C has only even degree vertices such that $G_C \in \mathcal{C}_n$ by Property P2. Thus, there exists a graph in $\mathcal{C}_n$ that differs from G by $k/2$ edges.

We prove the second part of the proposition by contradiction. Assume that

Received 19 April 2021; revision received 4 April 2023.

* Postal address: Mathematical Institute, University of Oxford, Wellington Square, Oxford OX1 2JD, United Kingdom. Email: natarajan@maths.ox.ac.uk

** Postal address: Yong Loo Lin School of Medicine, National University of Singapore, 10 Medical Dr, Singapore 117597.

*** Email address: vandenboom@nus.edu.sg

**** Email address: kristoforusbryant@u.yale-nus.edu.sg

***** Postal address: Yong Loo Lin School of Medicine, National University of Singapore, 10 Medical Dr, Singapore 117597. Email: mdi@nus.edu.sg

there exists a graph $G_C \in \mathcal{C}_n$ that differs from G in $l < k/2$ edges. Flipping these l edges in G to obtain G_C will affect the degree of at most $2l$ vertices since each edge involves two vertices. Thus, G_C shares at least $k - 2l > 0$ odd-degree vertices with G which contradicts $G_C \in \mathcal{C}_n$ by Property P2. $\square$

Proof of Proposition 2. Since the cycle space is a vector space with basis $C = \{c_1, \ldots, c_r\}$, each element of the cycle space is in bijective correspondence with a binary sequence of length r. Each binary element of this sequence corresponds to the inclusion of the respective cycle basis element. As the cycle inclusions are independent with probability 0.5, every binary sequence is equally probable. The marginal edge inclusion probability of 0.5 follows from Proposition 3 with $p = 0.5$. $\square$

Proof of Proposition 3. Denote $\{c_1, \ldots, c_r\}$ by C_e. The edge e is included if and only if an odd number of cycles from C_e are included. Suppose $C_k \subset C_e$ is a subset with the odd number of elements $|C_k| = 2k + 1$. Then, the probability that the cycles in C_k are included and the cycles in $C_e \setminus C_k$ are not is

$$p(C_k) = \left(\prod_{c_i \in C_k} p_i \right) \left(\prod_{c_i \in C_e \setminus C_k} (1 - p_i) \right).$$

By summing this probability over all $C_k \subset C_e$, we get the joint probability of edge e and including exactly $2k + 1$ cycles as $\sum_{C_k \subset C_e} p(C_k)$. Note that this is exactly the coefficient of x^{2k+1} in $f(x)$. Then, summing this coefficient over $k = 0, \ldots, \lfloor (r-1)/2 \rfloor$ yields the desired result.

Next, consider the special case where $p = p_1 = \cdots = p_r$. Then, $f(x) = (1 - p + px)^r = \sum_{i=0}^{r} \binom{r}{i} (1-p)^{r-i} (px)^i$ where the last equality follows from the binomial theorem. Thus, the probability of inclusion of the edge e is $\sum_{i \text{ odd}} \binom{r}{i} (1-p)^{r-i} p^i$. By the binomial theorem, $\sum_{i=0}^{r} \binom{r}{i} (1-p)^{r-i} p^i = (1 - p + p)^r = 1$ and $\sum_{i=0}^{r} \binom{r}{i} (1-p)^{r-i} (-p)^i = (1 - p - p)^r = (1 - 2p)^r$. Also,

$\sum_{i=0}^{r} \binom{r}{i}(1-p)^{r-i}p^i - \sum_{i=0}^{r} \binom{r}{i}(1-p)^{r-i}(-p)^i = 2\sum_{i \text{ odd}} \binom{r}{i}(1-p)^{r-i}p^i$ since

$$(1-p)^{r-i}p^i - (1-p)^{r-i}(-p)^i = \begin{cases} 0, & i \text{ even} \\ 2(1-p)^{r-i}p^i, & i \text{ odd} \end{cases}$$

Therefore, $\sum_{i \text{ odd}} \binom{r}{i}(1-p)^{r-i}p^i = \{1-(1-2p)^r\}/2$. □

Proof of Corollary 1. Let $\{v_0, \dots, v_{n-1}\}$ be the vertices such that the star tree generating the basis C is rooted at v_0. We refer to edges of the form (v_i, v_j) where $i, j \neq 0$ as *peripheral edges*. We refer to edges of the form (v_0, v_i) as *rooted edges*.

First, consider a rooted edge (v_0, v_i). Exactly $(n-2)$ edges in the complement $\overline{T}$ of T have v_i as an endpoint. Each of these edges correspond to their own unique basis element in C. No other basis elements in C contain (v_0, v_i). Thus, $r = n-2$ basis elements contain (v_0, v_i). On the other hand, each peripheral edge is in $\overline{T}$ and is thus contained in only $r = 1$ basis element due to the bijection between edges in $\overline{T}$ and elements of C. The required result for a given T now follows from Proposition 3.

For the result marginalising over the uniform distribution over all star trees, note that there are n star trees on n vertices. Each edge is rooted in 2 of these trees and peripheral in the other $(n-2)$. Thus, the marginal edge inclusion probability follows as

$$\begin{aligned}\mathbb{P}\{(i,j) \in E\} = \\ &\mathbb{P}\{(i,j) \text{ is peripheral}\}\mathbb{P}\{(i,j) \in E \mid (i,j) \text{ is peripheral}\} \\ &+ \mathbb{P}\{(i,j) \text{ is rooted}\}\mathbb{P}\{(i,j) \in E \mid (i,j) \text{ is rooted}\} \\ &= \frac{n-2}{n}p + \frac{2}{n}\{1-(1-2p)^{n-2}\}/2\end{aligned}$$

which simplifies to the required result. □

Proof of Proposition 4. Using the independence of the cycle inclusions, it is seen that for any $I \subset \{1, \dots, r\}$ the joint inclusion probability of the cycles

$C_I = \{c_i : i \in I\}$ is given by the coefficient of $\prod_{i\in I} t_i$ in f. The inclusion of C_I leads to the inclusion of the edges $\{e_{a(i)}, e_{b(i)} : i \in I\}$, notwithstanding cancellations due to intersecting cycles. The probability generating function $h \in \mathbb{R}[x_1, \ldots, x_m]$ for the edge inclusions (with numerosity and ignoring binary cancellations) is therefore given by replacing each t_i in f with $x_{a(i)}x_{b(i)}$. Binary cancellations are then handled by reducing all exponents modulo 2, which is equivalent to taking the image of h in the quotient ring. □

Proof of Proposition 5. For odd k, $\mathbb{P}\{\deg(v_i) = k\} = 0$ by Property P2. The remainder of this proof thus considers even k.

Note that each edge (v_i, v_j), $i, j \neq 0$, is present in exactly one cycle basis element whose vertices are $\{v_0, v_i, v_j\}$. Therefore, the inclusion of edge (v_i, v_j) is equivalent to the inclusion of the cycle $\{v_0, v_i, v_j\}$. Fix $i > 0$. If v_i has degree $k > 0$, then it is connected to either (i) k other vertices that are not the root vertex, or (ii) the root vertex and $k - 1$ other vertices.

(i) Consider the first case. For each of the k vertices connected to v_i via an edge, we must have the inclusion of the corresponding cycle, and for each of the $n-2-k$ vertices not connected to v_i via an edge, the corresponding cycle cannot be included. This happens with probability $p^k(1-p)^{n-2-k}$ where p is the inclusion probability for any cycle basis element. The inclusion of exactly k cycles containing the edge (v_0, v_i) ensures that this edge is not included since k is even. Since there are $\binom{n-2}{k}$ ways to pick k non-root vertices that are connected to v_i, the probability of v_i being connected to k non-root vertices is $\binom{n-2}{k}p^k(1-p)^{n-2-k}$.

(ii) In the second case, by a similar argument to the above, there are exactly $k-1$ cycles of the form $\{v_0, v_i, v_j\}$ that are included. This happens with probability $\binom{n-2}{k-1}p^{k-1}(1-p)^{n-k-1}$. Since k is even, this results in the edge $\{v_0, v_i\}$ being included.

Adding the probabilities of (i) and (ii) yields the required result for $k \geq 2$. For

$k = 0$, only (i) is possible such that $\mathbb{P}\{\deg(v_i) = k\} = \binom{n-2}{k} p^k (1-p)^{n-2-k} = (1-p)^{n-2}$. □

Proof of Proposition 6. Recall the definition of peripheral and rooted edges from the proof of Corollary 1. The lower bound $|E| \geq q$ follows from the observation that there are exactly q peripheral edges included in G. The remainder of this proof thus considers the upper bound.

The maximum of three edges per cycle ($|E| = 3q$) is attained when there are no intersections between the included cycles. Including a cycle $\{v_0, v_i, v_j\}$ corresponds to choosing a pair $\{v_i, v_j\}$ from the $(n-1)$ non-root vertices. Avoiding intersections between included cycles corresponds to choosing all such pairs disjoint. By the pigeonhole principle, there can be at most m mutually disjoint pairs. Thus, $|E| = 3q$ is attainable if and only if $q \leq m$. Also, $|E| \leq 3q$ for $q \leq m$.

By Property P2, the degree of v_0 and thus the number of rooted edges in G must be even. The largest even number less than the number $(n-1)$ of non-root vertices is $2m$. Additionally, there are q peripheral edges. Thus, $|E| \leq q + 2m$. □

S2. Edge union of spanning trees

We explore the prior induced by edge unions of spanning trees in this section. As seen in Højsgaard et al. (2012), Schwaller et al. (2019) and Duan & Dunson (2021), the spanning tree structure has proved relevant in graphical model inference. We thus consider defining a graph prior $p(G \mid k)$ where the graph $G = (V, E)$ is an edge union of k spanning trees, i.e. $E = \cup_{i=1}^{k} T_i$ where T_i is the set of edges of a spanning tree. We assume that the inclusion of each spanning tree is equally likely and independent a priori. Then, to infer the induced posterior distribution of the graphical model using Metropolis-Hastings, the following ratio needs to be computed:

$$\frac{p(E' = \cup_{i=1}^{k'} T_i \mid k')}{p(E = \cup_{i=1}^{k} T_i \mid k)} = \frac{|Y(G', k')| / \binom{|\tau(G')|}{k'}}{|Y(G, k)| / \binom{|\tau(G)|}{k}}$$

where the prime marks proposed values in the MCMC, $|Y(G, k)|$ is the number of ways to write G as the edge union of k distinct spanning trees, $k' \in \{k-1, k+1\}$, and $|\tau(G)|$ is the number of all spanning trees of G. Note that by Kirchoff's matrix tree theorem (Theorem S1 on page 7), $|\tau(G)|$ can be computed in $O(n^3)$ time. Hence, the term left to be computed is the ratio $|Y(G', k')|/|Y(G, k)|$.

The remainder of this section is structured as follows. Section S2.1 introduces notation. Then, Sections S2.2, S2.3 and S2.4 compute, lower bound and upper bound the ratio $|Y(G', k')|/|Y(G, k)|$, respectively.

S2.1. Notation

In this Section S2, all graphs considered are connected. Let $K \in \mathcal{G}_n$ be the complete graph on the vertex set V. For any graph G, denote the set of spanning trees of G by $\tau(G)$. For any $e \in E$, denote the set of spanning trees of G that contain e by $\tau_e(G) \subset \tau(G)$. Let $\binom{\tau(K)}{k}$ be the set with the $\binom{|\tau(K)|}{k}$ subsets of $\tau(K)$ of size k. There is a map

$$\Phi_k : \binom{\tau(K)}{k} \to \mathcal{G}_n$$

given by taking the edge union of the k trees in each element of $\binom{\tau(K)}{k}$. For $G \in \mathrm{im}(\Phi_k) = \{\Phi_k(\mathcal{T}) : \mathcal{T} \in \binom{\tau(K)}{k}\}$, define the set of combinations of k spanning trees that yield G as $Y(G, k) = \Phi_k^{-1}(G) = \{\mathcal{T} : \Phi_k(\mathcal{T}) = G\}$. Let $G_1 + G_2$ denote the graph obtained by taking the edge union of graphs G_1 and G_2. Our main aim is to estimate the ratio

$$\frac{|Y(G', k')|}{|Y(G, k)|} = \frac{|Y(G + T_0, k + 1)|}{|Y(G, k)|} \tag{S1}$$

for a fixed $G \in \mathrm{im}(\Phi_k)$ and $T_0 \in \tau(K)$. Note that here we define $G' = G + T_0$ and $k' = k + 1$.

S2.2. Direct approach

One approach to estimate this ratio is to compute $|Y(G,k)|$ and $|Y(G',k')|$ separately. First, consider the brute force strategy that enumerates every tree in G. This is infeasible given how the number of spanning trees $|\tau(G)|$ grows super-exponentially with $m=|E|$ (Greenhill et al., 2017) and the enumeration algorithm has complexity $O(n+m+|\tau(G)|)$ (Kapoor & Ramesh, 2000). In this section, we show an algorithm that improves on the super-exponential brute-force algorithm by counting the number of ways to write G as a union of its spanning trees in $O(2^m\, n^3)$ time.

Lemma 1. *Every edge union of k spanning trees of G is either equal to G or is an edge-induced subgraph of G that spans V and is connected.*

Proof. This follows by contradiction since the negation would mean that the spanning tree union contains an edge not in G. □

Proposition 1. *Let G_C denote any connected graph with as set of edges a proper subset of the edge set E of G. Then,*

$$|Y(G,k)| = \binom{|\tau(G)|}{k} - \sum_{G_C} |Y(G_C,k)|$$

if $|\tau(G)| \geq k$ and $|Y(G,k)| = 0$ otherwise.

Proof. Note that $\binom{|\tau(G)|}{k}$ is the number of possible combinations of k distinct spanning trees of G. Further, by Lemma 1, any union of spanning trees of G that is connected and not equal to G is a connected edge-induced subset of G. □

Hence, having proved the recursive formula in (1), we analyse the complexity of its calculation. We recall the following proposition due to Kirchoff (see Chaiken & Kleitman (1978) for a reference).

Theorem S1. (Matrix tree theorem.) *The number $|\tau(G)|$ of spanning trees of G is equal to any cofactor of the Laplacian matrix of G.*

This theorem implies that counting $\tau(G)$ is equivalent to finding the determinant of an $(n-1)\times(n-1)$ matrix which has complexity $O(n^3)$.

Now, we prove that we can compute $|Y(G,k)|$ in $O(2^m\, n^3)$ time.

Proposition 2. *$|Y(G,k)|$ can be calculated in $O(2^m\, n^3)$ time.*

Proof. Consider the recursive scheme implied by Proposition 1. Since every connected edge-induced subgraph of every G_C is also an edge-induced subgraph of G, using dynamic programming and memoisation, we only need to calculate every unique $|Y(G_C,k)|$ of which there are at most 2^m. Calculating $|\tau(G)|$ at each step involves calculating a determinant, per Theorem S1, which is $O(n^3)$. Hence, the overall complexity is $O(2^m\, n^3)$. □

Since G is constrained to be the union of only k spanning trees with each $(n-1)$ edges, $m \le k(n-1)$. Hence, the algorithm is $O(2^{kn}\, n^3)$.

S2.3. Lower bound

In this section, we derive a lower bound for (S1). Consider the equation

$$G' = G + T_0 = G + T \tag{S2}$$

for unknown $T \in \tau(K)$ and known T_0. Let the set of solutions T of (S2) be $\lambda(G,T_0)$. It is clear that

$$Y(G+T_0,k+1) \supset \{S \cup \{T\} : S \in Y(G,k),\, T \in \lambda(G,T_0)\}$$

and hence

$$|Y(G+T_0,k+1)| \ge |Y(G,k)||\lambda(G,T_0)| \implies \frac{|Y(G+T_0,k+1)|}{|Y(G,k)|} \ge |\lambda(G,T_0)|.$$

We therefore start by computing $|\lambda(G,T_0)|$. Consider the equivalence relation $\sim$ on V given by two edges being connected in $G'-G$ (whose vertices are V and edges are $E' \setminus E$). Let $G'' = (\mathcal{V},\mathcal{E})$ be the quotient graph $G'/\sim$. Let $\pi : G' \to G''$ denote the quotient map.

Lemma 2. *Suppose C is a simple cycle of G'. Then, $\pi(C)$ is either a single point, a single edge connecting a pair of vertices, or a simple cycle of G''.*

Lemma 3. *Suppose*

$$|\lambda(G, T_0)| = \sum_{\mathcal{T} \in \tau(G'')} \prod_{e \in \mathcal{E}(\mathcal{T})} |\pi^{-1}(e)|$$

where $\mathcal{E}(\mathcal{T})$ *is the edge set of* $\mathcal{T}$*.*

Proof. Let $\mathcal{T} \in \tau(G'')$. We construct a spanning tree $T \in \lambda(G, T_0)$ as follows. For each edge $e = (v, w) \in \mathcal{T}$ choose a representative $(v', w') \in \pi^{-1}(e)$ and set (v', w') to be an edge in T. We then add in all the edges of $G' - G$. Clearly the resulting graph satisfies (S2) because it contains every edge of $G' - G$. It is spanning and connected because it connects every connected component of $G' - G$ and contains every edge in $G' - G$. We show that it is acyclic. Suppose C is a cycle of T; without loss of generality we can assume it is simple. Note that $G' - G$ must be acyclic because its edge set is a subset of edges of the tree T_0, so C cannot be a subgraph of $G' - G$. Therefore $\pi(C)$ is not a single vertex. If $\pi(C)$ is a single edge, then C contains two distinct edges joining two components of $G' - G$, which is not possible by our construction of T. Therefore by the previous lemma $\pi(C)$ is a simple cycle of $\mathcal{T}$ which is a contradiction.

Every $T \in \lambda(G, T_0)$ can be constructed in this manner. Since T is a tree it must descend to a tree $\mathcal{T}$ in G''. An edge $(v, w) \in T$ corresponds to the edge $([v], [w]) \in \mathcal{T}$; by choosing v and w as the representatives in the construction above we get back the edge (v, w).

It is also clear that distinct trees $\mathcal{T}_1$ and $\mathcal{T}_2$ in H will give different trees in G'. For a single $\mathcal{T}$, each distinct choice of representatives for each edge will give a distinct tree T. Therefore $|\lambda(G, T_0)|$ is the number of ways to carry out the above construction, which is given by the formula above. □

We now give weights to the edges of G'': let each edge $e \in \mathcal{E}(G'')$ have the weight $|\pi^{-1}(e)|$. The weight of a tree $\mathcal{T} \in \tau(G'')$ is defined as the product of the weight of its edges:

$$w(\mathcal{T}) = \prod_{e \in \mathcal{E}(\mathcal{T})} w(e).$$

Then

$$\lambda(G, T_0) = \sum_{\mathcal{T} \in \tau(G'')} w(\mathcal{T}).$$

The expression above is called the weighted tree enumerator of G''. The weighted Kirchoff matrix tree theorem gives a way to compute this quantity.

Theorem S2. (Formula for $\lambda(G, T_0)$.) *Let L be the Laplacian of G'' taking into account the weights of the edges, i.e.*

$$L = D - A$$

where A is the weighted adjancency matrix and D is the diagonal matrix whose ith entry is the sum of weights of edges incident to vertex i. Then $|\lambda(G, T_0)|$ is the absolute value of any cofactor of L.

Example 1. If G and G' have the same edges then $G' - G$ has no edges and $G = G' = G''$. In this case $\lambda(G, T_0)$ is the number of trees of G and this agrees with the formula above (the empty product is defined as 1).

S2.4. Upper bound

In this section, we derive an upper bound for (S1) by bounding $|\lambda(G, T_0)|$. Suppose $G = (V, E) \in \mathrm{im}(\Phi_k)$, so that for $1 \leq j \leq k$ there exists $T_j \in \tau(K)$ such that $E = \bigcup_{j=1}^{k} \mathcal{E}(T_j)$. Define $G_j = (V, E_j)$ by $E_j = \bigcup_{i=1}^{j} \mathcal{E}(T_i)$. We therefore have

$$G = T_k + G_{k-1}.$$

Using the results from above we get

$$|Y(G, k)| \geq |Y(G_{k-1}, k - 1)| \, |\lambda(G_{k-1}, T_k)|.$$

By iterating we get

$$|Y(G, k)| \geq \prod_{j=1}^{k-1} |\lambda(G_j, T_{j+1})|.$$

Now note that the ordering of the T_j does not matter to the argument. Let S_k denote the kth permutation group and for $\sigma \in S_k$ let $G_{j\sigma}$ denote $\bigcup_1^j T_{\sigma(i)}$. Then

$$|Y(G,k)| \geq \max_{\sigma \in S_k} \prod_{j=1}^{k-1} |\lambda(G_{j\sigma}, T_{\sigma(j+1)})|. \tag{S3}$$

This bounds the denominator in (S1). Next we address the numerator. Suppose $S \in Y(G', k+1)$. For each edge $e \in G'$, at least one of the trees in S must contain e. Therefore,

$$Y(G', k+1) \subset \left\{ \{T\} \cup S : T \in \tau_e(G'), S \in \binom{\tau(K)}{k} \right\}.$$

This gives the inequality

$$|Y(G', k+1)| \leq \min_{e \in E(G')} |\tau_e(G')| \, |\tau(K)|^k. \tag{S4}$$

The computation of $\tau_e(G')$ can be carried out as follows: let G/e be the graph obtained by contracting the edge e, let π be the quotient, and for any edge e' in G/e assign the weight $\pi^{-1}(e')$ to e'. Then $|\tau_e(G)|$ is simply the weighted tree enumerator of G/e; the proof is similar to the proof for Lemma 3. Equations S3 and S4 together give an upper bound:

$$\frac{|Y(G', k+1)|}{|Y(G,k)|} \leq \frac{\min_{e \in E(G')} |\tau_e(G')||\tau(K)|^k}{\max_{\sigma \in S_k} \prod_{j=1}^{k-1} |\lambda(G_{j\sigma}, T_{\sigma(j+1)})|}.$$

Altogether, we have:

Theorem S3.

$$|\lambda(G, T_0)| \leq \frac{|Y(G', k+1)|}{|Y(G,k)|} \leq \frac{\min_{e \in E(G')} |\tau_e(G')||\tau(K)|^k}{\max_{\sigma \in S_k} \prod_{j=1}^{k-1} |\lambda(G_{j\sigma}, T_{\sigma(j+1)})|}.$$

Figure S1 shows that the lower and upper bounds derived here for $|Y(G', k+1)|/|Y(G,k)|$ can be multiple orders of magnitude different from $|Y(G', k+1)|/|Y(G,k)|$. As such, they do not provide usable approximations for a Metropolis-Hastings acceptance ratio.

S3. Posterior inference on the cycle space using MCMC

We describe two different MCMC algorithms for posterior inference on the cycle space. Firstly, we consider general posteriors arising from prior distributions on spanning trees and associated cycle bases. Secondly, we consider the uniform prior over the cycle space which allows for computationally less expensive MCMC.

S3.1. General spanning tree prior $p(T)$

We consider the following prior construction for the cycle space $\mathcal{C}_n$. Note that the distribution over $\mathcal{C}_n$ can change with the basis C used: see for instance Corollary 1. Thus, we do not constrain our inference to a single C. Let $p(T)$ be the prior distribution on spanning trees which each correspond to a C. Then, a prior distribution on basis element inclusions induces a prior $p(G \mid T)$ over $\mathcal{C}_n$. The joint posterior of G and T follows then from the likelihood $p(X \mid G)$ in (3) as $p(G, T \mid X) \propto p(T)\, p(G \mid T)\, p(X \mid G)$.

We use Metropolis-Hastings within Gibbs to sample from $p(G, T \mid X)$. Specifically, we update the graph G using a Metropolis proposal $q(G' \mid G, T)$ that involves flipping one random basis element in G, i.e. flipping all edges in one uniformly sampled element of C. Thus, this is a multiple-edge update. Also, the proposal is guaranteed to be in $\mathcal{C}_n$ because the latter is a vector space. Note that the support of $q(G' \mid G, T)$ varies with T which determines the basis elements C to choose from. In other words, which graphs are in the "neighbourhood" of G varies with T, resulting in differing exploration of the graph space depending on the current value of T.

For the spanning tree T, we use its prior distribution as an independent Metropolis-Hastings proposal, $q(T' \mid T) = p(T')$. A change in T corresponds to a change in basis C and thus requires the decomposition of G in terms of this new basis. This decomposition involves a matrix inverse over the field $\mathbb{Z}_2$ which is computationally expensive for a large number of vertices n as the size of the matrix to invert is super-exponential in n. Therefore, we repeat the update of G multiple times for each update of T. Algorithm S1 summarises the resulting MCMC algorithm.

Algorithm S1 MCMC step for $p(G, T \mid X)$ for a general spanning tree prior $p(T)$

1. Perform the Metropolis update for G a fixed number of times:

   (a) Sample a proposed G' from $q(G' \mid G, T)$.

   (b) Set $G = G'$ with probability $\min(1, \alpha_G)$ where

   $$\alpha_G = \frac{p(G', T \mid X)\, q(G \mid G', T)}{p(G, T \mid X)\, q(G' \mid G, T)} = \frac{p(G' \mid T)\, p(X \mid G')}{p(G \mid T)\, p(X \mid G)}$$

2. Perform the Metropolis-Hastings update for T:

   (a) Sample a proposed T' from $q(T' \mid T) = p(T')$.

   (b) Compute the decomposition of G in terms of the basis generated by T' for the evaluation of $p(G \mid T')$.

   (c) Set $T' = T$ with probability $\min(1, \alpha_T)$ where

   $$\alpha_T = \frac{p(G, T' \mid X)\, q(T \mid T')}{p(G, T \mid X)\, q(T' \mid T)} = \frac{p(G \mid T')}{p(G \mid T)}$$

To compute $\frac{p(X|G')}{p(X|G)} = \frac{I_{G'}(\delta+N, D+U)}{I_G(\delta+N, D+U)} \frac{I_G(\delta, D)}{I_{G'}(\delta, D)}$ in Step 1b of Algorithm S1, we use two different approximations. For $I_{G'}(\delta + N, D + U)$ and $I_G(\delta + N, D + U)$, we use the Laplace approximation from Lenkoski & Dobra (2011). For $\frac{I_G(\delta, D)}{I_{G'}(\delta, D)}$, we use the approximation for the ratio of G-Wishart normalising constants from Mohammadi et al. (2021). This is because the Laplace approximation is known to be inaccurate when the degrees of freedom δ is small (Lenkoski & Dobra, 2011). In our case, G and G' differ by at least three edges while Mohammadi et al. (2021) consider single edge updates. We therefore approximate $\frac{I_G(\delta, D)}{I_{G'}(\delta, D)}$ as the product of a sequence of (at least three) ratio approximations from Mohammadi et al. (2021) each involving a single edge change.

S3.2. Uniform prior over the cycle space

We now discuss the case with $p(G \mid T)$ being the uniform prior over $\mathcal{C}_n$ as it allows for simplifications. Then, the choice of spanning tree T is irrelevant per Proposition 2, negating the need for the expensive decomposition in Step 2b in Algorithm S1. Also, sampling a cycle of three vertices uniformly at random and flipping the corresponding edges in G as proposal yields $p(G' \mid G) = p(G \mid G')$. Note that $p(G')/p(G) = 1$ by construction. Thus, the Metropolis acceptance ratio follows as $\alpha = p(X \mid G')/p(X \mid G)$. We use this MCMC algorithm that proposes to change a uniformly sampled 3-cycle in Section 5.

We use this MCMC algorithm also for the uniform prior over all graphs $\mathcal{G}_n$ and over all decomposable graphs with the modification that the proposal $p(G' \mid G)$ follows from randomly flipping one instead of three edges for the results in Section 5. We run this algorithm for 10^6 iterations with 10^5 burn-in iterations for $\mathcal{C}_n$ and $\mathcal{G}_n$, and for 10^7 iterations with 10^6 burn-in iterations for the decomposable graphs: we use more MCMC iterations for the decomposable graphs since many Metropolis proposals G' resulting from flipping an edge are rejected because they are not decomposable, and thus have no prior nor posterior mass, resulting in worse mixing of the MCMC chain. Implementation of the MCMC algorithms and code used to generate the results in Section 5 are available from `https://github.com/kristoforusbryant/cbmcmc`.

References

Chaiken, S. & Kleitman, D. J. (1978). Matrix tree theorems. *J. Comb. Theory Ser. A* **24**, 377–381.

Duan, L. L. & Dunson, D. B. (2021). Bayesian spanning tree: Estimating the backbone of the dependence graph. arXiv:2106.16120v1.

Greenhill, C., Isaev, M., Kwan, M. & McKay, B. D. (2017). The average number of spanning trees in sparse graphs with given degrees. *European J. of Combin.* **63**, 6–25.

Højsgaard, S., Edwards, D. & Lauritzen, S. (2012). *Graphical Models with R.* Springer, New York.

Kapoor, S. & Ramesh, H. (2000). An algorithm for enumerating all spanning trees of a directed graph. *Algorithmica* **27**, 120–130.

Lenkoski, A. & Dobra, A. (2011). Computational aspects related to inference in Gaussian graphical models with the G-Wishart prior. *J. Comput. Graph. Statist.* **20**, 140–157.

Mohammadi, R., Massam, H. & Letac, G. (2021). Accelerating Bayesian structure learning in sparse Gaussian graphical models. *J. Am. Statist. Assoc.* **Advance online publication**.

Schwaller, L., Robin, S. & Stumpf, M. (2019). Closed-form Bayesian inference of graphical model structures by averaging over trees. *Journal de la Société Française de Statistique* **160**, 1–23.

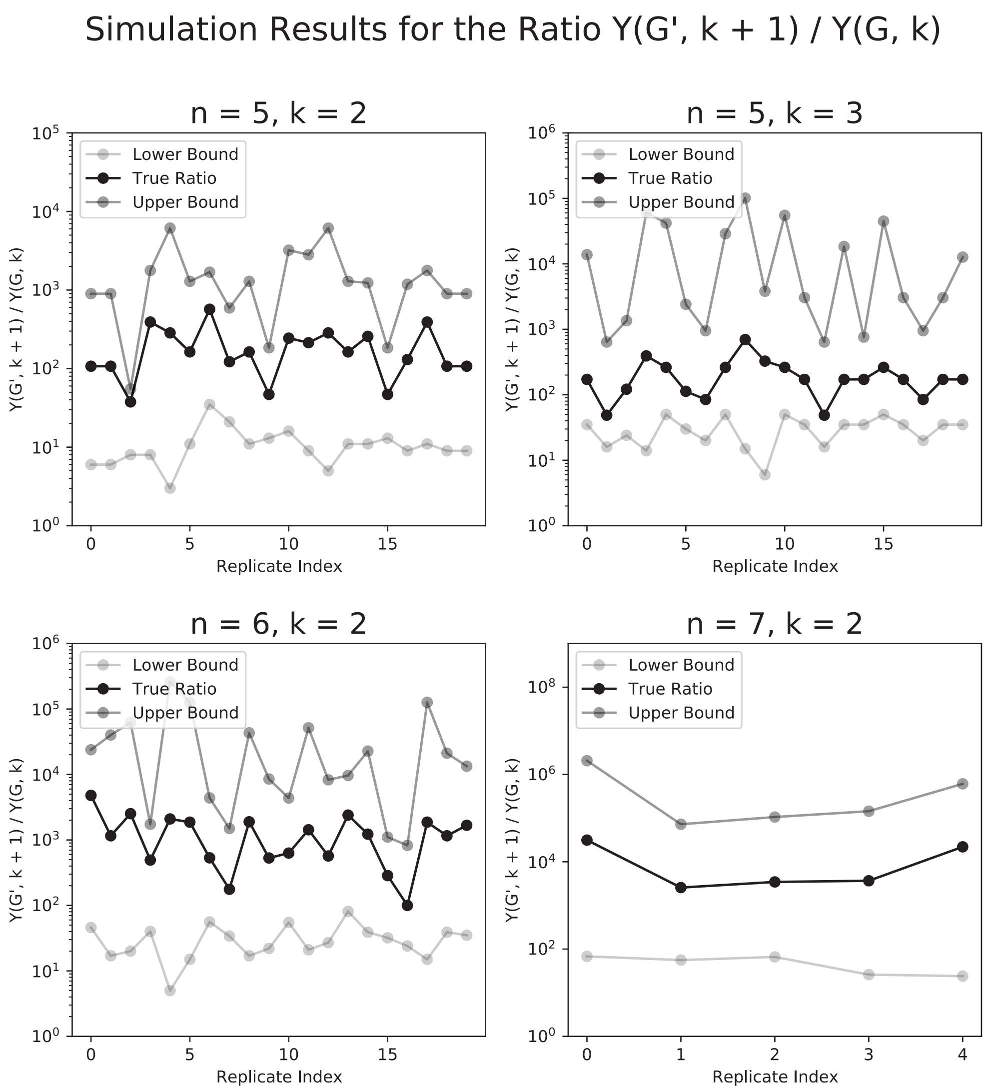

Figure S1: The value, lower bound from Section S2.3 and upper bound from Section S2.4 of $|Y(G', k+1)|/|Y(G,k)|$ on a logarithmic scale for fixed number of vertices n and number of spanning trees k. Here, the replicates consider randomly sampled G and G' where G is a union of k uniformly sampled spanning trees and $G' = G + T_0$ with T_0 a uniformly sampled spanning tree.